Why young children fail to understand *'before'* and *'after.'*

A thesis submitted to Lancaster University for the degree of Doctor of Philosophy in the Faculty of Science and Engineering

June 2016

Liam Blything, M.Sc.

Department of Psychology

Declaration

This thesis is my own work and no portion of the work referred to in this thesis has been submitted in support of an application for another degree or qualification at this or any other institute of learning.

Liam Blything 7th June 2016

Abstract

The goals of the thesis were to identify the development of 3- to 7-year-old children's comprehension and production of two contrasting temporal connectives - *before* and *after* - that signal the order of events in two-clause sentences, and to establish the reasons for difficulties with these linguistic devices. Chapter 1 reviews the literature that is considered relevant to the experimental work.

In the experimental work (Chapters 2 to 4), children's comprehension and production of two-clause sentences containing *before* and *after* was examined in separate groups of children aged 3 to 7 years. The sentence structures differed in their memory and also language demands. Independent measures of memory and language were related to performance. The design enabled a contrast of traditional memory capacity accounts (e.g., Just & Carpenter, 1992) versus more recent language-based accounts (e.g., Van Dyke, Johns, & Kukona, 2014) of why working memory explains variance in the processing of complex sentences. A capacity account predicts a direct relation between memory and sentence processing: specifically, that some sentence structures are more difficult to process than others because they require more information to be held in working memory than others. Alternatively, a language-based account proposes an indirect relation between memory and sentence processing, such that good language skills modulate the influence of memory on sentence processing, by influencing the accurate representation of information in verbal working memory.

Experiment 1 (Chapter 2) was a touch-screen comprehension paradigm. Children listened to two-clause sentences linked by a temporal connective, *before* or *after*, while viewing animations of the actions in each clause. After each sentence, they were asked to select the event that happened first to assess their understanding of the temporal connective. The pattern of results suggested that the memory demands of specific sentence structures

limited children's comprehension of sentences containing temporal connectives, supporting a memory capacity account.

Experiment 2 (Chapter 3) further investigated comprehension of these sentences focusing on how memory and language influence the ease of processing. Children were trained to make speeded responses to the sentence structures investigated in Experiment 1. The findings support Experiment 1: memory capacity best predicted comprehension of these sentence structures.

Experiments 3 and 4 (Chapter 4) examined production of the same sentence types. In two experiments (elicited production with blocked conditions, and sentence repetition), separate groups of children viewed an animated sequence of two actions, and were asked to describe the order of events. Instructions and practice trials were used to model the target sentence structures. In contrast to the comprehension experiments (Experiments 1 and 2), this work showed that children's individual differences in the production of two-clause sentences linked by *before* or *after* were related to variability in language skills, rather than poor memory capacity.

In Chapter 5, I conclude that Experiments 1-4 reveal a differential influence of working memory and language on children's comprehension and production of two-clause sentences containing *before* and *after*. I argue that the existing theoretical accounts of the influence of memory and language on sentence processing (e.g., Just & Carpenter, 1992; Van Dyke et al., 2014) require much more detailed investigation within the sentence structures examined here, and across other complex sentences that are also considered to differ in their memory and language demands. I present several suggestions as to how this might be accomplished in future work.

Acknowledgements

First and foremost, I feel privileged to have been supervised by Professor Kate Cain, who has been extremely generous with her expertise and time in providing me with every opportunity to develop the skills needed to pursue a successful future in academia. I also thank my second supervisor Dr Rob Davies for his valuable contributions to the Ph.D. work. Both Kate and Rob inspire me by their academic achievements, enthusiastic attitudes, and personability, to make the most out of life!

My thanks must also go to fellow colleagues and students. Dr Nicola Currie and Gill Francey have always made themselves available for advice, and have helped make our lab group meetings productive and enjoyable. Further, I have shared great research chats and humour with Matt Hilton and Ryan Blything, who both began their respective Ph.D.s at the same time as me in October 2012.

I also extend my gratitude to the Economic and Social Research Council for funding my Masters and Ph.D. work.

Finally, I thank my parents Don and Fiona, and my gran Bette, for keeping me well rounded and fully encouraging my enthusiasm and ambitions in life. These characteristics have shaped my positive attitude towards the Ph.D.

Contents

List of Tables

List of Figures

Author note: My submission for a PhD is via publication. Therefore, I would like to further acknowledge the contribution of my co-authors: Professor Kate Cain is a co-author for Chapters 2, 3, and 4. Dr Rob Davies is a co-author for Chapter 2.

1. Literature Review

1.1. Introduction and outline

We experience events in our everyday lives in the order in which they occur. In spoken or written discourse, however, temporal connectives such as *before* and *after* can be used to report the events in either their chronological order of occurrence, for example, '*He finished his homework, before he played in the garden*' or in reverse order, for example, '*Before he played in the garden, he finished his homework*'. Therefore, temporal connectives signal the actual order of occurrence.

It is well known that adults recognise the difference between *before* and *after*, using them accurately in both comprehension (Münte, Habets, & Jansma, 1998; Ye et al., 2012) and production (Habets, Jansma, & Münte, 2008; Ye, Habets, Jansma, & Münte, 2011). However, although *before* and *after* appear in spontaneous speech from around 3 years of age (Diessel, 2004), children display difficulties understanding and producing sentences containing these connectives up to at least early adolescence (e.g., Peterson & McCabe, 1987; Pyykkönen & Järvikivi, 2012). Therefore, it is not yet clear to what extent children can accurately use *before* and *after* to understand and signal the temporal relations between events (Cain & Nash, 2011; Pyykkönen & Järvikivi, 2012).

My thesis presents a series of experiments that have used accuracy and timing measures of comprehension and production to investigate children's understanding of two-clause sentences containing the temporal connectives *before* and *after* in 3- to 7-year-olds. The main aims were to identify the age at which early competence emerges, and to elucidate the reasons for why children continue to experience difficulties in the comprehension and production of sentences containing temporal connectives once they demonstrate an appreciation of the difference between *before* and *after*. In turn, this work enabled an investigation into whether performance in comprehension was similar to production. I next outline the relevant literature that motivates the experimental work presented in Chapters 2, 3, and 4.

1.2. Children's understanding and production of connectives

Connectives are cohesive devices such as *before*, *because*, and *although,* that help language users to establish coherence, because they signal the nature of the relation between events (Gernsbacher, 1997). Connectives are grouped into semantic classes that each signal a similar type of coherence relation. For example, temporal connectives (*before, after*) signal that the sequence of events relate in time; causal connectives (*so, because*) signal the causal relations between events, and adversative connectives (*but, although*) indicate information that is contrary to expectation.

Skilled comprehenders access word meanings and assemble these into meaningful clauses. They go beyond single clauses and integrate these to form coherent sentences, and to link the meanings of sentences to build a coherent mental representation of the combined meaning of those various propositions (Gernsbacher, 1990; Johnson-Laird, 1983; Van Dijk & Kintsch, 1983; Zwaan & Radvansky, 1998). This mental representation is typically referred to as a mental model or situation model. In the same way, speakers produce discourse in which the individual clauses and sentences are interrelated in meaning to refer to a coherent overall topic or theme.

Previous research shows that skilled adult comprehenders benefit from connectives in text. Take two events: '*Tom drove fast. He loves going to football matches*.' When linked by the causal connective *because*, adults read the second sentence more quickly than when no connective is present, because it signals the need to make an inference about the causal relation between the events (Cozijn, Noordman & Vonk, 2011; Sanders & Noordman, 2000; Traxler et al., 1997). Findings such as these indicate that connectives help adult comprehenders in the integrative and inferential skills that are necessary for linking clauses and establishing coherence (Zwaan, Langston, & Graesser, 1995).

In contrast to the facilitative effects seen for adults, research with children suggests that connectives might hamper, rather than help, children's text comprehension. Evidence for this comes from studies of 7- to 10-year-olds who have had difficulties with selecting an appropriate connective to join two clauses in cloze tasks (Cain & Nash, 2011; Cain, Patson, & Andrews, 2005). However, those studies found that children typically perform at above chance levels, which indicates some understanding about the function of connectives in a sentence and their specific meanings. Online processing paradigms have also indicated that children in this age range display knowledge of the function and meaning of connectives. In a self-paced reading paradigm, Cain and Nash (2011; Experiments 3 and 4) reported that 8-year-olds performed like adults: they read a clause more quickly when it was linked to a preceding clause by an appropriate connective (e.g., *Amy wanted a dog but she was not allowed one*) than when no connective was present (e.g., *Amy wanted a dog. She was not allowed one*) or when *and* was used to link the clauses. This and other studies also suggest developmental improvements: between 8 and 10 years, children demonstrate a more refined understanding for the function of connectives in both offline and online tasks (Cain et al., 2005; Cain & Nash, 2011).

Turning to production, corpus and experimental studies indicate that children can produce a range of connectives from around 3 years of age (e.g., Spooren & Sanders, 2008; Winskel, 2003). However, studies by Peterson and McCabe (1983; 1987; 1991) indicate that young children who appear to be using connectives in their speech, are not yet fully competent in their use. An analysis of personal narratives showed that 9-year-olds were just as likely as 4-year-olds to use *and* instead of a more appropriate connective that specifies a specific semantic relation between the events. The authors suggest that these age groups use connectives as general linking words between sentences rather than intending the same target meaning used by adults. Further, children frequently used connectives for pragmatic purposes such as narrative initiation or termination, rather than for semantic purposes, as demonstrated by examples i and ii below.

i. Experimenter: '*I bet you saw the sun come up this morning*.'

Child: '*But I saw the zoo…*'

ii. Child (following description of a car crash incident: '*…So they dead right now*.'

Peterson and McCabe explained this pattern of usage from a developmental perspective: young children start out using a connective to signal a range of relations between events and, through experience, develop a more refined understanding for the function of specific connectives. Therefore, this work in production also suggests that children do not fully understand how to use semantically restricted connectives in the same way as adults.

1.2.1. Why focus on temporal connectives?

Within each of the semantic classes already outlined (e.g., temporal, causal, adversative), the characteristics of the connectives themselves (e.g., Evers-Vermeul & Sanders, 2009; Spooren & Sanders, 2008), as well as their frequency of occurrence from environmental input (Crosson & Le Saux, 2008), may influence ease, and therefore age, of acquisition. Therefore there is a need to investigate the comprehension and production of specific connectives within a given semantic class, particularly from a developmental perspective whereby some connectives may be more easily acquired than others (Cain & Nash, 2011).

Temporal connectives typically appear in speech earlier than causal and adversative connectives (Bloom, Lahey, Hood, Lifter, & Fiess, 1980; Diessel, 2004; Shapiro & Hudson, 1991). Temporal connectives can also be divided into two subgroups: sequential connectives (*then*, *before* and *after*) signal that event x follows event y; whereas simultaneous connectives (*while*, *whilst* and *as*) signal that events happen at the same time. Sequential connectives are typically acquired earlier than simultaneous connectives (Atanassova, 2001; Keller-Cohen, 1981, 1987; Silva, 1991; Vion & Colas, 2004; Winskel, 2003, 2004, 2007). Therefore, an investigation into children's understanding of temporal connectives, particularly sequential ones, can indicate how early on in development children are able to use connectives to aid their comprehension and production of two-clause sentences.

The experimental work in this thesis examined children's understanding of *before* and *after* in two-clause sentences. It is unclear when children display the ability to comprehend and produce these sentences. As described earlier for comprehension, Cain and Nash (2011) reported that 8-year-olds performed above chance in offline comprehension tasks that required knowledge of *before* and *after*, and were faster in online tasks at integrating the relation between events when these connectives were used instead of a full stop (albeit the latter analysis did not examine specific connectives separately, but treated them as single class). Therefore, by 8 years, children display some level of competence in comprehending these connectives. However, those children were significantly less accurate compared to adults, and additional research has reported that children up to 12 years of age have difficulties in accurately comprehending two-clause sentences containing *before* and *after* (Pyykkönen & Järvikivi, 2012). Research on children's ability to produce these connectives tells a similar story to comprehension: both corpus and experimental work shows that children regularly produce sentences containing *before* and *after* from as young as 3-years-old (Diessel, 2004; Winskel, 2003), however, children display difficulties with correct production of these connectives up to at least 9 years (Peterson & McCabe, 1987; Winskel, 2003).

One reason for the lack of clarity about when children truly understand sentences containing *before* and *after,* is that these connectives can appear in a range of sentence structures, and that children find some of these more difficult than others (e.g., Clark, 1971; Pyykkönen & Järvikivi, 2012). Specifically, when these connectives are used to signal the order of events, the events can be written in either their chronological order of occurrence, for example, '*He finished his homework, before he played in the garden*', or in reverse order, for example, '*Before he played in the garden, he finished his homework*.' As a result of this language flexibility, the same underlying temporal information can be expressed in four different sentence structures: before-chronological, after-chronological, before-reverse, and after-reverse (see Table 1.1). This means that studies of children's competence for two-clause sentences containing *before* and *after* that do not consider the full range of possible sentence structures, may belie children's full competence. The remainder of this chapter will review the factors within these sentence structures that may influence sentence comprehension and production, focusing on accounts that can inform our understanding of development. First, I outline the processes and mechanisms involved in constructing a coherent mental representation, and then I explore the factors that may affect how accurately children use temporal connectives to construct a mental representation that encodes temporal relations between events (section 1.4).

Table 1.1.

Chronological and reverse order sentence structures containing before and after.

	Before	After
Chronological	He put on the sandals, *before* he ate the burger	*After* he put on the sandals, he ate the burger
Reverse	*Before* he ate the burger, he put on the sandals	He ate the burger, *after* he put on the sandals

1.3. Mental representations of temporal information

As noted, comprehenders construct a mental representation of the situation described by the text or discourse (Gernsbacher, 1990; Johnson-Laird, 1983; Van Dijk & Kintsch, 1983; Zwaan & Radvansky, 1998). The construction of the mental representation requires the simultaneous processing and storage of information in working memory, and is ultimately stored in long term memory. The majority of theoretical models (and empirical evidence) for how we mentally represent text and discourse focus on comprehension rather than production and, consequently, this section will describe the relevant processes with greater reference to comprehension. Of these theoretical models, the event-indexing model (Zwaan, Langston, & Graesser, 1995; Zwaan & Radvansky, 1998) speaks most to the present thesis because it separately indexes different dimensions, including time, which is the information provided by temporal connectives.

With relevance to temporal information, the event-indexing model explains that the comprehender foregrounds a mental substructure for initial temporal information (e.g., the first clause), and will more easily incorporate incoming information (e.g., the second clause) into an updated version of the foregrounded substructure if it is coherent, for example within the same timeframe.[1] Conversely, if incoming temporal information indicates a time shift from the foregrounded substructure, the comprehender will have difficulty integrating such information to their current representation. This would necessitate the encoding of a new foregrounded substructure, making the previous substructure less accessible in memory. The resulting mental representation is likely to consist of several branching substructures, each of which will differ by their levels of activation in memory. The strength of activation (or re-

---

[1] Note that the processes described here overlap in essence with Gernsbacher's (1990) structure building framework: foregrounding is similar to the concept of '*laying a foundation*'; the updating of incoming information is similar to '*mapping*'; and the building of new substructures when information is incoherent is similar to '*shifting*.' The event-indexing model is unique in its separate indexing of dimensions such as time (Zwaan & Radvansky, 1998)

activation) for new or previous substructures is determined by the continuity of incoming information with those substructures.

It is important to note that the construction of coherent mental representations of discourse meaning is also necessary for successful production. For example, studies of schizophrenic adults, who display disordered discourse, have reported that in comparison to controls, they take longer pauses between clause boundaries when describing a story (e.g., Ditman & Kuperberg, 2010; Rochester & Martin, 1978; Traxler & Gernsbacher, 1995). The longer pauses between clause boundaries are interpreted as difficulties in foregrounding and updating mental structures of the discourse meaning. This indicates that, as in comprehension, production difficulties can be attributed to how easily the speaker can construct a coherent mental representation of discourse meaning. Recent research has also forcefully argued that adult comprehension and production share similar cognitive processes (see Pickering & Garrod, 2007; 2013). For example, both involve the prediction of upcoming language, and the processing of the language itself is largely determined by accessibility to knowledge of sentence structure and words (Pickering & Garrod, 2013). In addition, both speakers and comprehenders process information incrementally (Brown-Schmidt & Konopka, 2015; Ferreira, 1996). If the processes are similar, then children’s patterns of difficulty in producing sentences containing *before* and *after* and the factors underlying performance might be analogous to comprehension. The next section of the thesis explores how the processes involved in comprehension and production that have been outlined here, might be influenced by several factors within the sentence structure of two-clause sentences containing *before* and *after*.

1.4. Accounts of developmental gains in connective comprehension and production of two-clause sentences containing *before* and *after*

1.4.1. A non-linguistic strategy

A critical factor that influences children's understanding of two-clause sentences containing *before* and *after* is the degree to which they have an appreciation for the temporal information signalled by the different connectives. When young children have a fragile understanding of the relation signalled by a temporal connective, they can use different strategies to understand and represent the relation between two events, rather than using the precise linguistic information provided by the connective itself (Clark, 1971). Depending on the sentence structure, these strategies will sometimes result in accurate comprehension.

There are two non-linguistic strategies that children may use to interpret temporal order: an order of mention strategy or a world knowledge strategy. Each strategy can assist children's comprehension accuracy for *before* and *after* in specific sentence structures, so can belie their understanding for the connective itself. No previous study has identified a preference for one strategy over the other.

In support of the order of mention account, a number of comprehension studies have reported that 3- to 5-year-olds interpret sentences containing *before* and *after* more accurately when events are mentioned in a chronological order such as '*He put on the jumper, before ate the cookies'*, compared to a reverse order such as '*Before he ate the cookies, he put on the jumper'* (Clark, 1971; French & Brown, 1977; Johnson, 1975). Specifically, the advantage for chronological sentences displayed by the younger children is a result of their below-chance accuracy for reverse order sentences. This pattern of performance suggests that young children may assume that the actual order of events corresponds to the order in which events are reported, rather than the order that is signalled by the connective.

A study by Clark (1971) confirms that an order of mention strategy is used by children when they do not appreciate the temporal information signalled by *before* and *after*. Clark mapped out three developmental stages in 3- to 5-year-olds' comprehension of chronological and reverse order sentences containing *before* and *after*, measuring performance using an act-out task. A 75% accuracy criterion was chosen to indicate competence for each sentence type, whereas 25% or below reflected inaccuracy. Children in the first developmental stage were accurate on chronological sentences but not on reverse order sentences. The mean age for children at this stage was 3 years and 7 months. Clark concluded that these children did not have a sufficient understanding of either connective. Children in stage two performed accurately on before-reverse sentences as well as chronological sentences, but were inaccurate on after-reverse sentences (M = 4 years and 3 months). This result suggests that these children were using an order of mention strategy only for *after* sentences, but had sufficient understanding of *before*. She attributed earlier competence for *before* relative to *after* to the semantic features of each term: *before* indicates the prior event, whereas *after* does not, making the latter more semantically complex. In the final stage, children (M = 4 years and 7 months) reached the accuracy criterion for all four sentences (before-chronological, after-chronological, before-reverse, and after-reverse), indicating an understanding of both *before* and *after* earlier than 5 years of age.

Other research has reported that children employ a different non-linguistic strategy, interpreting events on the basis of the typical sequence of those events (world knowledge). In support of this view, 3- to 5-year-olds have been shown to be more accurate at acting out the sequence of events in a sentence when the sequence is typical and thus supported by world knowledge (e.g., *He put on the socks, before he put on the shoes*), compared to when event order was arbitrary and could not be supported by world knowledge (e.g., *He put on the socks, before he ate the burger*) (French & Brown, 1977; Keller-Cohen, 1987). When sentences relate typical sequences of events the order can be inferred by world knowledge alone. When sentences relate events that share an arbitrary relation, the order cannot be inferred by world knowledge as the events can occur in either temporal order. Instead, the child must use the connective to identify the temporal order for accurate performance on all sentence types.

1.4.2. Factors that influence sentence processing in the immediate years that follow the emergence of an initial understanding of these connectives.

From around 5 years of age, children no longer rely on non-linguistic strategies to understand and represent order (Clark, 1971). There are several factors that might influence older children's comprehension and production of two-clause sentences containing *before* and *after*. These factors include the relation between the order of mention of the events and the order of events being described by the connective; the depth of understanding and ease of accessibility for the meaning of the connective; the position of the connective in a sentence, and the availability of world knowledge in a sentence. These may explain reasons for children's inaccurate comprehension and production of these sentences in the immediate years that follow the emergence of an initial understanding of these connectives. Each will be discussed in turn.

Order of mention. As noted, several studies have reported that 3- to 5-year-olds are more likely to be accurate in comprehension tasks for sentences in which the order of mention of events corresponds with the actual order of occurrence signalled by the connective. This has been used as evidence of an incomplete understanding of the meaning of *before* and *after* (e.g. Clark, 1971). However, even when an understanding for the connectives might be considered robust, as for adults, the order of mention of events appears to influence the processing effort that is required for sentence comprehension (Münte et al., 1998; Ye et al., 2012) and production (Habets et al., 2008; Ye et al., 2011).

CHAPTER 1: LITERATURE REVIEW

It is established that the mental representation of event order maps onto 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). As a result, the comprehender has a default expectation that new information (i.e., the event described in the second clause) will be temporally later than the most recent event in the current foregrounded substructure (i.e., the event described in the first clause) (Zwaan & Radvansky, 1998). Similarly, speakers also construct a mental representation of the events in the actual order that they occur (Bereiter & Scardamalia, 1987; Bos, Bjorn, de Koning, van Wesel, Boonstra, & van der Schoot, 2015). Therefore, reverse order sentences require more cognitive effort to comprehend and produce because the order of mention of events does not correspond with the linear mental representation of the events. This does not fully afford incremental word-by-word or clause-by-clause processing.

The extra processing effort required for comprehending and producing reverse order sentences in comparison to chronological sentences has been demonstrated in ERP and fMRI studies with adults (Münte and colleagues, 1998, 2008, 2011, 2012). Most recently, Ye et al. (2012) asked adults to read chronological sentences (*After the scientist submitted the paper, the journal changed its policy*) and reverse order sentences (*Before the journal changed its policy, the scientist submitted the paper*). Reverse order sentences were associated with stronger activation of the caudate network as a whole, which has been strongly associated with mental imagery processes such as visual rotation of pictures (Kucian et al., 2007). Moreover, the left middle frontal/precentral gyrus, which is associated with maintaining temporal order information in working memory (see Wager & Smith, 2003), was activated most for reverse order sentences. Together, these findings indicate that reverse order sentences require the comprehender (or the speaker) to revise their mental representation, and that this places extra demands on working memory.

Connective meaning. A second factor that is likely to influence the processing demands of the sentence structures is the differences between the meaning and function of the connectives. Previous theoretical work has argued that *before* should be more difficult to acquire than *after*. According to Clark's (1971) *semantic features* theory, words share a componential hierarchy and the positive value of a component is learned earlier than the negative value. Clark explained that *before* and *after* share three hierarchical components (+/-Time, +/-Simultaneous, +/-Prior). At the top of the hierarchy, both *before* and *after* represent time and so are classified as '+Time.' Second, both *before* and *after* represent a serial time relation, and so are classified as '-Simultaneous.' The third function classifies the direction for the serial sequence described by *before* and *after*: *before* indicates the 'Prior' event (+Prior) whereas *after* does not. Thus, according to Clark, *before* should be easier to understand than *after* because they differ in the direction of the serial relation that they signal. That is, *before* is semantically simpler than *after* because it carries more positive values.

In addition *after* may be more difficult than *before* because it functions more broadly as a grammatical device. Word frequency counts by Leech et al. (2001) from *The British National Corpus* (BNC) indicate that *after* functions more often (927 out of 1160 counts) than *before* (577 out of 882 counts) as a non-connective such as a preposition or adverb (e.g., *The dog chased after*). For that reason, it might be harder to learn the use of *after* as a connective.

Words that are typically more difficult to learn are processed by adults more slowly and less accurately than their less difficult to learn counterparts. For example, these difficulties have been demonstrated for words with a late age-of-acquisition (Carroll & White, 1973; Juhasz, 2005), a low frequency of occurrence (Coltheart, Rastle, Perry, Langdon, & Ziegler, 2001; Reichle, Pollatsek, Fisher, & Rayner, 1998) or a high ambiguity in meaning (Gunter, Wagner, & Friederici, 2003). Importantly, these processing costs are more pronounced in comprehenders with low working memory span compared to comprehenders with high working memory span, which suggests that more difficult words place additional demands on working memory resources (Gunter et al., 2003). Together, this literature indicates that even once the child acquires an appreciation for the difference in meaning between *before* versus *after*, their processing of sentences containing these connectives may still be affected by the connective because they are likely to find it more taxing on their working memory resources when activating their knowledge of *after* as a temporal connective compared to *before*.

Sentence position of the connective. Manipulating the order of mention for clauses that start with *before* and *after* also varies the position of the connective in the sentence. For example, when events are spoken in a chronological order, the connective appears in an initial position when *after* is used to establish the link (e.g., *After she combed her hair, she put on her gloves*), but appears in a medial position when *before* is used (e.g., *She combed her hair, before she put on her gloves*). The opposite applies to reverse order sentences. Children and adults display a preference for using temporal connectives in a medial position rather than an initial position (Diessel, 2004; 2008). A medial position is between the successive clauses, so the connective provides the linking information at a point close to when the events can be integrated during incremental processing (Cain & Nash, 2011; Traxler et al., 1997). Conversely, an initial position is associated with higher memory demands because it requires the individual to hold the connective information in working memory from the beginning of the sentence until the point at which the events can be integrated (Diessel, 2004; also see Hawkins, 2004).

World knowledge. World knowledge may also influence the ease with which these sentences are processed in the immediate years that follow the emergence of an initial understanding of these connectives. Reading time studies show that adults (e.g., Cozijn et al.,

2011) and children (e.g., Barnes, Ahmed, Barth, & Francis, 2015) make use of their world knowledge of the typical order of events to support sentence comprehension. For example, Barnes et al. reported that 11- to 18-year-olds were faster at reading sentences in which the events were typically related by a causal sequence (e.g., *Jane took the aspirin. Her headache went away*) than sentences in which the events were not typically associated (e.g., *Jane looked for an aspirin. Her headache went away*).

The reason why world knowledge supports sentence processing can be explained by theoretical models of mental representations of text and discourse (e.g., Zwaan & Radvansky, 1998). When incoming information is coherent with previous information it is highly activated in memory so is incorporated more easily into a foregrounded mental substructure. Therefore, when events follow a typical sequence, world knowledge can be used to infer order, for example that ketchup is typically poured on prior to eating a burger. For such an example, the order of events that is signalled by a temporal connective can be checked against world knowledge. In that way, world knowledge can support accurate sentence processing (Graesser et al., 1994; Metusalem, Kutas, Urbach, Hare, McRae, & Elman, 2012).

1.5. Memory capacity vs. language-based accounts of sentence processing

As noted, the construction of a mental representation of a sentence (or longer discourse) draws on working memory resources. Specifically, the processes of foregrounding and updating a mental representation require comprehenders (and speakers) to actively maintain the relevant substructures of meaning in working memory so that they can be integrated with incoming discourse and also with world knowledge (e.g., Zwaan & Radvansky, 1998). Indeed, a large number of studies have demonstrated that working memory predicts unique variance in comprehension for both adults (Carretti, Borella, Cornoldi, & De Beni, 2009; Daneman & Merikle, 1996) and children (Cain, Oakhill, & Bryant, 2004).

The classic theory concerning the role of working memory in sentence processing is the memory capacity account (e.g., Just & Carpenter, 1992). This posits that the structural demands of some sentences require a greater amount of information to be held in working memory, making it harder for the comprehender to retain a full and accurate representation. It follows that individuals with high working memory span should be more capable of maintaining relevant information to arrive at an accurate mental representation when comprehending text.

The majority of research on whether the memory demands of different sentence structures influence sentence processing has been related to comprehension rather than production. Consequently, the present section gives greater reference to comprehension, but is nevertheless considered equally applicable to production, as noted by Carpenter, Miyake, and Just (1994):

> '*Although we have focussed on the role of working memory in language comprehension, clearly the demands for concurrent computation and storage are equally crucial in language production* (p. 1112).'

The memory capacity account would predict that each of the factors outlined in section 1.4.2 influence sentence processing by varying the working memory demands. That is, working memory demands should be increased when a sentence has a reverse order of mention of events (see Ye et al., 2012), more difficult vocabulary (see Gunter et al., 2003) such as a later acquired connective (*after*; see Clark, 1971), an initial position of the connective (see Diessel, 2004), and when the order of events cannot be predicted by world knowledge (see Zwaan & Radvansky, 1998). Table 1.2 shows that before-chronological sentences should be easiest because their grammatical structure has no additional memory load (chronological order, earlier acquired connective, medial position), whereas the other structures each have two factors that increase the amount of information that must be held in working memory: before-reverse (reverse order, initial position), after-chronological (initial position, later acquired connective), and after-reverse (reverse order, later acquired connective). In addition, since the availability of world knowledge in sentences is associated with lower working memory demands and can support accurate processing, the influence of working memory on sentence processing may be more likely in sentences where the events cannot be predicted by world knowledge (i.e., in world knowledge absent sentences). Critically, if these factors exceed the available working memory capacity of the comprehender or speaker, then the correct interpretation will decay and be forgotten. Therefore, independent measures of working memory would be expected to explain unique variance in the performance differences across sentence structures.

Table 1.2.

Sentence structures and their additional working memory load as influenced by a reverse order of mention of events, a later acquired connective, and an initial position of the connective.

	Additional working memory load			
	Reverse order	Later acquired connective	Initial position	World knowledge absent
Before-chronological (world knowledge present)	No	No	No	No
After-chronological (world knowledge present)	No	**Yes**	**Yes**	No
Before-reverse (world knowledge present)	**Yes**	No	**Yes**	No
After-reverse (world knowledge present)	**Yes**	**Yes**	No	No
Before-chronological (world knowledge absent)	No	No	No	**Yes**
After-chronological (world knowledge absent)	No	**Yes**	**Yes**	**Yes**
Before-reverse (world knowledge absent)	**Yes**	No	**Yes**	**Yes**
After-reverse (world knowledge absent)	**Yes**	**Yes**	No	**Yes**

Given that the amount of information that can be held in working memory is often far less than the length of a complex sentence, it has been argued that working memory capacity cannot alone be an adequate explanation of sentence processing difficulties (McElree, 2006; Van Dyke & McElree, 2006). Alternatively, a language-based account argues that memory limitations on sentence processing are determined by whether language knowledge is rich enough to allow target concepts to be easily retrieved from long term memory (e.g., Kidd, 2013; Klem, Melby-Lervåg, Hagtvet, Lyster, Gustafsson, & Hulme, 2015; Van Dyke et al., 2014). That is, memory limitations are driven by the *quality* of the knowledge that must be retrieved from long term memory (e.g. Van Dyke et al., 2014), rather than by the *quantity* of retrieved information that can be maintained within working memory (e.g., Just & Carpenter, 1992). This argument draws on the framework that, rather than being separate systems, 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, rich language knowledge should support the role of memory in building a robust mental representation because the comprehender (or speaker) quickly accesses and accurately retrieves the target concepts, freeing up resources for maintaining an accurate representation of information in memory. Conversely, poor knowledge of language is likely to result in a more fragile mental representation because the comprehender (or speaker) is less able to suppress competing concepts which share some similarity to the target concept, using up processing resources.

Such a framework contrasts the memory capacity account (Just & Carpenter, 1992) outlined earlier, which draws on the framework that there are separate systems for working memory and long term memory (e.g., Baddeley, 2003; Baddeley & Hitch, 1974) to argue that the accurate representation of information is driven by the availability of processing resources that are specific to the working memory system. That is, memory limitations are driven by the quantity of information that can be stored within working memory. Since working memory is assumed to be an independent system, this effect should not be influenced by processes involved in the retrieval of information from long term memory. In support of a language-based account for sentence processing effects, recent research with adults has demonstrated that the specificity or distinctness of retrieval cues in the text (e.g., 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), rather than the quantity 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 & Johns, 2012; Van Dyke et al., 2014; Van Dyke & McElree, 2006). In these experiments, adult readers display greater processing difficulties for complex sentences (e.g., object relative clause sentences; *It was the boat that the guy who lived by the sea fixed after 2 day*s) when they are preceded by a list of recall words that have similar meanings to the target cues in the sentence. For the given example, the recall words *table, sink,* and *truck* would make it more difficult to retrieve the target meaning of the object (*boat*) of the verb *fixed* because each could semantically serve as objects of the verb. Critically, processing difficulties are not displayed when the recall words do not share similar features to the target cue, nor do they vary by the quantity of recall words. Also, independent measures of working memory capacity do not predict performance once measures of language knowledge are additionally incorporated (e.g., vocabulary, see Van Dyke et al., 2014).

Together, these findings indicate that performance is not driven merely by whether an additional load of information can be maintained within working memory. Instead, these experiments demonstrate that processing difficulties are determined by how well comprehenders use their language knowledge in the retrieval of the target information of the retrieval cues in the sentence. Thus according to this account, the influence of memory on processing difficulties is moderated by language knowledge.

If language-based accounts (e.g., Van Dyke et al., 2014) can explain developmental differences in children's comprehension and production of sentences containing *before* and *after*, then knowledge of the connective and of other words in the sentence should predict how well children perform with sentence structures that require more cognitive effort. In this thesis, knowledge of the connective was measured by manipulations within the experimental sentences (*before* versus *after*). A vocabulary test was used to assess language and served as a proxy measure for knowledge of the words within the sentence. Therefore, sentences with greater cognitive demands (i.e., reverse order sentences) would be expected to be more difficult when linked by *after*, which has greater semantic complexity and is later acquired, and this pattern of performance should also be driven by 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). For example, the memory capacity account predicts that since *after* is more complex than *before*, the presence of *after* in any sentence will increase the quantity of information that has to be actively maintained within an independent system of working memory. Therefore, the effect or interacting effect of the connective should be driven by working memory capacity alone. Conversely, the language-based account takes the perspective of a unitary storage system for long term memory and working memory (McElree, 2006), to argue that more processing resources are likely to be allocated to retrieval operations in long term memory for discriminating the meaning of *after* compared to *before*, so are less likely to be allocated to accurately representing complex sentences in memory (i.e., reverse order). Therefore, due to the association between language knowledge and working memory, the effects of working memory that are proposed by a memory capacity account (Just & Carpenter, 1992) would be expected to serve as a proxy for the processing difficulties that are moderated by language knowledge (e.g., Van Dyke et al., 2014).

To date, only one study has related children's difficulties with these sentences to additional processing demands during the construction of a mental representation, but it was not designed to disentangle the effects of memory versus language. Pyykkönen and Järvikivi (2012) asked 8- to 12-year-olds to read a sentence and to then circle whichever verb that they thought was first to temporally occur. Most notably, 8- to 12-year-olds displayed a difficulty with reverse order sentences but only when linked by the connective *after*. The authors argued that children had difficulty with after-reverse sentences because the connective did not signal a reverse order until half-way through processing the sentence, thus disrupting construction of the mental representation. In contrast, interpreting a reverse order sentence linked by *before* may not require revision of the mental representation, as reverse order of mention is signalled from the beginning of the sentence. Therefore, the authors argued that after-reverse sentences require additional working memory computations to hold the concepts for later reconstruction of the mental representation. It must be noted however, that the difficulty with after-reverse sentences may also align to the prediction of language-based accounts that were outlined earlier, especially considering that a significant disadvantage for *after* over *before* was reported. As noted, to help dissociate memory accounts, the present experimental work used independent measures of memory and language, in addition to manipulating these as factors within the experimental sentences.

1.6. Methodological limitations with previous research

1.6.1. Definitions of competence

Definitions of competence with temporal connectives can vary enormously. For example, overall competence could be explicated as (i) whether children are performing above chance overall (i.e., collapsed across conditions); (ii) whether children no longer display use of a non-linguistic strategy (i.e., they are performing above chance in each specific sentence structure); or (iii) whether children are performing near ceiling and therefore applying a full range of sentence processing skills. The majority of previous studies have been conducted in relation to the first two levels of competence outlined here (e.g., Clark, 1971). However, identifying the age that children perform above chance-level or when a non-linguistic strategy disappears can, at best, only inform of us of the age that children begin to improve upon their fragile understanding for the meaning of the connective. Competence cannot be concluded by this alone, as children continue to have difficulties up to early adolescence (Cain & Nash, 2011; Peterson & McCabe, 1987; Pyykkönen & Järvikivi, 2012). This motivated my experimental work to determine the reasons for why children continue to have difficulties with these sentences in the immediate years that follow the emergence of an initial understanding for the meaning of the connective. To do this, I contrasted the roles of memory and language skills to investigate how the structural demands of specific sentence structures influence processing.

One way to examine the role of memory versus language demands on sentence processing is to consider how they are manipulated within the structural demands of the sentences (i.e., order of mention and connective). This consideration is absent in studies of the production of sentences with temporal connectives: no previous study has used a paradigm that examined all possible sentence structures. In addition, no previous comprehension or production study has used independent measures of memory or language to gain insight into whether these skills can predict variation in performance for sentences that carry different structural demands on processing (e.g., Just & Carpenter, 1992; Van Dyke et al., 2014). I now outline other limitations in previous studies of children's comprehension and production of temporal connectives, which motivate the experimental work of the present thesis.

1.6.2. Comprehension tasks

Limited sample sizes. Of the 11 previous studies on children's comprehension of sentences containing *before* and *after* (Amidon & Carey, 1972; Clark, 1971; Crain, 1982; French & Brown, 1977; Gorrell, Crain, & Fodor, 1989; Hamburger & Crain, 1982; Keller-Cohen, 1987; Johnson, 1975; Pyykkönen & Järvikivi, 2012; Stevenson & Pollitt, 1987; Trosborg, 1982), 10 have less than six items per condition (the exception being Clark, 1971).

Studies are typically recommended to have a minimum of eight items per condition (Field, 2009), so these previous studies were short of reliable statistical power. This is important in the light of recommendations for more psychological studies to provide adequate power so that false positive and false negatives is minimised (see Cohen, 1992).

Additional task demands. A major limitation of previous comprehension studies is that the paradigms used may have underestimated performance because of additional demands. Of the 11 previous studies on children's comprehension of temporal connectives, 10 have used an act-out task (the exception being Pyykkönen & Järvikivi, 2012). However, the requirement to act-out a sentence with toys creates additional memory demands because children must store both clauses in memory whilst planning and acting out the sequence (Ambridge & Lieven, 2011). Indeed, Amidon and Carey (1972) and French and Brown (1977) cited this as a likely reason for why 3- to 5-year-olds often only acted-out one of two clauses that were linked by *before* or *after,* in their studies.

In addition, the concept of the act-out task may be difficult for children to understand. For example, to satisfy the presupposition in the instruction '*Before you move the red plane, move the blue plane*', the intention for the action in the subordinate clause should be established. To do this, Hamburger and Crain (1982) established a playing context prior to each item by asking the child to choose which toy they would like to use. That toy would then feature in the subordinate clause of the instruction (e.g., *Before you move the red plane*). Results indicated that the satisfaction of presuppositions vastly improved performance and children no longer appeared to fall back on a non-linguistic strategy. Further confusion over the purpose of act-out tasks may be caused by the materials used. For example, Keller-Cohen (1987) included materials that required children to suggestively act-out events such as opening a toy can and pouring the (unopened) toy can even though the state of the object was unaltered (e.g., the lid did not open, and soup did not pour out the can). This may have caused children to give a default response where they ignored the wording of the sentence and instead played with the toys in the order they considered to be most sensible (i.e., their natural tendency). For these reasons, previous research may have underestimated competence.

One comprehension study that did not use an acting-out task, was Pyykkönen and Järvikivi's (2012) written judgement paradigm, which was described in section 1.5. However, their study was of older children (aged 8 to 12) so does not examine the immediate years that follow the emergence of an initial understanding of the meaning of the connective. In addition, their study measured reading comprehension rather than listening comprehension. Therefore, word reading ability may have influenced performance, as it is a crucial requirement in reading comprehension (Cain et al., 2004; Gough & Tunmer, 1986; Juel, Griffith, & Gough, 1986). Further, in their task, children read the sentence, and then had to read a choice of answers (i.e., the first versus the second verb) and circle the one that they considered to have happened first. Similar to in act-out tasks, this may have increased processing demands because the child had to process additional information whilst maintaining their representation of the sentence in working memory. Together, these factors may provide an inaccurate picture of children's early competence.

One means to reduce the memory demands, is to use forced-choice paradigms such as picture-sequencing or touch-screen technology. Previous studies have examined children's basic understanding of temporal order by asking them to organise pictures into a temporal sequence (Brown & French, 1976; Brown & Murphy, 1975; Trosborg, 1982). An even simpler requirement for children would be to use a touch-screen to indicate their understanding of the temporal order, for example, touching the first mentioned event. This reduces the additional memory demands that are associated with the time and effort required in making a response. Indeed, Friend and Kelpinger (2008) reported that vocabulary performance in young infants was significantly better when using touch-screen technology compared to an analogous conventional picture identification task, indicating that the touch-screen task had fewer additional demands.

Furthermore, no previous developmental research has examined processing strategy for sentences containing *before* and *after*. Online research with children and adults has provided insights into the benefit of connectives in two-clause sentences and the processes involved in accurate reading comprehension (Cain & Nash, 2011; Cozijn, Noordman, & Vonk, 2011; Traxler et al., 1997). Only two previous studies have used online measures of children's comprehension of two-clause sentences containing connectives, both choosing a self-paced reading task (Cain & Nash, 2011; Mouchon, Fayol, & Gaonac'h, 1995). Both of those studies reported that 8-year-olds read the second clause faster when it was linked by an appropriate connective to the previous clause compared to when it was linked by no connective or by an inappropriate connective. However, both studies examined the presence versus the absence of appropriate connectives in general, so do not speak directly to the focus of the thesis on *temporal* connectives. Self-paced reading cannot be used with the age group of interest in this thesis because it involves non- and beginner-readers (3- to 7-year-olds). Instead, a listening comprehension task with a two-option picture choice on a touch-screen was used to record accuracy and response times, which have been previously used with children as young as three as a reflection of their mental representations (Möhring, Newcombe, & Frick, 2014).

1.6.3. Production tasks

The majority of studies examining children's production of sentences containing temporal connectives have used an elicited production paradigm. In this, the child is asked to describe the events after being shown a sequence of pictures (Silva, 1991; Vion & Colas, 2004; Winskel, 2007) or after viewing an acting-out of events with toys (e.g., Atanassova, 2001; Clark, 1971; Weist, Lyytinen, Wysocka & Atanassova, 1997). Other more naturalistic work has examined the use of connectives in corpora of children's language (e.g., Diessel, 2004). The problem with elicited production and corpus studies is that they cannot experimentally manipulate whether the speaker uses the connectives in a specific sentence structure (Ambridge & Lieven, 2011). As noted, in order to display true competence, children must be able to produce *before* and *after* in each sentence structure (Ferreira, 1996).

A sentence repetition task is one way to provide experimental control over target sentence structures, so that the design can be analogous to that used in comprehension studies (Ambridge & Lieven, 2011). This has been used by Winskel (2003; also see Keller-Cohen, 1981) to elicit target sentence structures containing temporal connectives. Children aged 3 to 7 years were asked to repeat sentences containing one of eight temporal connectives. The main analysis by Winskel focused on a research question that is not directly relevant to this thesis (acquisition of connectives marking sequential versus simultaneous order) and it did not include reverse order sentences. However, one finding is of direct interest. Children made frequent substitution errors for both before-chronological (45% by 4-year-olds; 38% by 5-year-olds; 15% by 6-year-olds; 20% by 7-year-olds) and after-chronological (27% by 4-year-olds; 27% by 5-year-olds; 43% by 6-year-olds; 10% by 7-year-olds) sentences. This finding reinforces the view that children have difficulties producing two-clause sentences containing *before* and *after* well beyond the period that they first occur regularly in their speech. Further, it motivated my experimental work to use sentence repetition as a measure of children's ability to produce *before* and *after* in chronological and reverse order sentences.

An additional advantage with sentence repetition is that, even when few errors are made, latency times can be used to tap into online processing constraints associated with the respective conditions (e.g., Bannard & Matthews, 2008). However, the task is not a measure of spontaneous production and the child also has to comprehend and store the narrated sentence in memory prior to production. These extra requirements may reduce the sensitivity of this task as a measure of production. The only other paradigm for examining the production of specific sentence structures containing temporal connectives reported in published work, was used by Ye et al. (2011). Adults were presented with red and yellow cues that denoted which sentence structure was required. Whilst this method would be too demanding for young children, my experimental work drew on this concept to design blocked training conditions for each sentence structure (e.g., Huttenlocher, Vasilyeva, & Shimpi, 2004). Blocked conditions can be used to examine spontaneous production, whilst removing the additional comprehension and memory demands associated with sentence repetition. The major limitation with a blocked design is that the child may remember the rules from one training phase better than for others, creating a bias for a specific block condition (see Müller, Heller, & Ziegler, 1995; Reali, Spivey-Tyler, & Terranova, 2006). Therefore, the experimental work in the present thesis uses both sentence repetition and a blocked design to examine children's production of two-clause sentences containing *before* and *after* (see Chapter 4).

1.7. Overview of the research

1.7.1. Summary of literature review

Previous research shows that connectives appear in children's speech from a young age (Diessel, 2004), but that even in early adolescence, they are not fully understood (Cain & Nash, 2011; Pyykkönen & Järvikivi, 2012) or produced accurately (Peterson & McCabe, 1987). Further, work with adults indicates that sentences with temporal connectives are more difficult to process in some sentence structures than others (e.g., Ye et al., 2012). For language users to take full advantage of the information signalled by the connective, they must not only develop an initial understanding of the connective so as to no longer rely on a non-linguistic strategy to understand and represent order (Clark, 1971), but also must have the necessary sentence processing skills to make full use of the function of the connective in relation to the surrounding discourse (Cozijn et al., 2011; Traxler et al., 1997). This means that there is more than one reason for why performance may be inaccurate. First, young children are likely to make errors because they have a fragile understanding of the meaning of the connective. This is shown in comprehension tasks when children engage in the use of a non-linguistic strategy, rather than using the information provided by the connective (e.g., Clark, 1971). Second, when children no longer display a non-linguistic strategy, they may still have difficulties with sentences that carry additional structural demands on processing effort (Pyykkönen & Järvikivi, 2012).

Note that the influence of connective knowledge that is proposed by a non-linguistic strategy hypothesis (Clark, 1971) differs to that proposed by accounts of sentence processing (Just & Carpenter, 1992; Van Dyke et al., 2014). A non-linguistic strategy 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 having a fragile understanding for the meaning of the connective. Conversely, an account of sentence processing (Just & Carpenter, 1992; 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 initial understanding for the meaning of the connective, which is a later period of interest to the non-linguistic strategy hypothesis (Clark, 1971) 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) and a language-based account (Van Dyke et al., 2014).

Turning to production, it may be that children's ability to produce temporal connectives maps onto the developmental sequence that is predicted for comprehension. Whilst comprehension and production are related and draw on many of the same cognitive processes (Pickering & Garrod, 2013), we do not know how the additional planning demands of language production influence accurate use of sentences expressing different temporal orders of events. The problem with corpus and elicited production studies cited here is that they cannot experimentally manipulate whether the speaker uses the connectives in a specific sentence structure. The flexibility to produce these connectives in any given sentence structure without measuring competence in other sentence structures, may belie children's true competence (Ferreira, 1996). There are two paradigms used in the present thesis to elicit specific sentence structures: sentence repetition (e.g., Winskel, 2003) and blocked conditions in elicited production (e.g., Huttenlocher et al., 2004).

Taking comprehension and production together, the essential research question that my experimental work examines is why children continue to have difficulties processing two-clause sentences containing *before* and *after* in the immediate years that follow the emergence of their use in speech. There are a number of limitations in previous research. One standout issue is that, whilst some research attributes variation in performance to the cognitive effort required for specific sentence structures (e.g., Pyykkönen & Järvikivi, 2012), there is a need for research to disentangle the specific contributions of memory and language. One way to do this, is to acknowledge the experimental manipulations of these factors within the sentence structure: order of mention manipulates the memory processing demands of the sentence whereas the connective manipulates language knowledge. A fundamental question is whether the influence of these factors on performance is predicted by memory capacity or vocabulary, as the decay of information from working memory that is proposed by the memory capacity account (Just & Carpenter, 1992) may be driven by language knowledge (e.g., Van Dyke & Johns, 2012; Van Dyke & McElree, 2006; Van Dyke et al., 2014). No previous research has directly related performance on these sentences to an independent measure of working memory or vocabulary.

In addition, there are a number of paradigms that can now be used to gain a clearer insight into early competence without the additional demands imposed by the act-out tasks (e.g., Friend & Kelpinger, 2008). An insight into when children use connectives to guide their understanding of multi-clause sentences and the factors that limit this is fundamental to the development of theoretical models of listening comprehension, critical skills for educational success.

1.7.2. Objectives of the thesis

The first aim of the present research was to determine the age that children can use *before* and *after* to understand and produce two-clause sentences with a sequential temporal order of events (Clark, 1971). The second aim was to elucidate the reasons for why children continue to have difficulties processing two-clause sentences containing connectives in the immediate years that follow the emergence of an initial understanding of the function of the connective. The third aim was to identify whether the pattern of difficulty for different sentence structures was the same or different for comprehension versus production. In the light of theoretical advances and experimental methods, these aims were investigated using accuracy and timing measurements of comprehension and production.

Each experiment was designed to contrast two accounts of why working memory can influence variance in the processing of complex sentences: a memory capacity account (e.g., Just & Carpenter, 1992) and a language-based account (e.g., Van Dyke et al., 2014). The influence of memory capacity was investigated in two ways: by manipulating the structural demands of the sentences while holding vocabulary constant, and also by including an independent measure of memory in the analyses to better understand any effects. A memory capacity account (e.g., Just & Carpenter, 1992) would predict that before-chronological sentences should be performed most easily because the sentence structure has no additional load on working memory capacity (chronological order, medial position, earlier acquired connective), whereas the other structures each have two factors that increase the amount of information that must be held active in working memory: before-reverse (reverse order, initial position), after-chronological (initial position, later acquired connective), and after-reverse (reverse order, later acquired connective). Critically, the independent measure of working memory would be expected to explain unique variance in performance across sentence structures.

The role of language knowledge was also investigated in two ways: by manipulating connective difficulty (*before, after*), and also by including an independent measure of vocabulary to examine whether it drives performance effects. A language-based account (e.g., Van Dyke et al., 2014) would predict that language knowledge (i.e., connective, vocabulary) modulates how well children perform with sentence structures that require more cognitive effort. More specifically, children would be expected to have most difficulty for reverse order sentences linked by *after*, and the pattern of performance should be driven by vocabulary knowledge.

Experiment 1 (Chapter 2) investigated 3- to 7-year-olds' comprehension of temporal connectives using a forced-choice touch-screen paradigm, designed to simplify the demands of the task. Children aged between 3 and 7 years listened to two-clause sentences linked by a temporal connective, *before* or *after*, while viewing animations of the actions in each clause. In addition, to examine whether world knowledge supports early competence, temporal order was either predictable from world knowledge information in the sentence, or was not (*He brushed his teeth, before he went to bed*; vs. *He brushed his teeth, before he walked in the rain*). Following each sentence, children were asked to select the event that happened first to assess their understanding of the temporal order. The wide age range allowed the study to pinpoint the age at which children typically: (i) display a non-linguistic strategy when they do not yet possess robust understanding of these connectives, and (ii) display processing difficulties that emerge in the immediate years that follow an initial understanding for the connective.

In relation to the first point above, previous evidence suggests that young children who do not understand the meaning of the connective use a non-linguistic strategy to comprehend sentences containing connectives. However, no previous study has identified if there is preference for the two possible non-linguistic strategies that were investigated in the present thesis (order of mention, world knowledge). In relation to the second point above, the experiment built on recent research on adults' processing of sentences containing connectives, which suggests that the structural demands of some sentences require greater cognitive effort to process (Münte et al., 1998; Ye et al., 2012). On that basis, this experiment examined whether a memory-capacity constrained (Just & Carpenter, 1992) or a language-based (Van Dyke et al., 2014) account best predicted comprehension accuracy.

Experiment 2 (Chapter 3) used the same paradigm and materials as Experiment 1 (Chapter 2), but additionally assessed speed of response. Children were trained to make speeded responses to touch the thing that happened last. The use of a timed response measure, in addition to response accuracy, provided a sensitive means to assess whether different sentence structures differ in processing ease, as has been previously found for adults. It also provided an opportunity to replicate the findings of Experiment 1 (Chapter 2). Therefore this experiment provided an insight into how young children construct temporal representations of the meaning conveyed in spoken and written discourse.

Experiments 3 and 4 (Chapter 4) examined production. Children viewed an animated sequence of two actions, and were asked to describe the order of events. Instructions and practice trials were used to model the target sentence structures. Accuracy and response latency were recorded. This work aimed to investigate what factors influence children's production of two-clause sentences linked by *before* or *after*, specifically whether any difficulties are best attributed to low working memory capacity or to weak language skills.

Together the four experiments provided several forms of measurements to test the predictions of two accounts of the role of working memory on the processing of complex sentences containing temporal connectives. Online and offline measures were used to examine comprehension and production. The focus of the Discussion (Chapter 5) concerns which findings converge across the studies and which do not, with particular interest in whether performance converges for comprehension versus production.

2. Young children's comprehension of temporal relations in complex sentences: the influence of memory on performance

Text as it appears in: 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.

CHAPTER 2: TEMPORAL RELATIONS IN CHILDREN'S COMPREHENSION.

Abstract

We investigated 3- to 7-year-olds' (N=91) comprehension of two-clause sentences containing the temporal connectives *before* or *after*. The youngest children used an order of mention strategy to interpret the relation between clauses: they were more accurate when the presentation order matched the chronological order of events: '*He ate his lunch, before he played in the garden*' (chronological) versus '*Before he played in the garden, he ate his lunch'* (reverse). Between 4 to 6 years, performance was influenced by a combination of factors that influenced memory processing load: connective type and presentation order. An independent measure of working memory was predictive of performance. We conclude that the memory demands of some sentence structures limits young children's comprehension of sentences containing temporal connectives.

Keywords: temporal connectives, listening comprehension, incremental processing, memory, language acquisition.

CHAPTER 2: TEMPORAL RELATIONS IN CHILDREN'S COMPREHENSION.

2.1. Introduction

Successful comprehension results in an integrated and coherent mental representation of the state of affairs described in a text, rather than a verbatim record of the specific words or syntactic structures (Johnson-Laird, 1983; van Dijk & Kintsch, 1983). Critically, adult readers and listeners encode the relations between events on several dimensions, including temporality, the order in which events occur (Gennari, 2004; Zwaan, 2008; Zwaan & Radvansky, 1998). Temporal connectives such as *before* and *after* are one source of linguistic information that specifies the order of events and, as a result, they aid the comprehension of two-clause sentences and the construction of an accurate and coherent meaning-based representation (Costermans & Fayol, 1997). Although, temporal connectives are produced in children's speech from around 3 years of age (Diessel, 2004), children have difficulty on tasks designed to assess the comprehension of these connectives up to at least 12 years of age (Pyykkönen & Järvikivi, 2012). That is, young school-aged children produce temporal connectives before they can comprehend them in spoken language.

In this research, I focus on the development of comprehension of sentences containing the temporal connectives *before* and *after* in 3- to 7-year-olds. Our findings indicate the age at which competence emerges in the use of connectives, and how this is related to different sentence structures. Our observations advance understanding of the development of competence in temporal connectives by revealing the influence of memory skills in the improvements in performance evident during early childhood.

When children do not understand a temporal connective, they can use different strategies to understand and represent the relation between two events in a two-clause sentence containing a temporal connective, rather than using the precise linguistic information provided by the connective itself (Clark, 1971). Two strategies that we consider are a world knowledge strategy and an order of mention strategy. World knowledge may support correct interpretation of event order when the events typically occur in a set order, for example, '*She put on her boots, after she put on her socks*.' The order of events in such sentences can be understood without using the information provided by the connective. When there is no typical order for two events, as in *'She put on her hat, after she put on her scarf',* language comprehenders can only interpret the order correctly if they understand the relation signalled by the connective. Between 3 to 5 years of age, children appear to rely on world knowledge, rather than knowledge of the connective: they are better at comprehending the sequence of events expressed in sentences when the sequences are typical, and thus supported by world knowledge, compared to when event order is arbitrary (e.g., French & Brown, 1977; Keller-Cohen, 1987).

Children may also construct a correct interpretation of the sequence of events expressed in a sentence by assuming that the event sequence corresponds to the order in which the events were mentioned: an order of mention strategy (Clark, 1971). If young children are using this strategy, they should find it easier to comprehend sentences in which the order of mention corresponds to the order of events, as in chronologically ordered sentences such as *'She put on her hat, before she put on her scarf'* compared to reverse order sentences such as *'She put on her scarf, after she put on her hat.'* An order of mention strategy will result in an incorrect interpretation of event order in the latter. Between the ages of 3 to 5 years, children perform more accurately on chronological sentences than on reverse order sentences (Clark, 1971; French & Brown, 1977; Johnson, 1975). This finding indicates that young children employ an order of mention strategy to comprehend the temporal order of events in multiple clause sentences. Thus, children can resort to two strategies, world knowledge or order of mention, to respond appropriately to connectives without fully understanding them.

These studies inform us that 3- to 5-year-old children do not have full understanding of the meaning of *before* and *after* and provide us with an insight into the strategies that young children might use to process complex sentences that include a temporal connective. However, as mentioned earlier, even 12-year-olds do not perform at adult levels in studies designed to assess the comprehension of sentences containing temporal connectives (Pyykkönen & Järvikivi, 2012). The question we ask is this: What factors drive the comprehension of complex sentences containing temporal connectives once children have developed an appreciation for the meaning of *before* and *after*?

The extant literature suggests that three key factors may influence the comprehension of sentences that include connectives: the relative familiarity of the connective in terms of its frequency of occurrence in a child's linguistic experience; the relation between the order of mention of the connective and the order of events being described by the connective; and the position of the connective in a sentence. Each effect can be explained in relation to the impact of variation in the demands on processing capacity imposed by sentences including connectives. Developmental improvements would be predicted by capacity theories of comprehension which propose that comprehenders with low working memory capacity are less likely to retain a full and accurate representation of a sentence during comprehension, particularly when that sentence carries high memory demands (e.g., Just & Carpenter, 1992).

To establish the motivation for our study, we review relevant research. One factor that should be expected to affect comprehension performance is the relative familiarity of different temporal connectives according to the language experience of the child. Clark (1971) found earlier competence for *before* than for *after* in 3- to 5-year-olds. She attributed this difference in age of acquisition to the semantic features of each term: *before* indicates the prior event, whereas *after* does not, making the latter more semantically complex. Another reason for earlier competence for *before* relative to *after* is differential exposure to these temporal terms. As is evident in large language corpora such as the British National Corpus (Leech, Rayson, & Wilson, 2001), *after* occurs more often than *before* as a preposition or adverb, as in '*The dog chased after the ball*', in addition to its use as a temporal connective. As a result, it may be more difficult for children to activate their knowledge of *after* as a temporal connective compared to *before*.

Another factor that may influence performance is the relation between the order of mention of the connective and the order of events being described by the connective. As noted, children who do not understand the semantics of a temporal connective are more likely to be accurate at comprehending sentences in which the order of mention of events is congruent with the chronological order of occurrence of the events (e.g. Clark, 1971). Importantly, once a competent understanding of the connective itself emerges, a processing difficulty for reverse order sentences may persist as a function of high demands on working memory (Ye, Kutas, St. George, Sereno, Ling, & Münte, 2012).

It has long been known that the mental representation of a two-clause sentence encodes its meaning, not specific words or syntactic structures (Bransford, Barclay & Franks, 1972). For a chronological order sentence, information about the sequence of events specified in two clauses linked by a connective can be assimilated into a congruent meaning representation for the sentence incrementally, as the events are mentioned. In contrast, the comprehender cannot incrementally construct a correct interpretation of the sequence of events for a reverse order sentence such as '*Before she put on her scarf, she put on her hat'*, but must wait until the second clause is presented. The greater demands on memory imposed by sentences in this account, can be expected to cause comprehension problems for young children. Consistent with this prediction, even adults find sentences with an initiating connective harder to process when the events are presented in reverse order (Münte, Schiltz, & Kutas, 1998; Ye et al., 2012).

The position of the connective in the sentence was not a factor directly manipulated in our study, but we consider it here because it will vary as a function of the connective (*before* vs. *after*) and manipulation of order (chronological vs. reverse). Temporal connectives can appear in either a sentence medial position, as in '*She put on her hat, before she put on her scarf.*' or a sentence initial position, as in '*Before she put on her scarf, she put on her hat*.' In an analysis of children's natural language production, Diessel (2004) found a strong preference for the sentence medial position for temporal connectives in the productions of children aged between 2 and 5 years (see also Diessel, 2008, for similar work with adults). This preference can be explained by noting that if a connective occurs in a sentence medial position, incremental word-by-word processing of the sentence meaning is afforded, but that when a connective occurs in the sentence initial position, the comprehender (or producer) cannot simply process (or plan) the sentence word-by-word. Thus, the position of the connective in the sentence may influence comprehension through the variation in working memory demands that arise through sentence position. When processing sentences that contain connectives in the sentence initial position, the comprehender must maintain the information provided by the connective in memory while processing the event of the first clause, and then use the stored connective information to link the event specified in the first clause correctly with the event specified in the second clause.

When processing a sentence medial connective, the information required to link events specified in the first and second clauses will be available roughly when it is required, reducing the period during which the content of the first clause must be maintained in working memory prior to linkage with the second clause. The assumption is therefore that connectives in the medial position are preferred because they can be processed accurately while making fewer demands on memory. Consistent with this account, studies of older children and adults have indicated the general use of an incremental processing strategy for sentences joined medially by connectives (Cain & Nash, 2011; Traxler, Bybee, & Pickering, 1997). For young children, who have low working memory capacity, a connective appearing in the sentence initial position may therefore be harder to comprehend.

Only one study to date speaks to these three factors in relation to children's (and adults') mental representation. This study by Pyykkönen and Järvikivi (2012) found that, for 8- to 12-year-olds, chronologically ordered sentences that could be processed incrementally (before-chronological) were easier to comprehend than reverse order sentences that also had a connective in the medial position, but which could not be processed incrementally (after-reverse). Sentences in which the connective appeared in the initial position (before-reverse and after-chronological) were of similar and intermediate difficulty for the children, whereas adults performed at ceiling on all sentence types. Pyykkönen and Järvikivi's study clearly demonstrates the need to consider that differences in sentence position, which will arise through the manipulation of connective and order, might influence the comprehension of sentences with temporal connectives. However, Pyykkönen and Järvikivi's task allowed re-reading and reflection on the sentence. For that reason, their findings cannot be interpreted directly in terms of the differing processing demands imposed by sentences with different structures involving temporal connectives. We set out to advance understanding of young children's comprehension of connectives by considering the impact of order, connective type, and position, by using a task that promoted response types that would allow interpretation of effects in terms of demands on working memory.

The present study

Previous research has identified the strategies that very young children might use to process two-clause sentences containing temporal connectives, but has not investigated why these sentences remain hard for children to process for several years after they appear in their spoken language productions. We compared consecutive age groups between 3 to 7 years of age to pinpoint the moment of developmental change. Our aim was to determine when children shift from using strategies such as order of mention or world knowledge to comprehend the chronological order of events in sentences that contain temporal connectives, to using the connective itself as a linguistic device that signals order. Further, we aimed to elucidate the reasons why these sentences are often misunderstood even after children appreciate the different orders signalled by *before* and *after.* We compared comprehension of two-clause sentences joined by *before* and *after* and manipulated whether the event sequence was presented in chronological or reverse order. In this way, position of connective varied as a function of these two factors. Thus, the design included the following sentence types: before-chronological order (medial position); before-reverse order (initial position); after-chronological order (initial position); and after-reverse order (medial position). We also manipulated whether the events in the two clauses typically occurred in a set order (world knowledge present) or not (world knowledge absent). The manipulation of world knowledge in conjunction with these other factors allowed us to identify whether children used an order of mention strategy or relied on world knowledge when they did not possess robust working knowledge of the connective.

Our interest in the language processing demands posed by connectives led us to select a task that had low cognitive performance demands. The majority of previous studies examining young children's comprehension of temporal connectives have used an act-out task, which has high cognitive demands (Ambridge & Rowland, 2013). Here, to capture early competence and to minimise the processing demands, we assessed comprehension with a simple forced-choice task. Children listened to a two-clause sentence in which the order of two events was signalled by a connective (*before* or *after)* while viewing an image of each clause on a touch-screen monitor. After each sentence, they selected which of the two events happened first. The use of images depicting the events in sentence stimuli reduces memory load (e.g., Vion & Colas, 2005). Previous successful use of touch-screen technology for capturing early comprehension competence has been reported with children as young as 18 months (Friend & Keplinger, 2008).

Knowledge of *before* should be acquired earlier than *after* according to both the semantic complexity and frequency of exposure accounts. Therefore, in general, *before* sentences should elicit a greater number of accurate responses than *after* sentences. We hypothesised that the youngest children's pattern of performance would indicate that they did not have robust knowledge of the temporal relation signalled by the connective (in line with the previous research detailed above) and would rely on a strategy, using either order of mention or world knowledge to comprehend sentences. Previous research has not identified a preference for either strategy, so we did not make specific predictions on this point. We predicted that the older children would generally perform above chance on both connectives, because they had more secure knowledge of the specific meaning of the connectives.

However, the previous literature discussed earlier motivated us to predict that older children's performance would be affected by the processing demands of different sentence types (e.g., Pyykkönen & Järvikivi, 2012; Ye et al., 2012). Taken together, this literature identifies three key factors that vary the grammatical structure of sentences including connectives, namely: connective type, the order of events, and connective position. This variation may also impact the demands on processing capacity. In line with a memory capacity theory of comprehension, we expected that children would perform worse on sentences that inflict high demands on working memory during clause integration (e.g., Just & Carpenter, 1992). For example, before-chronological sentences such as *'She put on her hat, before she put on her scarf'* were expected to elicit the most accurate level of performance because the order (chronological) and connective position (medial) combined to allow word-by-word incremental processing. In comparison, before-reverse (initial position, reverse order), after-chronological (initial position, later acquired connective), and after-reverse (reverse order, later acquired connective) sentences would elicit less accurate performance because such sentences each carry two features that add to the amount of information that must be held in working memory.

Given the potential explanation of performance patterns in terms of processing load, we included an independent measure of memory in our analysis of comprehension performance to examine if the influence of sentence structure on comprehension would be modulated by children's memory capacities. We predicted that memory would be a significant predictor of performance, in general.

2.2. Method

Participants

Ninety-one children aged 3–7 years participated in the study. All were native English speakers from schools and preschools that served mixed socio-economic catchment areas in the North West region of England. No children had reported language disabilities. Children were in four different school year groups: 22 3- to 4-year-olds (aged 3;3-4;4, 13 boys), 21 4- to 5-year-olds (aged 4;5-5;5, 14 boys), 24 5- to 6-year-olds (aged 5;5 to 6;5, 11 boys), and 24 6- to 7-year-olds (aged 6;5 to 7;4, 13 boys). A further two 3- to 4-year-olds participated but were excluded from the analysis because of either refusing to touch the screen (i.e., select an answer; N = 1) or being unresponsive after corrective feedback (N = 1). Data collection took place between January and July 2013. Written parental consent was obtained for all children, and children provided oral assent before each session. All children had age appropriate receptive language assessed using the British Picture Vocabulary Scales – III (Dunn, Dunn, Styles, & Sewell, 2009). Full details are reported below.

Materials and Procedure

All children completed three assessments: a connective comprehension task, a measure of memory, and a measure of receptive vocabulary. The connectives task was administered over two separate sessions. Each session lasted no longer than twenty 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 paradigm. Sixty-four two-clause sequences were constructed, each representing two events that were related by world knowledge. All sequences referred to one actor and two objects. Each of the 64 items was counterbalanced into one of eight lists. In each list there were 32 sentences assessing eight conditions (shown in Table 2.1) that resulted from three manipulated factors: presence or absence of world knowledge to support the relation between the two events; the temporal connective (*before* vs. *after*); and the presentation order of events (chronological vs. reverse). The manipulations of connective type and order of events in turn resulted in sentences in which for both *before* and *after* the connective could appear in either initial or medial position. Thus, for *before* sentences, the connective appeared in the medial position (as shown in Table 2.1) when events were presented in chronological order, and in the sentence initial position when the events were presented in reverse order. The reverse was true for sentences containing *after.*

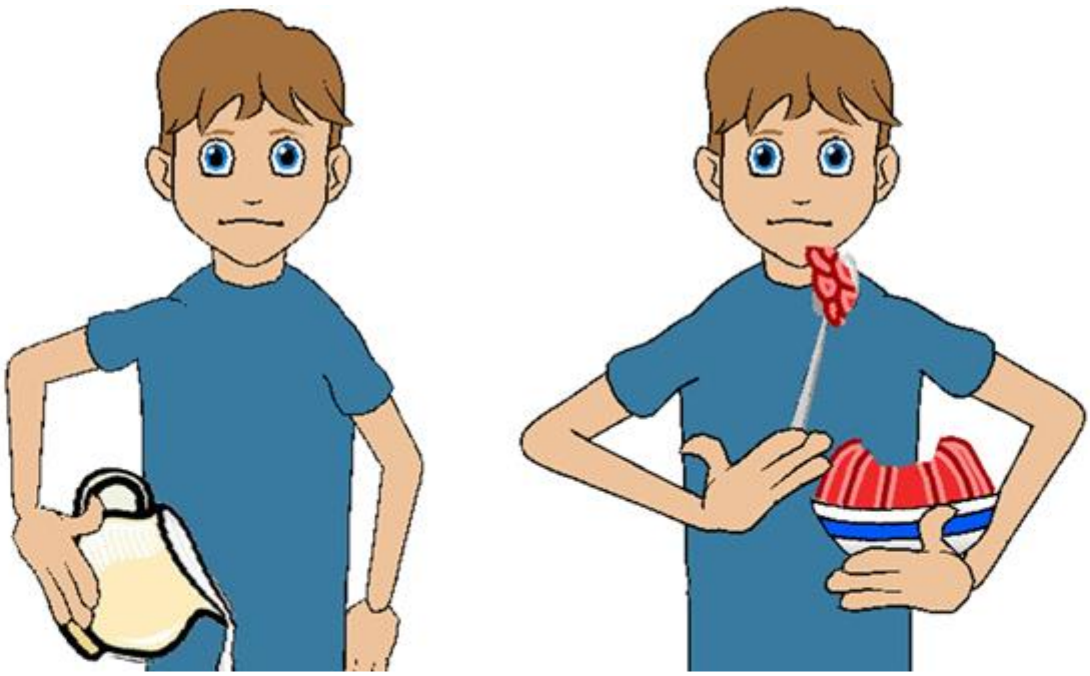
Table 2.1.

Sentence conditions

	Before		After	
	Chronological	Reverse	Chronological	Reverse
World knowledge present	He poured the ketchup, before he ate the burger	Before he ate the burger, he poured the ketchup	After he poured the ketchup, he ate the burger	He ate the burger, after he poured the ketchup
World knowledge absent	He put on the sandals, before he ate the burger	Before he ate the burger, he put on the sandals	After he put on the sandals, he ate the burger	He ate the burger, after he put on the sandals

For each clause, an animated cartoon was created using Anime Studio Pro 9.1 (Smith Micro Software, 2012). Each cartoon depicted the actor, action and object of a clause (e.g., Tom pouring a ketchup bottle; Tom eating a hotdog). Each animated segment lasted for three seconds and explicitly encapsulated only the object (e.g., hotdog) from one clause, whilst the object (e.g., ketchup) from the other clause was not present. Each animation ended with a freeze-frame judged by the first and third authors to best represent the action of that clause. An example freeze frame is provided in Figure 2.1. Each visual stimulus (e.g., Tom eating a hotdog) was 486 pixels in height and did not exceed the left or right half of the presentation (486 x 872 pixels).

Figure 2.1. Example presentation of an animation freeze-frame (cream, jelly).



Children first saw the two animations, shown sequentially. The animation on the right hand side of the screen was shown first, followed by the animation on the left hand side of the screen. Children were instructed to: '*Listen carefully and touch the thing Tom/Sue did first*' (name selected to match gender of child) and the narration of the sentence was played (over headphones). A response window was opened with a short beep and was closed by a blank screen once the child had responded. Both order of appearance and side of presentation for the visual representations of the target and non-target clauses were counterbalanced across trials.

The experiment was run using the PsyScript 3.2.1 (Slavin, 2013) scripting environment on a Macintosh laptop connected to a touch-screen monitor with items presented in a random order. Correct responses were recorded as those items for which the child touched the target action that was the first event to occur in the sentence.

Children practiced all four sentence types in which world knowledge was absent (see Table 1.2). Example items were not used in the test phase. Therefore, each child completed a minimum of four practice trials to ensure that they understood the procedure before the test phase. Children were excluded from the study if they were uncooperative in selecting an answer for each of the training items (i.e. they did not touch the screen on any trial), or if they were unresponsive after receiving corrective feedback. Practice trial instructions emphasised the importance of making judgements based solely on the narrated sentence, not the visual stimuli. Paired sample t-tests revealed no significant effect of order or side of presentation on accuracy (for all comparisons, $p > .1$; data was reflected and log transformed for the two oldest age groups because their data was not normally distributed).

Vocabulary. Receptive vocabulary was assessed using the British Picture Vocabulary Scales – III (Dunn et al., 2009) to ensure that the sample had age-appropriate vocabulary skills. In this task, the child has to point to one of four pictures that best illustrates the meaning of a word spoken aloud by the researcher. Testing is discontinued when a specified number of errors have been made. 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=109.27 (10.37); 4- to 5-year-olds =111.76 (6.16); 5- to 6-year-olds

=105.83 (8.67); 6- to 7-year-olds =103.88 (8.02). Thus, no children were excluded for weak receptive language skills.

Memory. Each child completed the digit span task from the Working Memory Battery for Children (Pickering & Gathercole, 2001) to assess memory. In this task, the child hears a string of digits, read out by the researcher, and is then asked to recall the digits in the same order. The easiest level comprises strings of two digits, and the number of items in the string is increased until the child cannot recall all of the digits after three successive attempts. 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). Raw scores were used for the analysis. The test-retest reliability reported in the manual for children aged 5-7 years is $r = .81$.

Design

A 4 x 2 x 2 x 2 mixed design was used. The between-subjects independent variable was year group (3-4, 4-5, 5-6, and 6-7 years) and the within-subjects variables were world knowledge (present, absent), connective (*before*, *after*), and order (chronological, reverse order). The dependent variable was accuracy.

2.3. Results

A total of 5824 responses were recorded. Before analysis, data were screened to remove potential distortions from the norm. Three children from the oldest age group were removed (192 responses, 3.3%) because they were identified in by-age box plots as outliers who were performing at floor level in accuracy. Therefore, 5632 responses were included. The removal of these participants did not affect the pattern of the reported results.

Analysis strategy

A series of Generalised Linear Mixed-effects models (GLMMs) (Baayen, Davidson, & Bates, 2008; Barr, Levy, Scheepers & Tilly, 2013) were fitted to the data in the R statistics environment (R Core Team, 2014) using the lme4 package (Bates, Maechler & Bolker, 2014). This method is essentially an extension of logistic regression, such that a GLMM analysis estimates the fixed effects due to experimentally manipulated variables while taking into account random error variance due to differences between participants or between stimulus items sampled for the study. We followed the recommendations of Barr et al. (2013) by estimating fixed effects in models that included random effects terms corresponding to both random differences between participants or items in overall accuracy of responses elicited (random intercepts) and random differences between participants or items in the slopes of the effects of world knowledge, connective and order condition. As a maximal random effects model did not converge, we used the likelihood ratio test (Barr et al., 2013; Pinheiro & Bates, 2000) to test whether the inclusion of fixed or random effects was warranted by superior model fit to data. That is, we added as many slopes as were found to be warranted. Each of the final models incorporated random intercepts for both participants and item effects, and by-participant random slopes for both connective and order effects.

The raw memory scores [mean (+/- SD)] demonstrated age-related improvements: 3- to 4-year-olds = 15.36 (3.50); 4- to 5-year-olds = 19.67 (2.94); 5- to 6-year-olds = 22.67 (4.02); 6- to 7-year-olds = 25.42 (4.48). In addition, the standardised scores of memory were within the normal range of 85-115 for each age group: 4- to 5-year-olds = 91.90 (10.35); 5- to 6-year-olds = 97.75 (15.15); and 6- to 7-year-olds = 100.96 (20.42). Standardised scores are not provided for 3- to 4-year-olds.

In the following, we first describe the optimum model for the full dataset, with age, order, and connective entered as fixed effects (Model 1, Table 2.2). We then further examined the significant interaction between age, connective, and order, found in the full dataset model, by conducting simple interaction analyses of the effects of connective and order for each group separately. Table 2.3 presents the analysis with different age groups to determine their use of comprehension strategies. Finally, in Table 2.4 (Model 2), we returned to our analysis of the full dataset to examine whether a model with memory included as a fixed effect fitted the data better than a model without. In each analysis, world knowledge had no significant main effects nor any significant interactions (all *p*s > .14), and the fit of the model was improved upon its removal, $\chi2(8)= 22.53$, $p < .01$. Therefore, following recommendations for obtaining an optimal model by Barr et al. (2013), the effect of world knowledge is not included in the models that we present.

The inferential statistics for each model are presented in Tables 2.2, 2.3, and 2.4 respectively. These summarise the main effects and interactions of age, order, and connective. The first column provides the coefficient estimates of effects (*b*) due to experimental conditions, which is the change in the log odds accuracy of response associated with each fixed effect. A positive coefficient indicates that the effect of differences between conditions was to increase the odds that a response would be correct while a negative coefficient indicates that the effect of a factor was to decrease the odds that a response would be correct.

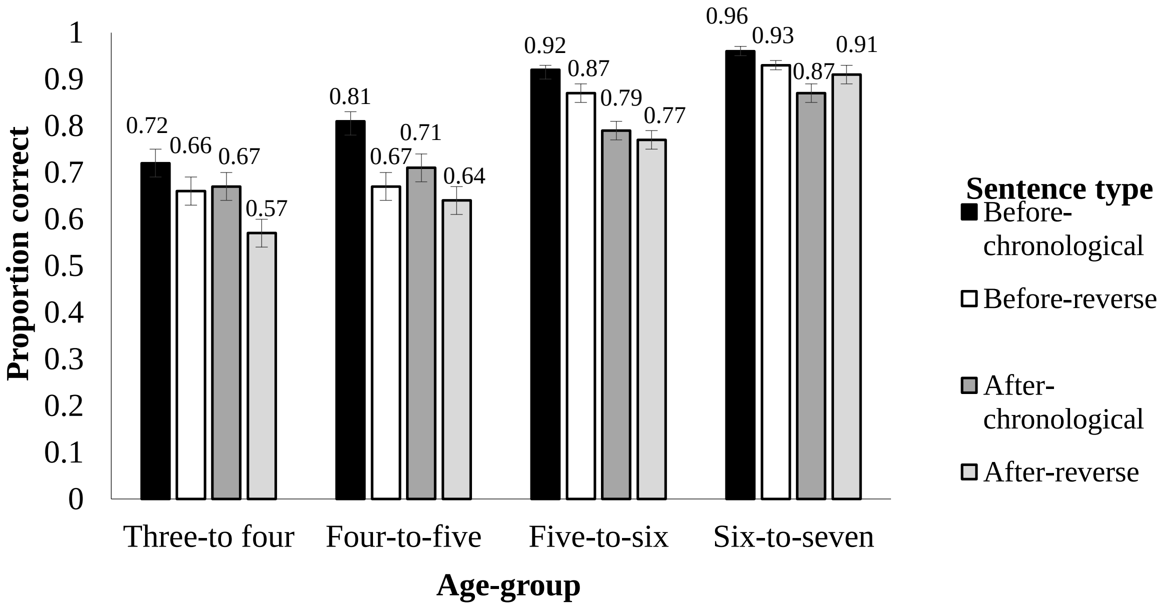
Main analysis

Model 1 (Table 2.2) shows that accuracy of response was significantly affected by participant age, indicating a developmental improvement in accuracy from three to seven years. In general, chronological sentences were comprehended as well as reverse order sentences. Similarly, there was no difference between accuracy for *before* and *after* sentences. Order and connective effects did not interact with each other, or by age. There was a significant three-way interaction between age, order, and connective. Figure 2.2 shows the mean accuracy scores for all observations to each experimental condition (collapsed over world knowledge).

Given the significant interaction between the effects of age, order, and connective conditions, we conducted simple interaction analyses to examine the effects of order and connective on the responses for each age group considered separately. These are reported next and summarised in Table 2.3.

Analyses of individual age groups. A main effect of order, only, was found in the analysis of the 3- to 4-year-olds' data, because this youngest age group comprehended chronological sentences more accurately than reverse order sentences. There was no main effect of connective for the two youngest age groups, indicating that they comprehended *before* and *after* equally poorly. In contrast, there was a main effect of connective in the analysis of the data from the 5- to 6-year-olds and a similar, but non-significant, effect in the analysis of the 6- to 7-year-olds' data. The interaction between order and connective was not significant for the youngest age group. In contrast, this interaction was significant for each of the three oldest age groups.

Figure 2.2. Mean proportion correct (with standard error bars) for each experimental condition by age group.



The order by connective interaction for each of the three older age groups was explored further by examining performance for the *before* and *after* items separately (see also Figure 2.2). The two middle age groups displayed a main effect of order with *before* items, but not with *after* items. That is, before-chronological sentences were comprehended better than before-reverse sentences, whereas after-chronological sentences were comprehended as well as after-reverse sentences. Therefore, 4- to 6-year-olds displayed a significant preference for *before* sentences that were presented in chronological order. These effects were not significant for the oldest age group, for whom performance was much higher in general.

Table 2.2.

Summary of GLMM for the log-odds of accuracy responses: Effects and interactions of age, order and connective.

Fixed Effects	**Estimated Coefficient (*b*)**	**SE**	**t**	**p(>|z|)**
Intercept	-2.42	0.66	-3.67	0.01
Age	**0.06**	**0.01**	**5.62**	**< 0.01**
Order	1.11	0.61	1.81	0.07
Connective	-0.63	0.73	-0.86	0.39
Order:Connective	-1.16	0.69	-1.67	0.09
Age:Order	-0.01	0.01	-1.47	0.14
Age:Connective	0.02	0.01	1.44	0.15
Age:Order:Connective	**0.03**	**0.01**	**2.19**	**0.03**

Note (1) Fixed effects labels: Age = effect of age (in months); Order = effect of order, chronological (reference level) vs. reverse order; Connective = effect of connective, before (reference level) vs. after. (2) Interactions are reported as colons. (3) Bold = predictor is significant at $p<.05$ or better. (4) See Table A.7.1 for zero order correlations (appendix).

Table 2.3.

Summary of GLMMs (per age group) for the log-odds of accuracy responses: Effects and interactions of order and connective.

	Age 3-4				Age 4-5				Age 5-6				Age 6-7			
	(*b*)	SE	z	*p*	(*b*)	SE	z	*p*	(*b*)	SE	z	*p*	(*b*)	SE	z	*p*
Full models																
(Intercept)	0.17	0.22	0.76	0.44	0.78	0.33	2.39	0.02	1.52	0.33	4.65	<0.01	3.18	0.37	8.64	<0.01
Order	**0.56**	**0.26**	**2.18**	**0.03**	0.36	0.37	0.97	0.33	0.43	0.29	1.46	0.15	-0.67	0.37	-1.78	0.08
Connective	0.28	0.19	1.45	0.15	0.10	0.39	0.26	0.80	**0.73**	**0.34**	**2.18**	**0.03**	0.70	0.40	1.74	0.08
Order: Connective	-0.41	0.28	-1.45	0.15	**0.81**	**0.36**	**2.27**	**0.02**	**2.95**	**0.81**	**3.66**	**<0.01**	**2.16**	**0.89**	**2.44**	**<0.01**
Simple effects models																
Before models																
(Intercept)	–	–	–	–	0.86	0.33	2.65	0.01	2.20	0.25	8.68	<0.01	3.79	0.49	7.81	<0.01
Order	–	–	–	–	**1.11**	**0.30**	**3.72**	**<0.01**	**3.75**	**0.77**	**4.88**	**<0.01**	1.69	0.91	1.85	0.06
After models																
(Intercept)	–	–	–	–	0.94	0.39	2.43	0.02	1.54	0.33	4.65	<0.01	2.87	0.33	8.62	<0.01
Order	–	–	–	–	0.17	0.38	0.45	0.65	0.25	0.21	1.18	0.24	-0.50	0.28	-1.82	0.07

Note (1) Fixed effects labels: Age = effect of age (in months); Order = effect of order, chronological (reference level) vs. reverse order; Connective = effect of connective, before (reference level) vs. after. (2) Interactions are reported as colons. (3) Bold = predictor is significant at $p < .05$ or better.

Model 2 (Table 2.4) indicates that when working memory is incorporated in the model of the full dataset, it predicts performance over and above age. Indeed, this model was a significantly better fit to the data than when the same model was run without working memory, $\chi^2(4) = 13.93$, $p < .01$. There was a significant two-way interaction between order and connective. Of particular note, the three-way interaction between age, order and connective was no longer significant but, instead, the three-way interaction between working memory, order, and connective neared significance ($p = .06$). This latter finding provided converging evidence that age effects were partly a proxy for memory, a conclusion corroborated by the strong correlation between these two variables ($r = .71$).

Of course, an alternative explanation of any working memory effects in developmental studies is variation in long-term knowledge of language (e.g., Kidd, 2013; Klem, Melby-Lervåg, Hagtvet, Lyster, Gustafsson, Hulme, 2015). To test this account, Model 3 tested the same three factors as Model 2 (age; order of events; connective) but with the receptive vocabulary scores included instead of performance on the assessment of memory. This model (Model 3) was not a significantly better fit to the data than Model 1 $\chi^2(4) = 8.82$, $p < .07$. A final model (Model 4) was also tested that included the same factors as Model 1 and with both memory and vocabulary included. Whilst this model was a significantly better fit than Model 1 ($\chi^2(8) = 20.02$, $p = .01$), it did not significantly improve the fit compared to Model 2 which included just working memory ($\chi^2(4) = 6.07$, $p = .19$). Together, these comparisons between models indicate that working memory, not vocabulary, is driving performance on our sentence comprehension task. For these reasons, we do not include the output for either of the models that incorporated vocabulary (Models 3 and 4: see appendix for this information).

Table 2.4.

Summary of GLMM for the log-odds of accuracy responses: Effects and interactions of memory, age, order and connective.

Main	**M (*b*)**	**SE**	**t**	**p(>|z|)**
(Intercept)	-2.77	0.66	-4.23	<0.01
Age	**0.03**	**0.01**	**2.10**	**0.04**
Memory	**0.10**	**0.04**	**2.68**	**<0.01**
Order	1.11	0.63	1.77	0.08
Connective	-0.42	0.74	-0.57	0.57
Order:Connective	**-1.44**	**0.71**	**-2.04**	**0.04**
Age:Order	-0.01	0.01	-0.98	0.33
Age:Connective	**0.03**	**0.02**	**2.09**	**0.04**
Memory:Order	<0.01	0.04	-0.08	0.94
Memory:Connective	-0.07	0.04	-1.46	0.14
Age:Order:Connective	<0.01	0.02	0.13	0.89
Memory:Order:Connective	**0.09**	**0.05**	**1.86**	**0.06**

Note (1) Fixed effects labels: Age = effect of age (in Age); Order = effect of order, chronological (reference level) vs. reverse order; Connective = effect of connective, before (reference level) vs. after. (2) Interactions are reported as colons. (3) Bold = predictor is significant at $p < .05$ or better. (4) See Table A.7.2 for zero order correlations (appendix).

2.4. Discussion

The aim of the present study was to identify the age at which children accurately use *before* and *after* to understand the temporal relation between two events in a sentence and to elucidate reasons for why sentences containing these connectives can be hard to process. Our findings extend understanding of young children's connective competence in several important ways. First, we show that at around 3 to 4 years of age, children perform above chance on these connectives, indicating that they have a basic understanding of these connectives, earlier than reported in previous research. Second, we find that between 6 to 7 years of age children are performing at high levels of accuracy, demonstrating the emergence of full competence earlier than has been reported previously (Cain & Nash, 2011; Pyykkönen & Järvikivi, 2012). Third, and critically, we demonstrate that children's competence is substantially modulated by variation in sentence structure and that an independent measure of working memory influences success. These findings indicate that both cognitive and language demands influence the emergence of young children's comprehension of temporal connectives.

The 3- to 4-year-old children performed above chance but demonstrated fragile understanding of the meanings conveyed by *before* and *after*. We hypothesised that the youngest children's pattern of performance would indicate that they did not have robust knowledge of the temporal relation signalled by the connective (in line with previous research detailed above) and would rely on a strategy of either order of mention or world knowledge to comprehend sentences. In line with this prediction, the pattern of performance for the youngest age group revealed that they relied on an order of mention strategy: the youngest age group were more accurate on chronological order than on reverse order sentences, consistent with some previous research (Clark, 1971; French & Brown, 1977; Johnson, 1975).

We can turn to adult literature to understand better the mechanisms underlying this developmental change. In the world around us, we experience events in a chronological order. Therefore, even adult comprehenders appear to expect that language will map onto that experience, displaying processing difficulties when such mapping is violated (Ye et al., 2012; Zwaan & Radvansky, 1998). Therefore, it is likely that children are mapping the events based on their assumption that language maps onto order, rather than focusing on the linguistic information of the connective itself.

In contrast to some previous studies (French & Brown, 1977; Keller-Cohen, 1987), there was no evidence that the children relied on world knowledge to process these sentences:

within each age group, children performed comparably whether the order of the two events was supported by world knowledge or not. We believe that our finding for reliance on order of mention rather than world knowledge is robust because the large participant sample size and use of a linguistic sample considerably larger than the norm for this area of research vouchsafed a fair opportunity to observe a world knowledge effect if one were to be found. In addition and critically, our use of a task that minimised cognitive load in producing responses (compared to act out tasks, e.g., French & Brown, 1977; Keller-Cohen, 1987) was designed to ensure maximum sensitivity.

Our findings show that the comprehension performance of 4- to 6-year-olds was governed by the varying demands on working memory by different sentence structures (Diessel, 2004; Pyykkönen & Järvikivi, 2012; Ye et al., 2012). Children were more successful in understanding the sequence of events if those events were presented in sentences in chronologically ordered clauses rather than in a reverse order, but this order effect was apparent for sentences containing the *before* connective not for sentences containing *after*. This interaction cannot be attributed to a lack of understanding of the connective *before*. Given that it is acquired earlier than *after* (Clark, 1971), it cannot be the case that understanding a *before* sentence is more susceptible to the effect of order because *before* is not understood as well.

As noted in the Introduction, higher demands on working memory have been associated with when the sentence elements are presented in reverse order (Ye et al., 2012), when the connective is later acquired (*after*; Clark, 1971; Leech et al., 2001), and when the connective is in the sentence initial position (Diessel, 2004), The difference due to order is revealed only for *before* sentences because the before-chronological sentences make the lowest demands on memory. They remain easier to comprehend than before-reverse (initial position, reverse order) sentences consistently throughout the age range in our sample of children. Like the before-reverse sentences, the after-chronological (initial position, later acquired connective) and after-reverse (reverse order, later acquired connective) sentences each possess two features taxing children's working memory capacity. We note that previous studies that have reported adult processing difficulties for reverse order sentences have only included order as a factor (Münte et al., 1998; Ye et al., 2012). That is, connective position was held constant and connective type was a confounding variable. Our interest in developmental acquisition motivated us to include connective type in addition to order as a statistical factor, which in turn enabled us to discuss how different connective positions may have influenced findings.

Our findings add to a growing body of research that has reported age-related differences in children's understanding of temporal connectives (e.g., Clark, 1971). Of notable interest was the high accuracy of responses to before-chronological sentences which supports our prediction that chronological order, a more familiar connective (*before*), and a medial connective position are factors that do not tax additional working memory resources. Our findings indicate that children as young as four years of age process two-clause sentences accurately in this way. The main effect of working memory on accuracy indicates that children with higher memory capacities comprehend two-clause sentences more accurately. Critically, we found that memory could explain why children displayed sentence specific performance. This finding supports previous research which informed us that children's comprehension of two-clause sentences containing *before* and *after* can be influenced by whether their memory capacity is sufficient to cope with the variability in the processing demands of our sentence structures (Just & Carpenter, 1992; Pyykkönen & Järvikivi, 2012; Ye et al., 2012).

The observed effect of memory suggests that where age effects are found they may, as here, reflect a contribution due to the development of memory capacity. This finding highlights the difficulty of distinguishing the impact of the development of memory capacity from the impact of more general and language development. In addition, our results also highlight the need to study specific connectives within a single temporal class (Cain & Nash, 2011; Crosson & Le Saux, 2013). Our observations suggest that knowledge of *before* was more robust than knowledge of *after*. This may be due to their differences in semantics (Clark, 1971) or to their differing frequency of occurrence as temporal connectives (Leech et al., 2001).

Recent literature suggests that working memory tests that have been used to support capacity theories might instead tap into long-term knowledge of language (Kidd, 2013; Klem et al., 2015). For that reason, we tested whether our proposed memory effects were a result of long-term knowledge of language by running two models that incorporated vocabulary scores, one with and one without memory. We concluded from model comparisons that vocabulary did not significantly improve the fit compared to equivalent models that did not include vocabulary. The findings confirmed that working memory was driving performance. In addition, it is important to note that we manipulated sentence structures while holding vocabulary constant. That is, vocabulary did not vary across experimental conditions (other than for the specific connective itself), which runs counter to this alternative explanation that vocabulary knowledge could be a proxy for the reported memory effects (see also Cain, Patson, & Andrews, 2005, for an example of the separation between vocabulary knowledge and connective comprehension in young children).

We would not dismiss a language account completely. Of note our sentences differed by connective type and we believe that language knowledge variation could explain the general advantage for *before*. In addition, frequency of exposure to specific sentence structure can influence comprehension (Tomasello, 2003). Therefore, future work should consider both the frequency of our sentence structures and the use of *before* and *after* in parental input and examine whether this maps onto the pattern of development found in the present study.

Convergent with the memory capacity explanation of children's difficulties with temporal connectives in two-clause sentences, we found that the inclusion of an independent measure of working memory improved model fit. However, further evidence is needed to corroborate this account. We used the most sensitive behavioural measure of working memory that we could identify for our age groups, but believe that other techniques will support our findings and reveal critical pressure points in the moment-by-moment processing of these sentences. The extent to which the factors of event order, connective, and position influence the real time processing of connectives in young children may be studied with techniques that do not require a behavioural response, such as using eye-tracking within a visual world paradigm (Arnold, Eisenband, Brown-Schmidt, & Trueswell, 2000) or by recording ERPs to index processing difficulty, as has been done successfully in studies of adults' production of connectives (Habets, Jansma, & Münte, 2008). These techniques would provide fine-grained measures of processing efficiency and processing cost in critical regions of two-clause sentences (Cozijn, Noordman & Vonk, 2011). Response times or evoked potential differences for regions where the cognitive demands were greatest, in particular, might be more strongly related to independent measures of memory.

In addition to the limitations discussed above, we note that we used experimenter-constructed sentences, rather than sentences based on natural speech. Previous research on children's understanding of complex sentences show that difficulties can disappear when target sentences are based upon naturalistic speech (Kidd, Brandt, Lieven, & Tomasello, 2007; Rowland & Noble, 2010). However, we contend that these sentences parallel those found in naturalistic speech: Diessel reports examples of children's (2004) and adults' speech (2005, 2008) containing numerous examples of sentences containing temporal connectives that are supported by world knowledge or not, just like our experimental manipulation. Therefore we do not believe this is the reason for our findings although various pragmatic manipulations could be explored in future research. Further, we did not programme the randomisation of items to prevent potential priming when the same sentence structure is presented twice in a row. Given the number of items (N=64) we do not believe that this feature unduly influenced our findings. However, it would be interesting in future work to test if such features could be used to support language comprehension of more difficult syntactic structures, as has been found for language production (e.g., Allen, Haywood, Rajendran, & Branigan, 2011).

A final thought for future research is whether the same factors that influence comprehension of sentences with temporal connectives also influence production. Typically, in terms of language knowledge, comprehension generally precedes production for specific words and grammatical structures (Benedict, 1979; Fraser, Bellugi, & Brown, 1963). One reason for that difference may be the additional planning demands of language production (Diessel, 2004; MacDonald, 2013). However, comprehension and production are related and draw on many of the same cognitive processes (Pickering & Garrod, 2013). There is a clear need for comparison of children's comprehension versus production of complex sentences containing connectives to provide insight into common sources of difficulty. Critically, it would be important to determine whether the processing patterns for specific sentence types reported by the present study, map onto performance in a production paradigm.

In summary, the present study demonstrates substantial differences between 3- to 7-year-old children in their comprehension of two-clause sentences containing *before* and *after* and the factors that influence performance. The 3- to 4-year-olds demonstrated poor knowledge of the distinction between the meanings of these two temporal connectives and tended to interpret the event order as the order of mention of events. Older children's performance indicated adequate understanding of these connectives, but the poorer performance of 4- to 6-year-olds relative to 7-year-olds, together with the relations with memory, indicated that comprehension may fail when additional information must be held in working memory. Further research using online measures of sentence processing measures is required to identify the locus of difficulty in these sentences to elucidate the role of processing resources in children's comprehension of two-clause sentences.

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

Text submitted to *Developmental Psychology* (accepted).

CHAPTER 3: PROCESSING SENTENCES CONTAINING TEMPORAL CONNECTIVES.

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 a linearly 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.

CHAPTER 3: PROCESSING SENTENCES CONTAINING TEMPORAL CONNECTIVES.

3.1. Introduction

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 & Radwansky, 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 the processing of language. 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), and adults expend more cognitive effort when processing such sentences for meaning (Münte, Schiltz, & Kutas, 1998; Ye et al., 2012).

A developmental perspective identifies two reasons for difficulty with reverse order sentences. First, children may not have an appreciation for the meaning of the connective itself. If so, they will be more likely 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 shown that young children who display a poor knowledge of *before* and *after* use a strategy which assumes that the order of mention of events corresponds to the actual order of events, resulting in below-chance accuracy for reverse sentences (e.g., 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 already described in a chronological order (Münte, 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 thesis). This set of findings motivated the current study to contrast memory capacity (e.g., Just & Carpenter, 1992) and language-based (e.g., Van Dyke, Johns, & Kukona, 2014) accounts in relation to young children's comprehension of sentences containing temporal connectives.

A memory capacity 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 account would predict that individuals with low memory capacity would experience comprehension difficulties specifically for these constructions. Support for the memory capacity 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, the connective used (*before, after)* may also influence the demands on working memory resources for these two-clause sentences. 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 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 account (e.g., Just & Carpenter, 1992).

An alternative hypothesis, is that the influence of memory on the processing of complex sentences is actually driven by the quality of language knowledge rather than by the quantity of information 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 contrasts the memory capacity 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 for sentence processing effects, 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-old 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 type. Therefore, like others (e.g., Pyykkönen & Järvikivi, 2012), we can also relate our findings to connective position in the sentence.

We hypothesized an overall advantage for sentences linked by *before* compared to sentences linked by *after,* as the latter is more semantically complex (Clark, 1971) and is used less consistently as a temporal connective because of its use as an adverb, adjective, and preposition (e.g., Leech et al., 2001). In addition, the youngest children were expected to use an order of mention strategy to compensate for a fragile understanding of the connective. 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. 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).

Different patterns of performance are predicted by the memory capacity account (e.g., Just & Carpenter, 1992) and the language-based account (e.g., Van Dyke et al., 2014). According to a memory capacity account (e.g., Just & Carpenter, 1992) 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). Further, this pattern of performance will be predicted by an independent measure of memory.

A language-based account (e.g., Van Dyke et al., 2014) would predict that language knowledge, as measured by performance across connective (*before*, *after*) and by an independent measure of vocabulary, modulates 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.

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 investigation will advance our knowledge for the role of memory and language in children's processing and comprehension of two-clause sentences linked by sequential temporal connectives.

3.2. 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 3.1.

Table 3.1.

Sentence conditions

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

Children practiced each of the four sentence conditions (Table 3.1). Practice items were not included in the test phase. Each child fulfilled the training requirements by providing answers (touching a side of the screen) and by responding to corrective feedback (i.e., either explicitly saying that they understand or by demonstrating an understanding by this time touching the correct side of the screen). Practice trial instructions emphasized the importance of making judgments based solely on the meaning of the narrated sentence, not the visual stimuli. Before each practice item, the experimenter provided a similar instruction to the recorded instruction that children were about to hear, but emphasised speeded response ‘*Listen carefully and touch the thing Tom/Sue did last - as fast as you can!*’ The 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 (*p*s >.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 3.1). The dependent variables were accuracy and response times.

3.3. 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.3.1, 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 *p*s > .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.3.2, 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 A.3.2). Memory and vocabulary were strongly correlated (r = .69), so were both centred. Memory ($\chi^2(2) = 7.23$, $p < .03$) and vocabulary ($\chi^2(2) = 7.23$, $p < .03$) were added separately to the pruned model (Table A.3.2 in Appendix), and both improved the fit.

The inferential statistics are presented in Table A.3.2. 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 3.1. 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.3.1 and Table A.3.2, see appendix): these had reported a main effect of age, with each of the three older age groups performing significantly more accurate 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 3.2), 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.3.1 and Table A.3.2), all *p*s > .90.

Figure 3.1. Mean proportion correct (with standard error bars) for each experimental condition by 3- to 7-year-olds

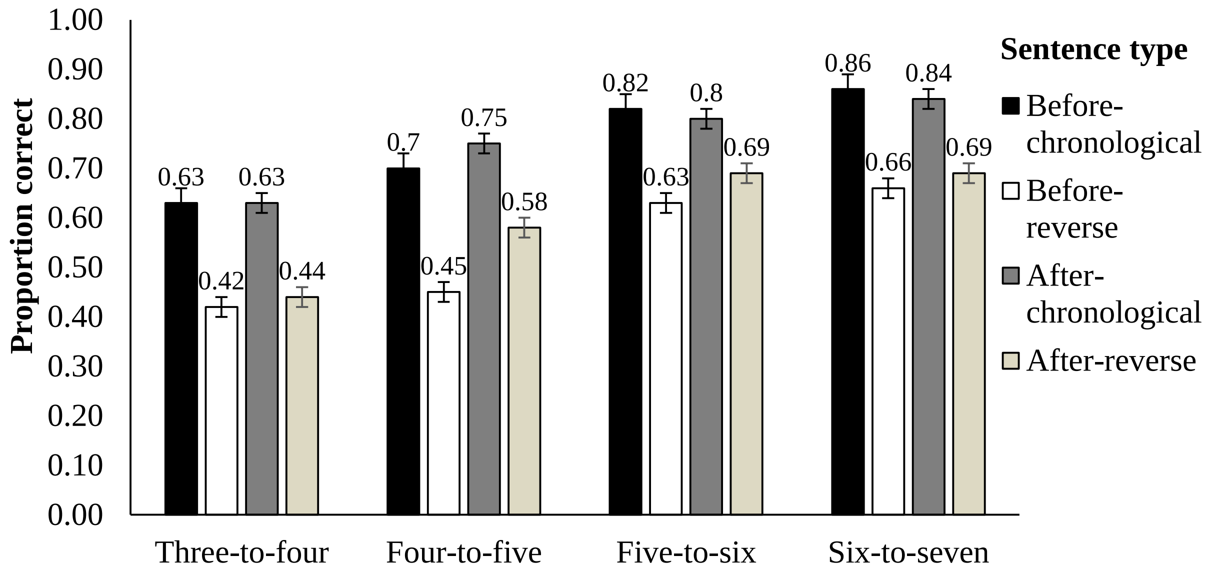


Table 3.2.

Summary of GLMM: 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 2.5%**	**CI 97.5%**	**p(>|z)**
(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	0.00	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

*Note: 1. Bold = predictor is significant at *p*<.05 or better. 2. See Table A.7.3 for zero order correlations (appendix).

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 better 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 3.2. 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.3.4, 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.3.3, Appendix), but our final model (Table 3.3) 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 3.2. Mean response times (with standard error bars) for each experimental condition by 4- to 7-year-olds

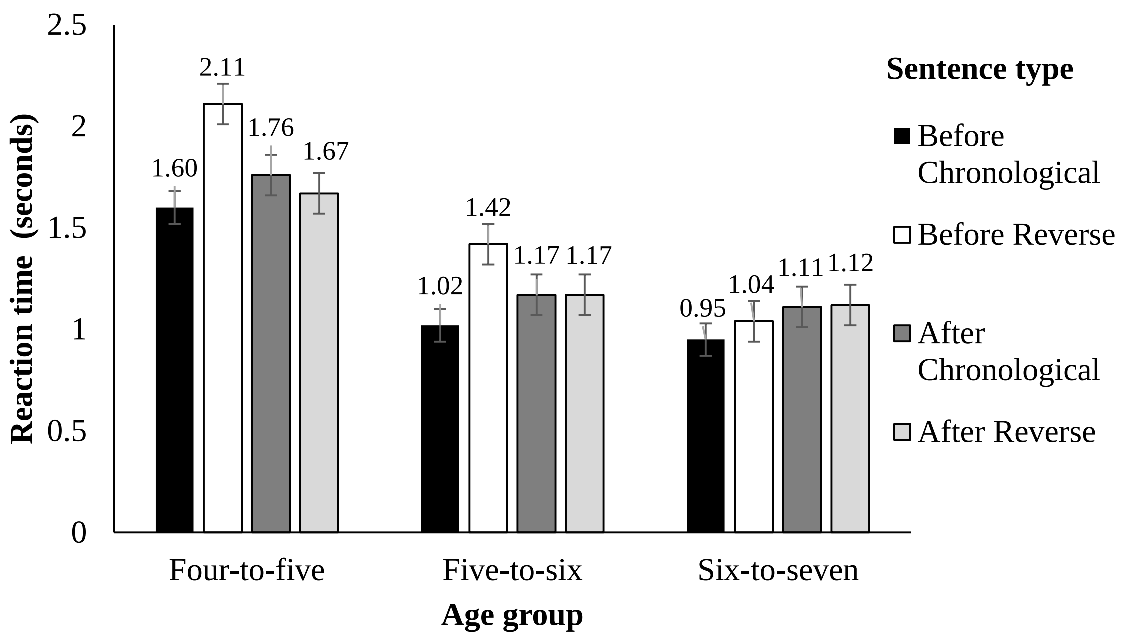


Table 3.3 summarises the main effects and interactions of memory, age, order and connective on response times. Similar to the accuracy analysis, there was no main effect of age once memory was added as a covariate, indicating that working memory was driving the developmental improvement in the processing 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 3.4 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, $\chi2(4) = 11.43$, $p = .02$. 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, $\chi2(4) = 6.53$, $p = .16$. 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 3.3

Summary of GLMM: 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%**
(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**	**<0.01**	**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	0.00
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**

*Note: 1. Bold = predictor is significant at $p<.05$ or better. 2. See Table A.7.4 for zero order correlations (appendix).

Table 3.4
Summary of GLMMs: 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															
(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
After															
(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

Notes. 1. Bold = predictor is significant at $p<.05$ or better.

3.4. 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. There were developmental improvements in performance, such that sentences were understood more accurately and processed more quickly by older children. Consistent with previous research (e.g., Clark, 1971), children were less accurate at comprehending reverse order compared to chronological sentences. In addition, response times were fastest for the chronological sentences that afford an incremental processing strategy. Responses were slowest for sentences that require the comprehender (or speaker) to maintain more information in working memory whilst constructing their mental representation. Critically, the developmental improvements, and the variation in performance across these sentence structures, were driven by children’s memory capacity. 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 comprehension of these sentences is best explained by a memory capacity 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 order 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 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 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 require a greater amount of information to be maintained in working memory. 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. 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 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 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, the finding that *before* sentences were responded to faster than *after* sentences means that we should not dismiss language effects *per se*. On its own this pattern of performance supports a simple form of a language-based account, such that language knowledge (of the connective) directly influences ease of processing. However, children with a higher working memory capacity were less likely to display such effects. Therefore, these connective effects can be interpreted in line with a memory capacity 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.[2] 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

[2] Older children did display a slight reductions 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.

supporting previous forced-choice accuracy studies that have attributed children's inaccurate comprehension to a difficulty in representing sentences that are more taxing on working memory (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 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). Our 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.

4. The role of memory and language ability in children's knowledge and production of two-clause sentences containing *before* and *after.*

Text submitted to *Cognition*.

CHAPTER 4: CHILDREN'S PRODUCTION OF BEFORE AND AFTER.

Abstract

We assessed 3- to 6-year-old's production of two-clause sentences linked by *before* or *after*. In two experiments (Experiment 3 and 4 in this thesis), we manipulated whether the presentation order matched the chronological order of events: '*He finished his homework, before he played in the garden*' (chronological order) vs. '*Before he played in the garden, he finished his homework*' (reverse order). Children were significantly less likely to accurately produce target sentences when the presentation order of the two clauses did not match the chronological order of events, specifically for target sentences linked by *after*. An independent measure of vocabulary appeared to be a stronger predictor of performance than an independent measure of working memory. We conclude that children's difficulties with the production of two-clause sentences linked by a sequential temporal connective arise because of weak language skills, rather than poor memory capacity.

CHAPTER 4: CHILDREN'S PRODUCTION OF BEFORE AND AFTER.

4.1. Introduction

We experience events in the world around us in real time as they occur. In the production of speech and text, however, the speaker or writer does not have to relate events in the order in which they occur; instead, linguistic devices such as the temporal connectives *before* and *after* may be used to refer to events in reverse order, for example, '*Before he ate the cookies, he put on his jumper.'* Although children produce sentences containing *before* and *after* from around 3 years of age (Diessel, 2004), they have difficulties with correct usage up to at least 9 years (Peterson & McCabe, 1987; Winskel, 2003). That is, children's production of sentences that include these expressions may belie their full competence, as they may have a better understanding of one construction over the other. In this study, we focus on 3- to 6-year-olds' production of two-clause sentences containing the connectives *before* vs. *after*. We demonstrate that language ability appears to have a stronger influence on performance than memory.

Successful understanding of multiple clause sentences results in an integrated and coherent mental representation of the state of affairs described, rather than the specific words or syntactic structures used to relate the events (Johnson-Laird, 1983; van Dijk & Kintsch, 1983). Importantly, a speaker's mental representation is linear with the actual temporal order of events: the event that occurs first in time is followed by the event that occurs second (Givón, 1991). That is, the speaker essentially imagines the events being acted out (Bereiter & Scardamalia, 1987; Bos, Bjorn, de Koning, van Wesel, Boonstra, & van der Schoot, 2015). Therefore, when a speaker chooses to narrate events in reverse order, as in *'She put on her gloves, after she had combed her hair.',* the language itself does not map onto the mental representation of events. It follows that incremental word-by-word or clause-by-clause planning and production is not afforded for a reverse order sentence. Instead, the speaker must draw upon greater processing resources because the clause sequence used for reverse order sentences deviates from the linear mental representation (Ye, Habets, Jansma, & Münte, 2011).

Despite their greater processing demands, sentences with reverse order structures occur frequently in speech, as well as in writing (Diessel, 2005, 2008). This may arise because other linguistic and structural features of sentences with temporal connectives, namely connective and position of connective, may in themselves contribute to processing ease or difficulty. For example, if a speaker chooses to narrate events in a chronological order, they can use either *before* or *after*, and this in turn influences whether the temporal connective occurs in an initial position of the sentence (e.g., *After she combed her hair, she put on her gloves*) or in the medial position (e.g., *She combed her hair, before she put on her gloves*). Our first question is how do these features – order of events, connective, position of connective – individually or in combination influence young native speakers' production of sentences containing temporal connectives? Our second question is what best explains any variation in performance between sentences and developmentally: a memory capacity account or a language-based account? We now consider these accounts, in turn.

The primary emphasis of a memory capacity account (King & Just, 1991) is that some sentence structures are more difficult to process than others because they require more information to be held in working memory than other sentence structures. It follows that such processing difficulties are expected to be more likely in individuals with a low working memory capacity because they have fewer available resources for holding additional information in their working memory.

The influence of working memory on children's processing of *before* and *after* sentences has recently been demonstrated in comprehension. There is an advantage for sentences in which the events are narrated in their chronological order and this is greatest for sentences with *before,* in which the connective occurs in the sentence medial position

(Blything et al., 2015). This finding has been related to the processing demands imposed by different syntactic structures and event orders and there is support for this account from the prediction of children's comprehension performance by an independent measure of working memory (Blything et al., 2015). The pattern of difficulty may apply to sentence planning for production. Comprehension and production are related and draw on many of the same cognitive processes (Pickering & Garrod, 2013). ERPs measured during sentence production reveal that adult speakers require greater processing resources to formulate reverse order sentences compared to chronological sentences (Ye et al., 2011). Also, processing resources may be increased when the speaker produces a connective in an initial position, because she has to hold the ordering information in working memory from the beginning of the sentence and throughout the first clause (Diessel, 2004; 2008). Conversely, a medially placed connective provides the linguistic information about temporal order at a point close to when the events can be integrated during incremental processing of language (Cain & Nash, 2011; Traxler, Bybee & Pickering, 1997).

We do not yet know how the additional planning demands of language production influence children's accurate use of sentences expressing different temporal orders of events. Developmentally, there are significant improvements in children's memory capacity between 3 to 6 years (Gathercole, Pickering, Ambridge, & Wearing, 2004). If a memory capacity account best explains children's difficulties with accurately producing particular sentence structures, and also developmental improvements, we should find that independent measures of memory best predict sentence and developmental effects.

An alternative account is that language knowledge underpins young children's production of multi-clause sentences. There are subtle but important differences between *before* and *after*, which may lead to differences production and comprehension of sentences containing these connectives. *Before* is used more consistently as a temporal connective than *after,* which is commonly used also as a preposition, as in '*Watch out, he is only after your money*' (see The British National Corpus: Leech, Rayson, & Wilson, 2001). In addition, *after* is more semantically and cognitively complex than *before* (Clark, 1971). Together, these features may result in later age of acquisition for *after* relative to *before.* Recent work on children's language comprehension shows that 3- to 7-year-olds find multiple clause sentences containing *before* easier to process than those in which the two events are linked by *after* (Blything et al., 2015). The authors concluded that the *before* advantage might arise because children have a greater appreciation of its meaning. For sentence production, we might therefore expect that language ability would directly influence the accuracy of production of sentences containing temporal connectives, and that developmental differences would also occur because older children would have more accurate knowledge of the distinction between *before* and *after*.

A third viewpoint is another language-based account of sentence processing that proposes an indirect relation between memory and sentence processing modulated by language. According to this view, language knowledge influences sentence processing because good language skills support the accurate representation of information in verbal working memory (Klem, Melby-Lervåg, Hagtvet, Lyster, Gustafsson, & Hulme, 2015). Given that the amount of information that can be held in working memory is often far less than the length of a complex sentence, it has been argued that memory capacity alone cannot be an adequate explanation of sentence processing difficulties (Van Dyke & McElree, 2006; Van Dyke & Johns, 2012; Van Dyke et al., 2014). In support of this account, recent research with adults has demonstrated that the specificity or distinctness of retrieval cues in the text, rather than the distance between individual text elements, can account for why some sentences are more difficult to process than others (Van Dyke & Johns, 2012 Van Dyke et al., 2014).[3] A

[3] Full examples on page 20 (Chapter 1).

developmental account might posit that interference from competitors (i.e., temporal connectives other than the target connective, and words that share a similar but different meaning to the rest of the target words in the sentence) would be reduced with meaningful exposure to language, as lexical representations become more precise and robust. On this basis, younger language users may experience difficulties with processing complex sentences (i.e., reverse order) because the quality of their lexical representations is weaker.

Overview of study aims, methods, and hypotheses

The aim of the experiments reported in this paper was to determine whether a memory capacity (e.g., King & Just, 1991) or language-based account (e.g., Van Dyke et al., 2013) best explains young children's production of sentences containing *before* and *after*. We manipulated the connective type (*before, after*), and whether the order of mention of events was chronological or reverse. As a result, the position of the connective was manipulated (medial or initial). Note that a reverse order sentence with *after* places the connective in the medial position, whereas a reverse order sentence with *before* places the connective in the initial position. Thus we manipulated language (choice of connective) and memory processing demands (chronological vs. reverse order) in our materials. To study the factors that influence performance, we also examined the extent to which independent measures of language (receptive vocabulary) and working memory explained variance in performance.

The majority of studies examining the production of sentences containing temporal connectives have used an elicited production paradigm. In this, the child is shown a sequence of pictures and asked to describe the events (Atassanova, 2001; Clark, 1971; Silva, 1991; Weist et al., 1997; Winskel, 2007). These studies demonstrate that children as young as three have the ability to produce semantically specific connectives in two-clause sentences. However, the design of these studies does not restrict the speaker to use a specific sentence structure (Ambridge & Rowland, 2013). We report two experiments (Experiment 3 and 4 of the thesis) designed to address this limitation: a sentence repetition task, and an elicited production task with a block design presentation.

If memory capacity is a critical influence on children's production of complex sentences, we would expect sentences that relate events in reverse order to be produced less accurately than those that relate events in chronological order. Specifically, before-chronological sentences should be produced most accurately because they are associated with the factors that do not increase demands on working memory (chronological order, medial position, earlier acquired connective), whereas the other structures each have two factors that increase demands on working memory: before-reverse (reverse order, initial position), after-chronological (initial position, later acquired connective), and after-reverse (reverse order, later acquired connective). Critically, our independent measure of memory should explain unique variance in performance. In addition, developmental improvements should serve as a proxy for memory once both age and memory have been incorporated into the model. That is, where age effects are found, they may be better reflected by the development of memory capacity.

A language-based account would predict that language knowledge, as measured by performance across connective and on our independent measure of vocabulary, explains unique variance in overall performance and modulates how well children can cope with the cognitive demands of difficult sentence structures. A simple form of the account would predict a general disadvantage for *after* sentences, as children's knowledge of *after* is more difficult to activate than *before.* A language-based account which acknowledges the influence of memory would provide a more specific prediction: an influence of memory (i.e., order of mention, memory capacity) would be expected, but would be indirect and modulated by language knowledge (Van Dyke, Johns, & Kukona, 2014). Children should display a difficulty in producing reverse order sentences only when they are linked by the connective *after*: the planning of a reverse order sentence should be disrupted more easily when it contains *after* because knowledge of *after* is more difficult to activate than *before*. Critically, the language-based account would predict that vocabulary would modulate processing difficulties as limited vocabulary would be more likely to disrupt planning and production stages because it does not provide strong support for the representation of information in working memory. Relatedly, a developmental perspective would expect that since younger children have poorer language knowledge, age effects should serve as a proxy for vocabulary knowledge.

4.2. Experiment 3

In this experiment, we assessed sentence production using a sentence repetition task. In sentence repetition tasks, the participant hears a target sentence and is asked to repeat it. Sentence repetition is a sensitive measure of processing difficulties because the participant is required to process the syntactic and the semantic information, and then formulate the sentence themselves using the same sentence production mechanisms as in spontaneous speech (see Boyle et al. 2013; Klem et al., 2013; Lust, Lynn, & Foley, 1995). In general, children are less accurate when repeating sentences with more difficult structures. Previous studies of children's production of temporal connectives using sentence repetition have contrasted sequential (e.g., *then, before*) and simultaneous (e.g., *whilst, when*) connectives (Keller-Cohen, 1981; Winskel, 2003). However, these do not speak to the issues addressed in this paper.

4.2.1. Method

Participants

Sixty-seven monolingual, typically developing 3- to- 6-year-old children were recruited from schools of mixed socio-economic status in the North West region of England.

The data was collected between May and August 2013. Children were in three different school year groups: 20 3- to 4-year-olds (aged 3;5 to 4;7, 13 boys), 23 4- to 5-year-olds (aged 4;9 to 5;9, 12 boys), and 24 5- to 6-year-olds (aged 5;9 to 6;8, 11 boys). Written parental consent was obtained, and children provided oral assent before each session.

Materials and Procedure

All children completed a sentence repetition task that was split into two sessions, each lasting no longer than twenty minutes. One session was followed by an assessment of receptive vocabulary, the other by an assessment of memory.

Sentence repetition. Thirty-two two-clause sequences containing *before* and *after* were constructed (N=32). Each of the 32 items conveyed the temporal order of two events that were arbitrarily related (e.g., *he put on the socks, before he ate the burger*). These items were counterbalanced across four lists so that they each represented one of four sentence constructions (shown in Table 4.1). The four constructions were the product of manipulations of the order of mention of events (chronological or reverse) and the connective (*before, after*).

We also created 32 filler sentences, in which the sequence of events in a sentence was typical and supported by world knowledge (e.g., *he put on the socks, before he put on the shoes*)*,* rather than arbitrary. Sentences that relate typical sequences (world knowledge present) may reduce the working memory demands of the task by scaffolding the structure of the sentence (Zwaan & Radvansky, 1998). Previous work shows a benefit for children's production of two-clause sentences in act-out tasks (French & Brown, 1977; Keller-Cohen, 1987). We included these sentences to enhance the likelihood that children would produce full sentences in the task and to maintain their confidence.

Table 4.1

Sentence conditions

	Before	After
Chronological	He put on the sandals, before he ate the burger	After he put on the sandals, he ate the burger
Reverse	Before he ate the burger, he put on the sandals	He ate the burger, after he put on the sandals

Each sentence was visually represented by cartoon animations, one for each clause and each lasting three seconds. These were created using Anime Studio Pro 9.1 (Smith Micro Software, 2012). Animations make children more likely to use the actor, action and object of the target sentence, thus reducing task demands (Ambridge & Lieven, 2011). Each animation segment explicitly showed an object (e.g., mustard) from one of the clauses, and the object (e.g., burger) from the other clause was not present. Each animation segment was followed by a freeze-frame judged by the researchers to best represent the action of that clause. Each segment (e.g., Tom eating a hotdog) was 486 pixels in height and did not exceed the left or right half of the presentation (486 x 872 pixels). The experiment was run using the PsyScript 3.2.1 (Slavin, 2013) scripting environment on a Macintosh laptop connected to a monitor with items presented in a random order.

Practice trials emphasised the importance of producing an exact copy of the narrated target sentence. Children practiced each of the four sentence types in Table 4.1. Children who were not able to repeat a sentence after four practice trials completed another set of four practice trials. With this level of practice, all children were able to perform the task above the exclusion criteria: Each child was able to copy at least one sentence. The animation on the left hand side of the screen was shown first, followed by the animation on the right hand side of the screen. An instruction began with: '*Can you say…*', and was followed immediately by the narration of the target sentence. A response window was signalled by a short beep. The presentation order of the animation segments corresponded to the actual order of events, rather than the narrated order. This minimised the processing demands by ensuring that children understood the target order of events when required use the target sentence construction. Responses were recorded, later transcribed and finally scored.

In scoring, an exact repetition was marked as correct. Based on recommendations by Lust, Lynn and Foley (1995), a response was also marked as correct if the only mistake was a minor change to the specific wording of a subject (e.g., Sue, she), verb (e.g., put on, putted on), and/or object (e.g., ketchup, tomato sauce). This lenient criterion was used was because marking such changes as incorrect would create unnecessary noise when the main point of interest was to evaluate the variance that was caused by the factors we had hypothesised to affect children's ability to accurately communicate the order of events using a temporal connectives. The time taken between the beep and when the child began producing their response was extracted using Audacity (Mazzoni, 2014). There were no significant differences between age groups or sentence types for correct responses, so these results are not reported.

Errors were first categorised into three broad types: sense maintained, sense changed, and incomplete. We categorised responses as a *sense maintained error* if the child inaccurately repeated the target sentence, but successfully communicated the order of events by using a temporal connective as a linguistic device. Of course, since the sense of actual order is maintained, this might not be interpreted as an error per se. Nevertheless, the sense maintained errors were counted as errors because at least one critical feature of the target sentence was missed out (connective, order of mention, or position, see p183). That is, children did not display competence in the target structure that they were trained to produce. A *sense changed* error was defined as an inaccurate repetition of the target sentence in addition to inaccurate communication of the order of events (i.e. they changed the sense). Responses were categorised as *incomplete* when the child failed to respond, omitted a clause, failed to use a connective, or used the connective '*and*.' Responses which used the connective *and* (42 total, 45% of the incorrect responses that were analysed, see results section) were categorised as incomplete because *and* does not explicitly specify order (Peterson & McCabe, 1987), so we were unable to categorise whether the error maintained or changed the sense of order. Within each of the three broad error categories, we coded the specific error or combination of errors that the child had made. Our supplementary materials include examples and frequency counts of each specific error type that falls within a broad error category (see Table A.4.1).

A second coder blind to the hypotheses randomly selected individual children so that at least 10% of the data from each year group could be assessed for reliability. Agreement between the coders was very good for both the main analysis (Agreement: 99%; Cohen's k=.96) and the error analysis (Agreement: 96%; Cohen's k=.80).

Memory. Working memory was assessed using the digit span task from the Working Memory Battery for Children (Pickering & Gathercole, 2001). In this task, children are required to retain and recall the order of a string of digits that are read aloud by the assessor. The number of digits in a string is increased until the child cannot successfully recall strings of that length on three separate trials. This assessment of memory was selected because it is most appropriate for our youngest children, who have been reported to perform at floor on more complex measures of working memory (Gathercole, Pickering, Ambridge, & Wearing, 2004). Raw scores were used for the analysis. The test-retest reliability reported in the manual for children aged 5 to 7 years is high, $r = .81$.

Vocabulary. Each child completed the British Picture Vocabulary Scale – III (Dunn, Dunn, Styles, & Sewell, 2009). In this task, children hear a word and are asked to point to one of four pictures that best illustrates the meaning. Testing is discontinued when a specified number of errors have been made.

Design

A 3 x 2 x 2 mixed design was used. The between-subjects independent variable was year group (3-4, 4-5, and 5-6 years) and the within-subjects variables were connective (*before*, *after*), and order (chronological, reverse). The position of the connective was manipulated as a function of the manipulations of connective type and order. The dependent variables were accuracy and error type.

4.2.2. Results

Method of analysis.

The main analysis was completed using Generalised Linear Mixed-effects models (GLMMs) (Baayen, Davidson, & Bates, 2008; Barr, Levy, Scheepers & Tilly, 2013), followed by further models of data subsets in the event of any significant interactions. These were conducted using the lme4 package from the R statistics environment (R Core Team, 2014) (Bates, Maechler & Bolker, 2014). We followed the recommendations by Barr et al. (2013) for obtaining an optimal model. Our maximum random effects models did not converge, and so the decision to incorporate random intercepts and slopes for participants and items was determined by incremental likelihood ratio tests of whether each specific random effect significantly improved the fit of the model to the data (Barr et al., 2013). We describe the optimum models for each respective dataset. The model statistics are reported in Tables 4.2, 4.3, 4.4, and 4.5, in which the first column always provides the coefficient estimates of effects (*b*) due to experimental conditions, the change in the log odds accuracy of response associated with each fixed effect. A positive coefficient indicates that the effect of a factor is to increase the odds that a response would be correct while a negative coefficient indicates that the effect of a factor is to decrease the odds that a response would be correct. Age, order, and connective were entered as fixed effects.

Memory.

The raw memory scores [mean (SD)] demonstrated age-related improvements: 3- to 4-year-olds = 21.65 (5.66); 4- to 5-year-olds = 22.65 (3.7); 5- to 6-year-olds = 25.42 (3.45). In addition, the standardised scores of memory were within the normal range of 85-115 for each age group: 4- to 5-year-olds = 101.39 (12.43); and 5- to 6-year-olds = 105.96 (12.19). Standardised scores are not provided for 3- to 4-year-olds.

Vocabulary.

Raw vocabulary scores demonstrated age-related improvements: 3- to 4-year-olds = 72.65 (26.16); 4- to 5-year-olds = 78.26 (9.76); 5- to 6-year-olds = 102.30 (8.59). 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 = 111.35 (13.08); 4- to 5-year-olds = 101.22 (9.14); 5- to 6-year-olds = 101.54 (9.10).

Analysis of accuracy data.

A total of 2144 responses were recorded. Figure 4.1 shows the means for each sentence structure by age. The 5- to 6-year-olds were excluded from the analysis because they performed at ceiling. For the two younger groups, 19 responses were removed because they were inaudible, leaving 1357 responses for analysis. Only 13 responses were judged to be inappropriate responses (nonsense or no response), thereby indicating that children understood the purpose of the task. Table 4.2 reports the inferential statistics and summarises the main effects and interactions of age, order and connective.

Figure 4.1. Mean percentage correct (with standard error bars) for each experimental condition in a sentence repetition paradigm by age group.

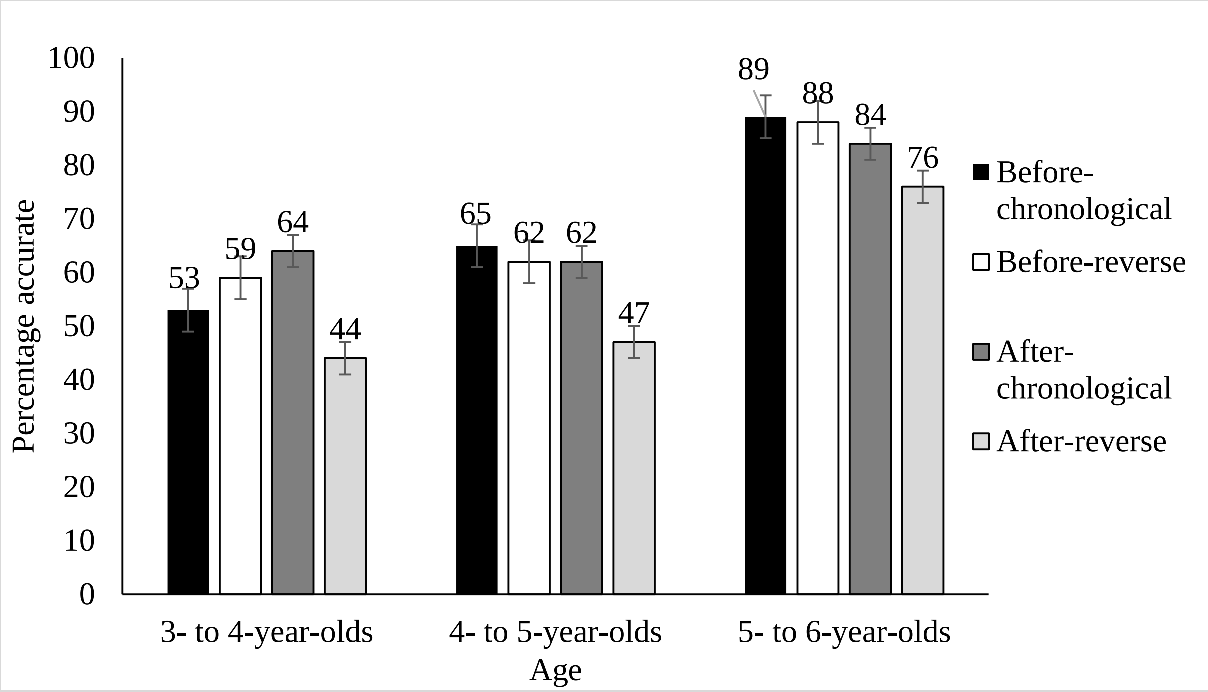


Table 4.2

Summary of GLMM: Main effect and interactions of age, order and connective on accuracy responses by 3- to 4- and 4- to 5- year-olds in the sentence repetition task.

Main model		**M (*b*)**	**SE**	**t**	**p(>|z|)**
	(Intercept)	-3.09	2.47	-1.25	0.21
	Age	0.05	0.04	1.11	0.27
	Order	1.21	1.97	0.62	0.54
	Connective	2.16	2.29	0.94	0.35
	Age:Order	<0.01	0.03	-0.02	0.99
	Age:Connective	-0.02	0.04	-0.54	0.59
	Order:Connective	**-5.26**	**2.25**	**-2.34**	**0.02**
	Age:Order:Connective	0.07	0.04	1.79	0.07
Before only					
	(Intercept)	0.52	0.44	1.17	0.24
	Age	0.06	0.60	0.11	0.92
	Order	-0.45	0.36	-1.28	0.20
	Age:Order	0.87	0.50	1.74	0.08
After only					
	(Intercept)	-0.39	0.52	-0.75	0.45
	Age	0.02	0.73	0.02	0.98
	Order	**1.20**	**0.42**	**2.85**	**<0.01**
	Age:Order	0.03	0.60	0.05	0.96

*Note: 1. Bold = predictor is significant at $p<.05$ or better. 2. See Table A.7.5 for zero order correlations (appendix).

Response accuracy was not significantly affected by participant age, indicating that performance did not significantly improve between 3 and 5 years. There was no main effect of order or connective. However, order and connective effects were involved in a significant two-way interaction. We further examined the significant interaction by conducting simple interaction analyses of the effects of order for each connective type separately. A main effect of order was displayed for *after* sentences, but not for *before* sentences. That is, children found it more difficult to accurately repeat after-reverse sentences compared to after-chronological sentences, whereas accuracy was equivalent for before-chronological and before-reverse sentences.

We also built upon the optimum model by incorporating memory and vocabulary as additional factors to age, order and connective (Table 4.3). The addition of these factors significantly improved the fit of the model, $\chi 2(8) = 45.05$, $p < .01$. Memory and vocabulary both significantly influenced performance, such that stronger sets of skills in both domains improved performance. The order by connective interaction remained significant, and there was no three way interaction with either memory or with vocabulary.

Table 4.3

Summary of GLMM: Main effect and interactions of age, memory, vocabulary, order and connective on 3- to 4- and 4- to 5- year old's accuracy responses in the sentence repetition task.

Main model	**M (*b*)**	**SE**	**t**	**p(>|z|)**
(Intercept)	-10.58	2.39	-4.43	<0.01
Age	-0.05	0.04	-1.41	0.16
Memory	**0.40**	**0.08**	**4.94**	**0.01**
Vocabulary	**0.06**	**0.03**	**2.22**	**0.03**
Order	3.06	2.54	1.20	0.23
Connective	2.79	2.89	0.96	0.33
Order:Connective	**-8.25**	**2.97**	**-2.78**	**0.01**
Age:Order	0.02	0.04	0.48	0.63
Age:Connective	<0.01	0.05	-0.07	0.95
Memory:Order	-0.08	0.09	-0.96	0.34
Memory:Connective	-0.13	0.10	-1.32	0.19
Vocabulary:Order	-0.01	0.03	-0.54	0.59
Vocabulary:Connective	0.02	0.03	0.53	0.60
Age:Order:Connective	0.04	0.04	1.00	0.32
Memory:Order:Connective	0.13	0.10	1.34	0.18
Vocabulary:Order:Connective	0.02	0.03	0.70	0.48

*Note: 1. Bold = predictor is significant at $p<.05$ or better. 2. See Table A.7.6 for zero order correlations (appendix).

Analyses of error types

The frequency of different types of error was investigated to determine whether particular types of error were associated with specific experimental conditions and/or age group. This provided the opportunity to examine additional support for either the memory or the language accounts, outlined in the Introduction. We excluded responses from the oldest age group due to ceiling performance. All 775 correct responses by the youngest two age groups were excluded, leaving a total of 582 errors for analysis. The majority of errors involved a change of sense to the target sentence's meaning (*sense changed* = 358: 61% of errors), whereas fewer errors maintained the sentence meaning (*sense maintained* = 131: 23%) or were incomplete (*incomplete responses* = 93: 16%).

To examine error type further, we calculated the percentage of *sense changed* errors that involved a connective, order, or a position substitution. For this error type, a notable observation was that connective substitution was the most common error (252; 70%), whereas position (109; 30%) and order (78; 22%) changes were fewer. These values add up to more than 100% because the error types are not mutually exclusive such that children could make more than one error per response. We conducted a further analysis to examine the variance in the percentage of the total errors for each experimental condition that was caused by the most common error type: sense changed errors that involved a substitution of the target connective. Because these categorical decisions are not independent, we first excluded errors involving connective substitutions other than *before* and *after* (e.g., *then, and then, when*). There were 218 remaining errors - substitutions of *before* instead of *after*, or *after* instead of *before*. These errors were involved in a significantly larger percentage of the overall errors for reverse order sentences (47%) compared to chronological sentences (33%). The errors did not make up a significantly different percentage of the total by-age group or by-connective errors, and did not feature in any significant interactions. A model summary for this error analysis is available in the supplementary materials (Table A.4.2).

4.2.3. Discussion

The sentence repetition task was successful at eliciting production of complete two-clause sentences linked by an appropriate temporal connective, yielding very few incomplete responses (0.7% of total responses by the two youngest age groups). Together our analyses demonstrate support for both the memory capacity (King & Just, 1991) and the language-based (e.g., Van Dyke et al., 2014) accounts of sentence production. First, our experimental manipulations of connective and sentence order demonstrated an influence of processing load and connective type: children demonstrated a difficulty in producing events in a reverse order, most notably for after-reverse sentences. Second, performance on independent measures of memory and vocabulary improved the overall fit of the model.

The difficulty with reverse order sentences was found only for sentences containing *after*. Taken together, these findings do not provide unanimous support for the memory capacity account. Instead they can be interpreted as support for a language-based explanation that proposes an indirect relation between memory and sentence processing modulated by language, because the difficulty with reverse order sentences appears to be modulated by connective type. On this view, young children's lexical representations for *after* are less precise and secure than those for *before,* because the former is used less consistently as a temporal connective. For that reason, it may be more difficult to accurately plan and maintain in memory multi-clause sentences linked by *after* sentences during language production, particularly when the event order is reversed. In that way, variation in language knowledge may lead to difficulties with sentences that have a high processing load.

We need to be cautious in this language-based interpretation, because there was no main effect of connective to confirm that *before* was generally easier than *after*. However, our analysis of error subtypes provided additional support for the language-based account, because connective substitution errors that used *before* and *after* to communicate an incorrect event order (sense changed errors) were more likely for reverse order than for chronological order sentences. This pattern indicates that an inaccurate representation of the connective itself is more likely to interfere with the production of reverse order rather than chronological order sentences.

4.3. Experiment 4

Experiment 4 sought to replicate these findings using a different method to elicit sentence productions. A limitation with the sentence repetition paradigm used in Experiment 3 is that there are additional demands on memory because the child has to store the narrated sentence prior to production. For that reason, sentence repetition may be an insensitive method to differentiate memory capacity and language-based accounts of children's difficulties with sentence production. Experiment 4 comprised four blocked sessions that were each independently assigned to elicit one of the four target sentence constructions (e.g., Huttenlocher, Vasilyeva, & Shimpi, 2004). These blocked conditions were designed to complement Experiment 3 by minimising the contributions of sentence comprehension and memory that are associated with sentence repetition, and elicit the spontaneous production of sentences.

4.3.1. Method

Participants

A new set of participants were recruited (N = 68). They were 23 3- to 4-year-olds (aged 3;8 to 4;11, 10 boys), 24 4- to 5-year-olds (aged 4;9 to 5;9, 13 boys), and 21 5- to 6-year-olds (aged 5;10 to 6;9, 10 boys).

Materials, Procedure, and Design

Children completed the same independent measures of memory and receptive vocabulary as in Experiment 3. Sentence production was assessed using an elicited production task with a blocked design over two separate sessions. Each session lasted no longer than twenty minutes. One session included the vocabulary assessment, the other the memory assessment.

Elicited production: Blocked design. The same stimuli from Experiment 3 were used. The 64 items (32 fillers) were split into four testing blocks, each preceded by a training phase in which children were instructed to use a specific target sentence structure. Depending on which block children were first counterbalanced to perform, the experimenter provided the instruction: '*In this game, I am going ask you to watch two videos and to say what happened using the word before/after.* I want you to tell me the order that he/she did these things, and[4] *I want you to use before/after in the middle/at the start of your sentence*. Corrective feedback was provided for all four practice items, and training was repeated. Three 3- to 4-year-olds and one 5-to 6-year-old were excluded from testing after this phase because they each failed to correctly produce a target structure.

As in Experiment 3, the order that the animations were presented corresponded to the order of events described by the target sentence. An instruction was narrated: '*Can you tell me the order that Tom did these things?*'. A response window was signalled by a short beep. The four blocked conditions were counterbalanced. Responses were recorded and were later transcribed and scored.

We employed the same criteria for scoring accuracy and categorisation of errors as Experiment 3. We did not analyse the time taken to start a response because this measure was shown not be sensitive in Experiment 3. Agreement between the coders was very good for

[4] ('*and*' replaced by '*but this time*' for the 2nd, 3rd, and 4th training block sessions)

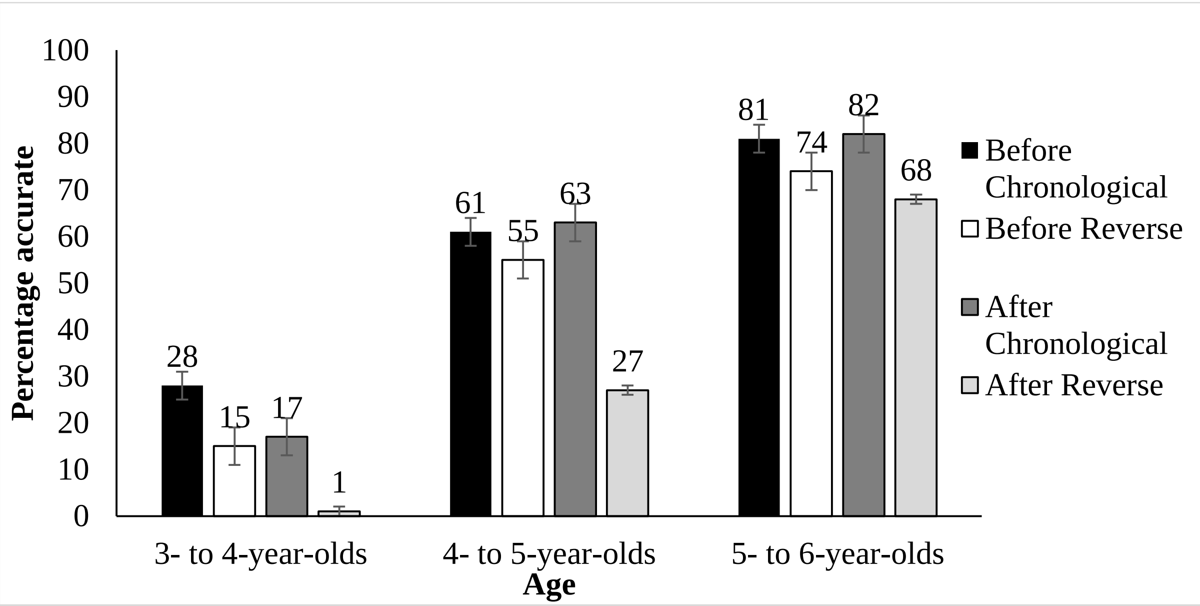
both the main analysis (Agreement: 99%; Cohen's k=.97) and the error analysis (Agreement: 96%; Cohen's k=.96).

4.3.2. Results

Data Extraction.

A total of 45 responses (2%) were excluded because they were inaudible or interrupted, leaving 1345 responses. Figure 4.2 reports the mean accuracy scores to each experimental condition, by age group. We report the optimum model with age, order, and connective entered as fixed effects (Table 4.4) and an additional model that incorporates memory and vocabulary as additional factors to age, order and connective (Table 4.5).

Figure 4.2. Mean percentage correct (with standard error bars) for each experimental condition in a blocked elicited production paradigm by age group.



Memory.

The raw memory scores [mean (SD)] demonstrated age-related improvements: 3- to 4-year-olds = 21.15 (2.12); 4- to 5-year-olds = 22.57 (2.12); 5- to 6-year-olds = 25.05 (4.95). In addition, the standardised scores of memory were within the normal range of 85-115 for each age group: 4- to 5-year-olds =100.52 (14.85); and 5- to 6-year-olds = 105.9 (12.73). Standardised scores are not provided in the manual for 3- to 4-year-olds.

Vocabulary.

The raw memory scores [mean (SD)] demonstrated age-related improvements: 3- to 4-year-olds = 70.85 (7.78); 4- to 5-year-olds = 82.48 (12.02); 5- to 6-year-olds = 90.95 (9.90). 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 = 111.75 (7.07); 4- to 5-year-olds = 100.45 (14.85); 5- to 6-year-olds = 102.15 (16.26).

Analysis of accuracy data.

The inferential statistics for the accuracy analysis are presented in Table 4.4. The pattern of data differs from that found in Experiment 3. There were main effects of age, order, and connective, and also significant two-way interactions between these variables. The effects were qualified by a significant three-way interaction. We examined the significant interaction by conducting simple interaction analyses for the effects of age and order for each connective type separately. For *after* sentences, there were main effects of age and order and these were also involved in a significant two-way interaction; for *before* sentences, only the main effect of age reached statistical significance (see Table 4.4 for a full breakdown of results and Figure 4.2 for graphs of these effects by sentence type). In sum, children found it more difficult to accurately repeat after-reverse sentences compared to after-chronological sentences; whereas accuracy was equivalent for before-chronological and before-reverse sentences. This effect was more pronounced in younger children.

We tested three additional models, as follows. The addition of memory to the original model significantly improved the fit of the data, $\chi2(4) = 20.11$, $p = .01$. Of note, there was a significant three-way interaction between memory, order and connective. This finding suggests that memory modulated the interaction between connective and order. In a second model, we added vocabulary to the original model and also found improved the fit compared with the original model, $\chi2(8) = 33.57$, $p = .01$. In a third model, we included both vocabulary and memory. The inclusion of vocabulary resulted in an improved the fit compared with the memory model, $\chi2(4) = 12.08$, $p = .02$. There was a main effect of vocabulary, but not memory. In addition, the memory by order by connective interaction was not evident when vocabulary was also present (see Table 4.5). We include a copy of the memory alone model in the supplementary material (Table A.4.3).

Table 4.4

Summary of GLMM: Main effect and interactions of age, order and connective on accuracy responses by 3- to 6- year-olds in the elicited production task.

Main model		**M (*b*)**	**SE**	**t**	**p(>|z|)**
	(Intercept)	-32.09	4.37	-7.35	<0.01
	Age	**0.46**	**0.07**	**6.90**	**<0.01**
	Order	**14.38**	**4.32**	**3.33**	**<0.01**
	Connective	**14.16**	**2.39**	**5.93**	**<0.01**
	Age:Order	**-0.17**	**0.07**	**-2.63**	**0.01**
	Age:Connective	**-0.19**	**0.04**	**-5.29**	**<0.01**
	Order:Connective	**-11.00**	**2.96**	**-3.72**	**<0.01**
	Age:Order:Connective	**0.14**	**0.04**	**3.17**	**<0.01**
Before only					
	(Intercept)	-33.62	8.20	-4.10	<0.01
	Age	**0.53**	**0.13**	**4.12**	**<0.01**
	Order	-4.19	11.92	-0.35	0.73
	Age:Order	0.10	0.19	0.52	0.60
After only					
	(Intercept)	-50.59	11.29	-4.48	<0.01
	Age	**0.72**	**0.17**	**4.29**	**<0.01**
	Order	**29.92**	**10.95**	**2.73**	**0.01**
	Age:Order	**-0.38**	**0.16**	**-2.35**	**0.02**

*Note: 1. Bold = predictor is significant at $p<.05$ or better. 2. See Table A.7.7 for zero order correlations (appendix).

Table 4.5

Summary of GLMM: Main effect and interactions of age, memory, vocabulary, order and connective on 3- to 6- year old's accuracy responses in the elicited production task.

Main model	**M (*b*)**	**SE**	**t**	**p(>|z)**
(Intercept)	-37.87	5.37	-7.06	<0.01
Age	**0.38**	**0.08**	**4.56**	**<0.01**
Memory	-0.12	0.17	-0.73	0.47
Vocabulary	**0.16**	**0.06**	**2.78**	**0.01**
Order	**17.26**	**5.55**	**3.11**	**<0.01**
Connective	**14.41**	**3.10**	**4.66**	**<0.01**
Order:Connective	**-11.50**	**3.72**	**-3.09**	**<0.01**
Age:Order	**-0.19**	**0.08**	**-2.22**	**0.03**
Age:Connective	**-0.29**	**0.05**	**-5.83**	**<0.01**
Memory:Order	0.12	0.17	0.70	0.49
Memory:Connective	**0.35**	**0.09**	**3.82**	**<0.01**
Vocabulary:Order	-0.06	0.06	-0.96	0.34
Vocabulary:Connective	-0.02	0.03	-0.71	0.48
Age:Order:Connective	**0.26**	**0.06**	**4.40**	**<0.01**
Memory:Order:Connective	-0.19	0.12	-1.65	0.10
Vocabulary:Order:Connective	-0.04	0.04	-0.97	0.33

*Note: 1. Bold = predictor is significant at $p<.05$ or better. 2. See Table A.7.8 for zero order correlations (appendix).

Analyses of error types

Responses by 5- to 6-year-olds were excluded because they made too few errors for analysis (152, 15% of total errors by the three age groups). We analysed the 867 errors made by 3- to 4-year-olds and 4- to 5-year-olds. The highest percentage of errors was sense maintained (410; 47%), followed by incomplete (305, 35%). Sense changed errors made up the lowest percentage of the total errors (152; 18%). These findings contrast with the sentence repetition task, which found a high percentage of sense changed errors. The different error types did not vary significantly by experimental conditions, although 3- to 4-year-olds (220; 42%) made a substantially larger number of incomplete errors compared to 4- to 5-year-olds (85; 25%).

To examine error type further, we calculated the percentage of *sense maintained* errors that involved a connective, order, or a position substitution. As noted in Experiment 3, these error types do not add up to 100% because more than one error can be made for a single response. A notable observation was that connective substitution was the most common type of sense maintained error (313; 76%), but position (237; 58%) and order (256; 62%) changes were also both involved in over half of the total errors that maintained the sentence meaning. Of the 313 sense maintained errors, there were only 129 connective substitution errors that involved the direct replacement of *before* for *after*, or *after* for *before*. Therefore, unlike Experiment 3, there were too few errors of this type for further analysis.

4.3.3. Discussion

The elicited production task replicated the main finding in Experiment 3: a pronounced difficulty in producing reverse order sentences linked by the connective *after*. That is, we again found that difficulty with reverse order sentences was modulated by connective type: this effect was limited to after-reverse sentences. In contrast to Experiment 3, there was a main effect of connective, because *after* was more difficult than *before,* in general. A critical difference between the two experiments was that inaccurate production of after-reverse sentences was not modulated by children's working memory capacity in Experiment 4, when we accounted for the variance of vocabulary. Together, these findings suggest that language knowledge, rather than memory, is the stronger determiner of accurate sentence production. Our error analysis revealed a lower proportion of errors involving a change of sense, compared with Experiment 3.

4.4. General Discussion

These two experiments set out to examine why young children have difficulties in producing two-clause sentences containing *before* and *after* in the developmental period that follows their emergence in spontaneous speech. The findings complement previous work (Diessel, 2005) by showing that, when 3- to 6-year-olds produce sentences containing *before* and *after* as temporal connectives, their competence is not yet robust. That is, children do not fully understand how to use semantically restricted connectives to specify the temporal relation between events in the same way as adults (Peterson & McCabe, 1987; also see Winskel, 2003). Specifically, in both experiments, children up to 6 years of age had difficulties in producing reverse order sentences linked by the connective *after*. The results also show developmental improvements in performance: the oldest children performed at ceiling in Experiment 3, and the difficulties reported in Experiment 4 were less likely with increasing age. Our experiments offer a significant advance in our understanding of the factors that influence young children's sentence production, demonstrating that difficulties are more likely attributed to weak language knowledge rather than limited memory capacity, thus supporting a language-based account of sentence production (Van Dyke et al., 2014).

We contrasted two opposing accounts for why some sentence structures are more difficult than others. Both accounts predict that reverse order sentences would be more difficult to produce accurately than chronological order sentences, in line with our findings. A memory capacity account (King & Just, 1991) explains this finding on the basis that such sentences require more information to be held active in working memory. Additional support for this account would be demonstrated by improved prediction of performance when an independent measure of memory capacity was fitted to the data. We did not find such an effect. The contrasting language-based account proposes that the effects of working memory are not direct, but rather the result of its relation with language knowledge (Van Dyke et al., 2014; see also Klem et al., 2015). According to this viewpoint, the ability to represent information accurately in short-term memory (a requirement for good performance on a sentence production task) is influenced by the quality of language knowledge. We found support for this account in several ways: a difficulty with reverse order sentences was limited to the connective *after,* and an independent measure of language ability explained performance over and above our independent measure of memory. Further, our error analysis in Experiment 3 demonstrated that a poor representation of our target connectives was significantly more likely to affect the production of reverse order than chronological order sentences.

Although the main prediction of language-based account, a difficulty for after-reverse sentences, was replicated across both our experiments, there are at least two reasons to remain cautious about accepting this explanation for children's difficulties with sentence production. First, although we found a significant interaction between connective and order in both experiments, the main effect of connective type was significant only in Experiment 4. A language-based account would predict this main effect, so we must consider the possibility that order effects are modulated by a confounding variable, connective position, rather than connective type. Second, we must address why the stronger influence of vocabulary over memory (determined by examining model fit) was apparent only in Experiment 4. These limitations are considered, in turn, below.

A natural consequence of our design was that the interaction between connective type and order was influenced by the fact that position of the connective also differs across sentence structures. For example, *after* is used in a sentence initial position when the order of events is presented in chronological order (*After he put on the socks, he ate the burger*), but is used in a sentence medial position when events are presented in reverse order (*He ate the burger, after he put on the socks*). The reverse applies to *before* sentences. Thus, an alternative explanation for a specific difficulty with reverse order sentences is that position of the connective modulates the effects of order. That is, a reverse order sentence in which the temporal sequence is cued by *before* may be easier to represent than its *after* counterpart, because the initial position of the connective signals from the beginning that events will be narrated in a reverse order. This viewpoint is supported by evidence that speakers have cognitive biases to highlight certain referents at the beginning of the sentence that act as cues to reduce ambiguity for the listener, in our case the temporal connective (e.g., Chafe, 1984; Grice, 1975; Myachykov, Garrod, & Scheepers, 2012; Silva, 1991). Conversely, reverse sentences that contain *after* may be more difficult to plan and narrate because the critical information about event order is provided midway through the sentence, which may place greater demands on representation in memory.

We believe that this account (that connective position rather than connective type modulates order effects) does not adequately explain our pattern of findings. First, if position accounted for our results, we would have expected that the difficulty for after-reverse sentences would arise because the late signalling of reverse order places greater demands on memory capacity. However, we found that vocabulary was a stronger predictor of performance than memory. Moreover, as cited in the Introduction, corpus work suggests that speakers have a preference for relating information using the connective in a medial position

(Diessel, 2005). Clearly, more experimental work is needed to investigate the role of connective position on sentence processing.

The second reason for caution in fully accepting a language-based account was that in Experiment 3, our independent measures of memory and vocabulary did not explain unique variance in children's specific difficulty with after-reverse sentences. This may be because sentence repetition (Experiment 3) requires the child to comprehend the target sentence before they plan and produce it, whereas elicited production (Experiment 4) does not. Therefore, the extra requirements of sentence comprehension and memory in sentence repetition tasks may make it an insensitive method to differentiate memory capacity and language-based accounts of children's sentence production, for the following reasons.

Production and comprehension share similar cognitive processes such that both involve the prediction of upcoming language. The processing of the language itself is largely determined by accessibility to knowledge of sentence structure and words (Pickering & Garrod, 2013). However, the retrieval process should be less easily disrupted in comprehension compared to production because the comprehender has already been given the appropriate structure and words, so their knowledge should be more easily activated and therefore accessible. Conversely, a speaker is required to search for the appropriate target structures and words as well as retrieve their meaning from memory (Gennari & MacDonald, 2009). Therefore, one can argue that, despite similar cognitive processes across the two domains, the effect of memory and processing demands on children's performance is influenced by language knowledge in an elicited production task than one that involves comprehension.

This explanation may also explain why the pattern of findings across sentence structures differs to that found in a recent paper using an analogous touch-screen comprehension study (Blything et al., 2015). Blything et al. reported a similar pattern of results to the present study for the same age groups (3- to 6-year-olds): reverse order sentences that contained *after* were the most difficult. However, in contrast to the findings of this production study, the difficulty for after-reverse sentences was not statistically significant in comprehension (Experiments 1 and 2); instead it was the advantage for before-chronological order sentences that drove the effect. Further, in the comprehension study, an independent measure of memory accounted for significant variance in performance, whereas an independent measure of vocabulary did not, in contrast to the current findings. Of course, these differences may be spurious. However, our replication for a difficulty with after-reverse sentences, in addition to stronger effects of vocabulary over memory capacity, suggests a different explanation is required to that used for comprehension.

Of course, it may be that production tasks are simply more demanding than forced-choice comprehension tasks (Pickering & Garrod, 2013). The touch-screen comprehension task used by Blything et al. (2015) minimised the processing memory load of the task itself because once children had decided on their response, they were required only to select their response by touching the target item on a computer monitor, from a choice of two visual representations. Both of our production tasks called on additional memory processing resources: as described earlier, once the child decided on an answer, there were additional demands involving utterance planning and formulation. Therefore, it is possible that children with weaker memory skills did not display benefits from the low memory load of before-chronological sentences because their working memory capacity was already limited by task demands. A simple way to test this would be to investigate whether the touch-screen comprehension results (Blything et al., 2015) are replicated under a series of conditions that increase task working memory load (e.g., increasing sentence length by adding additional words or clauses).

Limitations, implications, and future research

A strength of our work was the replication of the main finding across two different tasks: children up to 6 years of age had difficulties in producing reverse order sentences linked by the connective *after*. However, the error analysis highlighted the differences in the nature of our two experiments, which we believe is informative for researchers whom are considering a marriage of the two paradigms. First, incomplete errors contributed to 35% of the total errors that we analysed in the elicited production task (Experiment 4), compared to only 0.7% in the sentence repetition task (Experiment 3). This may be because a sentence repetition task may provide more scaffolding for the child and is an easier task to tap children's production skills. Another notable difference between the experiments was that sense maintained errors made up the largest percentage in elicited production (Experiment 4), whereas sense changed errors contributed to a large percentage of errors in the sentence repetition (Experiment 3). In sense-maintained errors, the children produced a temporal connective as a linguistic device to successfully communicate order, but did not use the target structure. This indicates that the elicited production paradigm, which elicited a high number of sense maintained errors, is more likely to lead children to revert back to a sentence structure that they are familiar with, when required to signal temporal order with a connective. Overall, the difference in error types, along with the differences in the nature of the tasks themselves, highlights that replication across both paradigms is highly corroborative for conclusions in the production domain.

Age differences in Experiment 4 persisted even when memory and vocabulary were incorporated in the model. This finding shows that the ability to produce these two-clause temporal sentences increases with age, and that the influence of language-knowledge was less likely in older children. The latter finding, coupled with ceiling performance and non-significant speech onset times by 5- to 6-year-olds in Experiment 3, motivates the need to conduct assessments of online processing during both comprehension and production. Therefore, sensitive measures of online processing are needed to reveal more subtle individual differences in the efficiency of incremental planning and production. A previous adult online study using fMRI demonstrated a difficulty for producing reverse order compared to chronological sentences (Ye et al., 2011). However, that study only examined the effect of order (not connective) on processing, so a future study using an online measure suitable for children is needed to extend that finding to children, and to use a full factorial design of all four sentence structures. In addition, another advantage of an online measure is that it would provide a more sensitive means to examine the connection between the two domains, and whether differences in results are due to differences in task demands or not.

Our two production experiments lend support to recent arguments for a language-based account of young children's sentence processing (Kidd, 2013; Klem et al, 2015; Van Dyke et al., 2014). As shown in both experiments, language knowledge influences the quality of the representation of information in verbal working memory, which in turn effects sentence processing. Therefore, children's representation during their planning and production of reverse order sentences was more fragile when linked by *after*. One critical implication is that a memory capacity account of sentence processing (King & Just, 1991) is likely too simplistic on its own and needs to factor in the influence of the specificity or distinctness of retrieval cues (i.e., language knowledge). Indeed, converging evidence has been provided in other areas of language development such as inference generation, which have reported that the effects of memory are indirect and modulated through vocabulary knowledge (e.g., Currie & Cain, 2015).

A next question for the language-based account regards what an individual has to acquire to consolidate their language knowledge. The central aspect of the account is that processing difficulties occur when the quality of lexical representations is less precise and robust. It follows that more precise lexical representations lead to superior language skills (Perfetti, 2007; Van Dyke & Johns, 2012). A straightforward assumption from a developmental perspective is that language representations become stronger as a result of frequency of occurrence in language (Wells, Christiansen, Race, Acheson, & MacDonald, 2009; Van Dyke & Shankweiler, 2014). Future corpus work would be beneficial for investigating whether some referential cues have more probabilistic regularities than others (e.g., *before* versus *after*), and whether regular exposure to those cues can reduce susceptibility to interference.

On a final note, it would be hoped that the present experiments encourage future studies to investigate the commonalities between production and comprehension, namely examining the role of memory and language across both domains. There is little research on the role of memory in sentence production (Hartsuiker & Barkhuysen, 2006), in contrast to the debate about the role of memory in sentence repetition and recall tasks (e.g., Klem et al., 2015). A likely reason for this is that production is inherently more difficult to study than simple repetition, because it is difficult for the experimenter to elicit constructions that the speaker may typically opt out of using. We believe that the present study highlights the advantages of using sentence repetition and blocked elicited production paradigms together in order to restrict speakers to use specific target sentence structures (Ambridge & Rowland, 2013). Most impressively, both experiments complement each other by using different ways to elicit speech whilst replicating the main finding of children's difficulty in producing reverse order sentences linked by the connective *after*.

In conclusion, 3- to 6-year-olds showed a significant difficulty in producing reverse order sentences that were linked by the temporal connective *after*. Our findings from independent measures of vocabulary and memory indicated that language knowledge, not working memory capacity per se, modulated these effects. Further experimental work is needed to understand how these factors influence the language retrieval processes involved in sentence planning and production, and to elucidate the commonalities and differences in their influence on language production and comprehension.

5. General Discussion

The general aim of this thesis was to investigate young children's understanding of the temporal connectives *before* and *after,* and the factors that influence performance. Previous research shows that usage of these connectives develops over a significant period of childhood: temporal connectives appear in children's speech from around 3 years of age (Diessel, 2004), but they are not fully understood (Cain & Nash, 2011; Pyykkönen & Järvikivi, 2012) or produced accurately (Peterson & McCabe, 1987) until early adolescence. To examine the development of connective competence, the experimental work aimed to (i) determine the age that children display an early appreciation for the meaning of *before* and *after* for understanding and producing two-clause sentences with a sequential temporal order of events; (ii) identify the reasons why children display difficulties with *before* and *after* in the immediate years that follow an appreciation for their meaning; and (iii) investigate whether the pattern of difficulty for different sentence structures is the same or different for comprehension and production. In this final chapter, I discuss the findings and the theoretical implications from my experimental work in relation to these three research aims. I then discuss the methodological and educational implications and, in doing so, I outline the motivation for future research. This is followed by my conclusions.

The experimental work was designed to contrast how two critical factors, working memory resources and language knowledge, influence children's comprehension and production of two-clause sentences linked by *before* and *after* using a cross-sectional developmental design. In addition to manipulating the memory and language demands of the experimental sentence conditions, performance in each experiment was directly related to independent measures of working memory and vocabulary.

Experiment 1 (Chapter 2) and Experiment 2 (Chapter 3) examined comprehension. To do this, 3- to 7-year-olds' understanding of *before* and *after* was investigated using a forced-choice touch-screen paradigm, designed to simplify the demands of the task. For each sentence, the child viewed two short animations depicting the two events and heard the sentence. The two events were then shown as two static clips from the animation. I manipulated whether the presentation order matched the chronological order of events or not, and also whether the temporal order was arbitrary or predictable from world knowledge. Children's understanding of temporal connectives was assessed by asking them to touch the picture showing what the character did first, rather than a verbal response or an acting out task, both used in previous research (e.g., Clark, 1971; Trosborg, 1982). In this way, task demands were minimised so that performance more directly reflected understanding of the sentence, rather than planning or production of a response. To replicate the findings of Experiment 1 (Chapter 2), the second experiment (Chapter 3) required children to identify what happened last. To extend the findings of Experiment 1, response times were recorded to indicate how efficiently children processed the different sentence structures. Training and practice were provided to encourage speeded responses.

Experiments 3 and 4 (Chapter 4) investigated 3- to 6-year-olds' production of two-clause sentences linked by *before* or *after*. In these two experiments, children viewed an animated sequence of two actions, and were asked to describe the order of events. Instructions and practice trials were used to model target sentence structures, which were analogous to those assessed in the comprehension experiments. In a sentence repetition task (Experiment 3), the participant heard a target sentence and was asked to repeat it. In an elicited production task (Experiment 4), four blocked conditions were each preceded by a training phase in which children were instructed to use one of the four target sentence structures.

5.1. Aim 1: When do children display an early appreciation for the meaning of *before* and *after*?

For comprehension, both experiments (Chapters 2 and 3) demonstrated that 3- to 4-year-olds performed at above chance levels in their comprehension of two-clause sentences containing *before* and *after*. This finding indicates an early understanding of *before* and *after*, which is at a younger age than reported in previous research (e.g., Clark, 1971). However, the results of both experiments indicated that the 3- to 4-year-olds used an order of mention strategy, because their overall performance was higher for chronological compared to reverse order sentences. Thus, it cannot be said that these children truly understood the meaning of the connective; rather their response pattern suggests that they defaulted to an expectation that language order maps onto the actual order of events (Clark, 1971). Further, in Experiment 2 where the instructions emphasised a speeded response, 4- to 5-year-olds also showed this response pattern. This differs from Experiment 1, where an order of mention strategy was only evident for 3- to 4-year-olds. I provide reasons for why the findings in these two experiments differ in a discussion of methodology (Section 5.4). Together, the findings of these two experiments suggest that an initial understanding for the temporal order specified by *before* and *after* is not apparent until at least 4 to 5 years of age. Note that these conclusions are not relevant to sentence processing accounts (e.g., Just & Carpenter, 1992; Van Dyke et al., 2014), which focus on a more fine-grained understanding of the connective in the immediate years that follow an initial understanding for its meaning.

The experimental work was also able to determine whether performance was enhanced when the sequence of events could be interpreted by world knowledge. Previous studies have reported better performance when materials conform to the likely order of events in the real world (French & Brown, 1977; Keller-Cohen, 1987; Trosborg, 1982). That finding was not replicated in the experimental work in this thesis. Of note, Experiment 1 found that when the order of events was typical and predictable from world knowledge (e.g., *He put on the socks, before he put on the shoes*), performance was not enhanced relative to when event order was arbitrary and not supported by world knowledge (e.g., *He put on the socks, before he ate the burger*). This finding was replicated in both production experiments (Experiments 3 and 4). Note that the world knowledge manipulation was not included in Experiment 2. The replication of the absence of an effect of world knowledge when included as an experimental manipulation, with different samples of children, strongly suggests that this finding is robust and an accurate reflection of sources of influence on young children's sentence processing.

Early competence with temporal connectives was also demonstrated in two production studies (Experiments 3 and 4, Chapter 4). That is, just as 3- to 4-year-olds performed above chance in comprehension tasks (Experiments 1 and 2), that age group also displayed an early form of appreciation for the meaning of the connective by frequently employing *before* and *after* as a linguistic device to accurately signal temporal order. The two production experiments reported in this thesis are the first to investigate production across each of the four sentence structures. Replication of this main finding in two experiments with different samples of children suggests that this finding is robust.

Unlike comprehension, in production there is no certain method for identifying the age at which children use information other than that provided by the connective (such as world knowledge). Specifically, in forced-choice comprehension tasks, children can be accurate for some sentences without using the information provided by the connective, for example, an order of mention strategy results in choosing the correct response for chronological sentences. Conversely, in the production tasks, children who do not have robust working knowledge of the connective will be inaccurate at all sentence structures, because the materials presented do not provide the target connective so the child might get the order correct but still produce a non-target response (e.g., *He ate the burger. He put on the sandals*).

None of the age groups (aged 3 to 6) displayed full competence at producing two-clause sentences containing *before* and *after,* that is none were at ceiling levels of performance. This is consistent with previous research which has suggested that children have difficulties in producing two-clause sentences up until at least 9 years of age (Peterson & McCabe, 1987). Therefore, as in comprehension, consideration of the processing demands of each sentence structure is needed in order to investigate how and when competence develops in the use of temporal connectives to signal temporal order.

5.2. Aim 2: The influence of memory capacity vs. language knowledge on sentence processing

Experiments 1 to 4 demonstrated that in the immediate years that follow an early appreciation for the meaning of the connective, performance was not uniform across sentence structures. One possibility is that performance on some structures was poor because of differences in the processing demands of those sentence structures. As noted, the design of the experimental work in this thesis contrasted the role of working memory resources versus language knowledge on processing. This enabled a test of whether processing difficulties were best explained by a memory capacity account (e.g., Just & Carpenter, 1992) or a language-based account (e.g., Van Dyke et al., 2014). A memory capacity account posits that performance will be driven by an individual’s working memory capacity and that processing ease will be influenced by the amount of information that must be held in working memory for a specific sentence structure. A language-based account posits that an individual’s performance is not dependent on their working memory capacity *per se*, rather, it is dependent on the support for an accurate representation of information in working memory that is provided by language knowledge. These accounts are discussed in turn.

5.2.1. Evidence in support of memory capacity accounts

In the two comprehension experiments (1 and 2), children aged 5 years and older were above chance for each sentence structure, so demonstrated an appreciation for the meaning of the connective. However, these children were typically most accurate (Experiment 1) and fast at processing (Experiment 2) sentences in a chronological order linked medially by before (e.g., *He put on the sandals before he ate the burger*) compared to the other sentence types. This superior performance for before-chronological sentences has been explained in relation to the additional demands that are made on working memory when a sentence has a reverse order (Ye et al., 2012), a more difficult vocabulary (Gunter et al., 2003), or an initial position for the connective (Diessel, 2004). Before-chronological sentences carry none of these additional demands, and afford incremental processing. Conversely, before-reverse (reverse order, initial position), after-chronological (initial position, difficult connective), and after-reverse (reverse order, difficult connective) sentences each carry two of these additional demands, so require the comprehender (or speaker) to maintain more information in working memory whilst constructing their mental representation (see Table 1.2, relevant section reproduced here for ease of reference).

Table 1.2.

Sentence structures and their additional working memory load as influenced by a reverse order of mention of events, a later acquired connective, and an initial position of the connective.

	Additional working memory load		
	Reverse order	Later acquired connective	Initial position
Before-chronological	No	No	No
After-chronological	No	**Yes**	**Yes**
Before-reverse	**Yes**	No	**Yes**
After-reverse	**Yes**	**Yes**	No

Critically, in both comprehension experiments, the advantage reported for before-chronological sentences was predicted by performance on an independent measure of working memory. Therefore, the overall findings for comprehension support a memory capacity account (Just & Carpenter, 1992), such that children's comprehension of two-clause sentences containing *before* and *after* is influenced by an individual's memory capacity resulting in poorer performance for sentence structures that carry additional memory load (Just & Carpenter, 1992; Pyykkönen & Järvikivi, 2012; Ye et al., 2012).

In contrast to the comprehension experiments, the production studies (Experiments 3 and 4) did not provide support for a memory capacity account: children did not display an advantage for before-chronological sentences, and the independent measure of working memory did not predict unique variance in performance. The production data support the predictions of a language-based account (e.g., Van Dyke et al., 2014), discussed further below.

5.2.2. Evidence in support of language-based accounts

Although the comprehension experiments provide strong support for a memory capacity account (Just & Carpenter, 1992), some of those findings can be explained by language knowledge. For example, the advantage for chronological sentences displayed by the younger children, 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. Also, a general advantage found for *before* in both comprehension experiments, supports a simple form of a language-based account, such that language knowledge (of the connective) directly influences ease of processing.

As detailed in the literature review (Chapter 1), a language-based account that acknowledges the influence of memory proposes that the relation between memory and sentence processing is indirect and that, rather, good language skills modulate the accurate representation of information in verbal working memory (e.g., Van Dyke et al., 2014). According to this account, working memory should not explain unique variance in performance after individual differences in language knowledge have been taken into account (Klem et al., 2015). The comprehension experiments did not support these predictions: in Experiments 1 and 2, an independent measure of working memory explained unique variance in performance over and above an independent measure of language knowledge.

In contrast to comprehension, the production experiments (Experiments 3 and 4) support a language-based account of the role of working memory in sentence processing (Van Dyke et al., 2014). In both a sentence repetition task (Experiment 3) and an elicited production task (Experiment 4), children were significantly less accurate at producing target sentences when the presentation order of the two clauses did not match the chronological order of events, but this difficulty was specific to target sentences linked by *after*. Further, in elicited production, vocabulary explained unique variance in this pattern of performance. Critically, as noted in Section 5.2.1, an independent measure of working memory did not explain unique variance in either of the production experiments. Together these findings suggest that language knowledge modulated children's ability to cope with the cognitive demands of producing reverse order sentences.

5.2.3. Does position of connective provide an adequate explanation for the difficulty in producing after-reverse sentences?

It should be noted that an additional influence on children's specific difficulty in producing reverse order sentences linked by *after* (Experiments 3 and 4)*,* may be the position of the connective. In reverse order sentences linked by *after,* the connective appears in the medial position (*He ate the hotdog, after he put on the jumper*); when linked by *before,* the connective appears in the initial position (*Before he ate the hotdog, he put on the jumper*)*.* Although a medial position is generally proposed to carry lower demands on working memory than an initial position (Diessel, 2005), speakers may have a pragmatic preference to place the connective in the initial position in order to foreground critical information that can guide the listener's understanding (see Chafe, 1984; Gernsbacher, 1997; Junge, Theakston, & Lieven, 2013; Myachykov, Garrod, & Scheepers, 2012; Silva, 1991). Therefore, there may be two reasons why a reverse order sentence presents difficulties only in production when it is linked by *after*. First, as argued by a language-based account, *after* is the later-acquired connective and is used less consistently as a connective than *before.* Second, a reverse order of events signalled by *after* is difficult to plan and produce because the speaker only provides the critical information that the events occur in reverse order, part way through the sentence.

This explanation fits the findings in Experiment 1 and 2 that children are most accurate at comprehending and processing sentences that fully afford incremental processing.

However, it does not account for the pattern of results in Experiments 3 and 4 as adequately as the explanation that is provided by the language-based account in section 5.2.2. The first reason for this was that corpus and experimental work suggest positioning the connective in the medial position enables incremental processing (Cain & Nash, 2011; Diessel, 2004). As a result, medially placed connectives in after-reverse sentences should not be problematic for the speaker (or comprehender). Second, if the late (medial) signalling of reverse order in *after* sentences is the true explanation for the results, it should tax working memory resources because more information has to be held in working memory while the mental representation is revised. However, as noted, working memory did not explain unique variance over and above language knowledge. I return to the issue of determining the influence of connective position on sentence processing in Section 5.3 and in Section 5.4.2.

5.3. Aim 3: Comprehension versus production

In this work, a number of commonalities, as well as differences, were apparent for children's comprehension versus production of two-clause sentences containing *before* and *after*. The most notable commonality was that 3- to 4-year-olds displayed an early competence in understanding and producing these sentences, and this was followed by a period of development extending (at least) 3 years in which performance was limited either by working memory capacity or language knowledge. The relation between task performance and the independent measures of memory and language were the most salient differences between comprehension and production. Working memory directly influenced comprehension, but did not directly influence production; instead language had a direct influence on production. As noted, this means that the comprehension data was best explained by a memory capacity account (Just & Carpenter, 1992), whereas the production data was best explained by a language-based account (e.g., Van Dyke et al., 2014). This section will explore the potential reasons for those differences.

CHAPTER 5: GENERAL DISCUSSION.

In the literature review, I described the general consensus that comprehension and production draw on similar cognitive processes (Pickering & Garrod, 2013). However, the two domains differ by their inherent input and output: comprehenders must map language form onto meaning, whereas speakers must map meaning onto language form (Grimm, Müller, Hamann, & Ruigendijk, 2011; Pickering & Garrod, 2013; Vigliocco & Hartsuiker, 2002). More specifically, comprehenders are provided with the appropriate language form and structure through input, and are required to create a meaningful mental representation of the state of affairs and events described. In contrast, speakers begin with a mental representation and are required to find the appropriate language form and structure in order to convey that meaning. Therefore, there may be greater demands on language knowledge in production compared to comprehension tasks: comprehenders receive input on the details of the form so only have to use their language knowledge to work out an interpretation of its meaning, whereas speakers must use their language knowledge to specify every level of detail of the form themselves (i.e., syntactic, morphological, phonological, and articulatory), so that it is mapped onto the intended meaning (see Garrett, 1980; Gennari & MacDonald, 2009; Vigliocco & Hartsuiker, 2002).

Support for the above explanation of the differences in comprehension versus production, can be provided by everyday life encounters of ‘tip of the tongue’ states. These states are a product of disruption in the retrieval processes that are involved in constructing language form (i.e. production), and occur more often than their comprehension counterpart which involves a difficulty in accessing the meaning of a word that has already been provided in form (Brown, 1991). This indicates that retrieval operations may be more likely to be disrupted in production compared to comprehension tasks, so the former should benefit most from high quality language representations.

Further support for this explanation can be provided by a computational acquisition model of how a child learns to comprehend and produce language (Chater, McCauley & Christiansen, 2016). The Chunk-based learner (CBL) model draws on Pickering and Garrod's (2013) model of adult language processing, that comprehension and production both rely entirely on the ability to process language, which is viewed as a unified skill across each domain. Critically, CBL is able to store chunks (i.e., one or more words) of distributional information and linguistic units that are learned from input. This inventory of learned chunks enables rapid incremental processing because the chunks can be activated 'just in time' so that they are accurately held in memory for when new material must be comprehended or produced.[5]

Despite using the same processes for each domain, the CBL exhibits a developmental lag in its ability to produce sets of randomly selected test utterances, relative to comprehension. This resembles the comprehension-production asymmetry that is typically displayed by young children across different aspects of language (Benedict, 1979; Fraser, Bellugi, & Brown, 1963). Chater et al. (2016) argue that these asymmetries are a result of the differences between the nature of the two tasks (i.e., input and goals). Specifically, the rudimentary understanding of grammatical constructions that is provided through chunking, allows the child (or CBL) to form a 'good enough' interpretation of an utterance to give the appearance that they are fully representing it in a comprehension task. This is commonly known as *shallow parsing*, whereby the final representation of the sentence is underspecified but is nevertheless a plausible reflection of the arguments and their verbs (see Ferreira, Bailey & Ferraro, 2002; Karimi & Ferreira, 2016). It applies equally that when the child (or CBL) is unable to model a fully detailed and accurate representation in sentence comprehension, they

[5] Such a framework overcomes the risk of the decay of an accumulation of existing representations in favour of the incoming new material. Note that this means that the framework can operate under a severely limited working memory capacity (i.e., it is unable to store the entire information of a complex sentence), thereby corresponding more to a language-based account (Van Dyke et al., 2014) than a memory capacity account (Just & Carpenter, 1992).

cannot appropriately retrieve and sequence chunks for production. Critically, due to the nature of production tasks, which require speakers to 'get the details of the form right in every instance' (Garrett, 1980, p. 216), children will display a developmental lag compared to their performance in comprehension tasks, despite both domains drawing on the same chunks of distributional information and linguistic units that were learned from input. Therefore, the asymmetry should effectively disappear once grammatical constructions are mastered in full. That is, the asymmetry should be overcome with sufficient linguistic experience, as predicted by a language-based account (Van Dyke et al., 2014).

5.3.1. An integrated framework for comprehension and production

An integrated framework is needed to explain the findings across the four experiments in this thesis. In each experiment, children performed better with chronological compared to reverse order sentences, and, before-chronological sentences were typically performed best whereas after-reverse sentences were typically performed worst. For example, although not statistically significant, the pattern of comprehension accuracy in Experiment 1 was strikingly similar to the pattern of results in both production experiments (Experiments 3 and 4): reverse order sentences that contained *after* were notably the most difficult. Similarly, although not statistically significant, the production experiments showed strong performance for the same sentences that comprehenders found easiest (before-chronological).

It is important to note that these two sentences (before-chronological, after-reverse) were never directly compared in statistical analyses. To follow this up, additional by-age group GLMM analyses are reported in the Appendix (Table A.5), in which the medial position is held constant and just these two sentence types compared. Each age group performed more poorly on the after-reverse sentences and the difference reached statistical significance for at least one age group in each experiment. This suggests a degree of commonality across comprehension and production.

It is also worth note that, since these two sentences were both compared in the experimental work to sentences with a connective in the initial position (i.e., before-chronological vs. before reverse; after-reverse vs. after-chronological,), it is possible that the different findings for the two domains are due to position effects, rather than order effects. That is, children may have performed differently in their comprehension compared to their production of sentences with a connective in the initial position (i.e., after-chronological, before reverse). For comprehension, sentences linked by the connective in an initial position (before-reverse, after-chronological) were both performed equivalently to after-reverse sentences, which has the connective in the medial position. In contrast, for production, these sentences were both performed equivalently to before-chronological sentences, which has the connective in the medial position. This difference may be linked to the possible pragmatic preferences outlined earlier for using an initial position specifically in speech production, so further research on sentence position is clearly needed (see Section 5.4.2).

Another common finding across the experiments is that, when considered alone, the independent measure of working memory explained variance across sentence structures in Experiment 1 (comprehension accuracy), Experiment 2 (comprehension response times), and Experiment 4 (elicited production). However, the influence of memory across sentence structures in production was not significant in Experiment 3 (sentence repetition), and was not significant over and above language knowledge in Experiment 4. This finding is supported by a language-based account (e.g., McElree, 2006; Van Dyke et al., 2014), which argues that the specificity or distinctness of retrieval cues in the text, rather than the number of individual text elements that must be held active in memory (e.g., Just & Carpenter, 1992), can account for why some sentences are more difficult to process than others. Importantly, these two apparently contrasting accounts do have a common core: they seek to explain why memory limitations effect sentence processing. Of course, an alternative explanation for these differences is that they arise because of methodological differences between the comprehension and production tasks. This is explored in the next section.

5.4. Methodological implications

This work has highlighted emerging competence for *before* and *after*. Specifically, children display a long period of development, throughout which competence most notably varies by accuracy (e.g., demonstration of an appreciation for the meaning of the connective versus full range of sentence processing skills), sentence structure (e.g., competence in one structure does not mean competence in another), and the domain of language use (e.g., comprehension versus production). An implication for the study of language competence in general, is that researchers should not define accuracy based on a single choice of definition or measurement, but instead should take advantage of a broad range of measures that can together explicate the multifaceted skills involved in language competence. In this section, I will highlight the advantages of the methodologies used in my comprehension and production experiments, and will also suggest methodological improvements for future research.

5.4.1. Comprehension

The touch-screen paradigm used in this work was intended to minimise the task processing load compared to previous comprehension tasks (e.g. Clark, 1971). Children were required only to select their response by touching the target item on a computer monitor from a choice of two visual representations (e.g., Friend & Kelpinger, 2008) rather than having to produce a response that would require additional processing such as an act-out task (e.g., Clark, 1971) or verbal response (Trosborg, 1982). This technique has revealed an earlier competence than has been reported previously for children’s understanding of *before* and *after,* exemplified by the finding in Experiment 1 that 3- to 4-year-olds were above chance overall and 6- to 7-year-olds were at ceiling.

Another advantage of the touch-screen methodology is that it enables speed of response to be measured. Critically, in Experiment 2, response times mapped onto the pattern of findings for accuracy: children took longer to respond to sentence structures for which they demonstrated poor accuracy both in this thesis (Experiment 1 and Experiment 2) and in other work (Pyykkönen & Järvikivi, 2012). Overall, this informs future research that response times in touch-screen paradigms can be used to identify difficult sentence structures that require extra time to represent mentally (also see Möhring et al., 2014).

Surprisingly, the findings of Experiment 2 indicated that an order of mention strategy was used by 4- to 5-year-olds, which contrasts with Experiment 1 where this immature strategy was evident only for 3- to 4-year-olds. This difference is attributed to the instructions for children to (i) produce speeded responses, and (ii) to select ‘What happened last.’ Whilst speeded instructions were necessary for measuring the processing ease of sentence comprehension, a unique aspect of Experiment 2 was that these instructions did not allow extra time for the comprehender to reflect on the representation that they had constructed and stored in memory (see Marinis, 2010). As a result, the children in Experiment 2 who did not fully comprehend the meaning of the text during initial parsing, may not have engaged in post-presentation processing, because of the need to make fast responses. This task requirement may make them more likely to default to a non-linguistic strategy.

An additional reason for this difference between experiments could be attributed to the different questions: ‘What happened first?’ in Experiment 1 versus ‘What happened last?’ in Experiment 2. In these experiments, children may have been more likely to select the last mentioned event because it was the most recent. This maps onto the correct answer for chronological sentences in Experiment 2, but onto the correct answer for reverse order sentences when children are asked ‘What happened first?’ (Experiment 1). This would boost the advantage for chronological over reverse order sentences in Experiment 2, thus providing a reason for why it reported that 4- to 5-year-olds showed evidence of an order of mention strategy and Experiment 1 did not.

Experiment 2 is the first study that we know of to provide a timing measure that indicates how efficiently children process two-clause sentences containing *before* and *after*. The time to touch the screen was chosen as a processing time measure because it provided the same method for measuring accuracy that was used in Experiment 1 and provided an opportunity to test the replicability of those findings. However, the finding that children were fastest at processing before-chronological sentences, in turn motivates the need to gain a full picture by using more sensitive measurements of sentence processing as it happens. For example, measurement of processing ease in critical regions of two-clause sentences would remove the need for a behavioural response, and might provide insight into difficulty when it occurs.

Adult studies that have used ERP and fMRI to study sentence processing suggest that greater memory processing resources are required for reverse order sentences compared to chronological sentences (Münte et al., 1998; Ye et al., 2012). These measures reflect processing costs (e.g., semantic or syntactic violations) by displaying the changes in the electric potentials (ERP) or in blood oxygenation levels (fMRI) that occur during particular regions of the sentence (Kuhl & Rivera-Gaxiola, 2008). ERP is typically considered more child friendly than fMRI because the latter requires children to remain perfectly still, and because the loud MRI sounds are considered distracting (Dehaene-Lambertz, Dehaene, & Hertz-Pannier, 2002). In an ERP study, similar findings to those reported by Münte et al. (1998) would be expected with children, such that ERP components should elicit a progressively larger negativity over left frontal regions (associated with memory processing resources) for reverse compared to chronological sentences. However, adult studies to date (including the fMRI study by Ye et al., 2011) have used stimuli in which the connective was presented only in the sentence initial position, such that connective (*before, after*) was confounded with event order. From a developmental perspective, a full factorial design of all four sentence structures is essential, as children display developmental differences in their understanding of *before* and *after* (Clark, 1971).

More sensitive timing measurements of sentence processing can also be used to determine how early young children (e.g., 3- to 4-year-olds) are able to use temporal connectives to support discourse comprehension. An option would be a visual world paradigm, which presents children with visual stimuli on a screen and uses eye tracking to examine where the child looks in relation to a narrated sentence (e.g., Nation, Marshall, & Altmann, 2003). Children would be informed that their task is to listen to spoken sentences and to judge whether a picture that follows the narration is of the first event that occurred in that sentence (yes/no). Following the sentence narration, an additional narration would be played: '*First, Tom did…*'. The narrations would play whilst the screen displays the two animations of the actions in each clause (counterbalanced for left and right screen sides). However, the additional narration ('*First, Tom did…*') would be followed by a picture on its own in the centre of the screen. On half the trials, this would be a picture of the thing that happened first (i.e., the correct picture) and in the other half of trials, this would be a picture of the thing that happened second (i.e., the incorrect picture). The children would be informed that the computer is supposed to show a picture of thing that happened first, but that sometimes it gets it wrong. They would be trained to make their judgements using 'yes' or 'no' buttons, which interfere less with eye tracking compared to verbal judgement.

Typically, visual world paradigms might examine where the child looks whilst listening to the sentence itself. However, due to the nature of the experimental sentences, eye gaze would not provide meaningful information during sentence presentation. For example, if narrated '*Tom ate the ice cream before he put on the boots*', the child is likely to look at the respective pictures upon immediate hearing of the nouns *ice cream* and *boots*. Instead, the additional narration '*First, Tom did…*' is an ambiguous cue phrase that is expected to focus gaze on the event that the child considers to have occurred first, therefore reflecting whether they understand the order signalled by the connective. If children understand the temporal connective, they would be expected to look immediately at the event that occurred first. The experiment would additionally measure whether this understanding, as demonstrated by eye gaze, is also demonstrated in the judgement task.

5.4.2. Production

As already noted, in comparison to comprehension, it is inherently more difficult to measure production of specific sentence structures because it is difficult for the experimenter to elicit specific target constructions (Ambridge & Rowland, 2013). My work (Experiments 3 and 4) highlighted the advantages of using sentence repetition (Experiment 3) and blocked elicited production (Experiment 4) paradigms to elicit the production of these structures in young children. Critically, these two paradigms provided converging evidence: the main finding of a greater difficulty in producing reverse order sentences linked by the connective *after,* was found in both experiments. It is hoped that this work inspires future studies of production so that research can continue bridging the gaps between measuring comprehension and production.

Surprisingly, vocabulary knowledge predicted the difficulty for after-reverse sentences in the elicited production task (Experiment 4) but not in the sentence repetition task (Experiment 3). One reason for this may be that sentence repetition requires the comprehension and storage of the narrated sentence prior to production, so those additional requirements may provide an inaccurate account of the role of memory and language skills in children's sentence production. In turn, this would support the explanation outlined earlier in

Section 5.3, that production tasks may place greater demands on language knowledge retrieval processes relative to comprehension.

An important benefit of eliciting children's production of the full target range of sentence structures is that it provides an opportunity to compare performance to comprehension. However, the differences found between comprehension and production might be because paradigms used to measure production are generally considered to have additional task-related demands compared to paradigms used to measure comprehension (Treiman, Clifton, Meyer, & Wurm, 2003). The touch-screen comprehension paradigm used in Experiments 1 and 2 required children to respond in a certain way, so they had a one in two chance of responding correctly. This minimises task-related demands, and also increases the likelihood of false positives. Conversely, the production paradigms were less limited in the range of possible answers than the comprehension paradigms, which means that false positives are less likely but that additional task-related demands are high. These additional demands may be too demanding for a child's limited processing resources. Therefore, it is possible that children with weak memory skills did not display benefits from the low working memory load of before-chronological sentences because their working memory capacity was already limited by the additional demands of the production paradigms. As outlined in Experiments 3 and 4, a simple way to test this would be to investigate whether the touch-screen comprehension results (Experiments 1 and 2) are replicated under a series of conditions that increase task working memory load (e.g., increasing sentence length by adding additional words or clauses, see Cain, 2007).

Future research should use production paradigms with simplified demands in an attempt to make a fairer comparison with comprehension. One option would be a structural alignment paradigm, in which participants take turns with an experimenter in describing a set of events (see Pickering & Branigan, 1998; Rowland, Chang, Ambridge, Pine & Lieven, 2012). Participants typically show a preference to align their choice of sentence structure with that of the experimenter (Bock, 1986). This paradigm could be used as an additional test for replication: support for the accuracy data in Experiments 3 and 4 would be provided with less frequent production of sentence structures that are associated with higher demands on language knowledge. Critically, this improves the ecological validity of the task because children do not receive explicit instructions in the training nor in the test phase. The disadvantage is that, as outlined in Chapter 1, less explicit instructions in production do not match the constrained experimental manipulations offered by comprehension tasks and fewer target sentence structures may be produced. Nevertheless, this task may provide the means to allow future research to investigate whether the difficulty in producing after-reverse sentences (driven by poor language knowledge) holds even when task-related demands are minimised.

In addition, more fine grained comparisons between comprehension and production would be enabled if timing measures for children's production are improved so that they are more sensitive to processing costs in critical regions. The measurement of onset times in the sentence repetition task (Experiment 3) did not predict performance across different sentence structures. It would be problematic to measure processing times once the child has begun her utterance (e.g., total time to produce the sentence), as young children often alter certain words in the sentence that are not part of the central interest of the study (e.g., *tomato sauce* instead of *ketchup*; see Lust et al., 1995). An alternative for providing a timing measure is to examine older children so that there is enough data where the precise words have been produced. This could potentially be extended into an fMRI study, a research tool that can be used for both comprehension and production (e.g., Ye and colleagues, 2011, 2012). For example, after-reverse sentences would be expected to be the greatest activator of brain areas such as the left medial frontal gyrus (associated with maintaining temporal order information in working memory; see Wager & Smith, 2003) and the left medial temporal gyrus (associated with the retrieval of language knowledge; see Indefrey & Levelt, 2004).

It is also worth note that the measurement of processing costs induced in critical regions of sentences for comprehension versus production may provide a better understanding of whether preferences for connective position differ across the two domains, as outlined earlier as an alternative reason for the difficulty in producing after-reverse sentences (see 5.2.3, and also 5.3.1).

5.4.3. Methodological improvements for future studies

The experimental work was designed to disentangle the role of memory and language in the acquisition of two-clause sentences containing *before* and *after*. As noted, the main difference between the findings for comprehension and production was that memory capacity was a stronger predictor of children's performance in comprehension than in production; whereas for the latter, language knowledge was the more significant influence. I have explained this asymmetry by drawing on Chater et al.'s (2016) framework that the two domains share the same cognitive processes but that they differ by the nature of the tasks themselves (i.e., different input and goals), which may place different demands on language retrieval processes. However, at this point it is important to recall that a key difference between the memory capacity (Just & Carpenter, 1992) and the language-based (Van Dyke et al., 2014) accounts is their viewpoint on the architecture of the memory systems: the former assumes that short term and long term memory are independent systems, whereas the latter assumes that they are unified. Critically, if we assume that comprehension and production share the same processes and mechanisms (Chater et al.; Pickering & Garrod, 2013), then they must share the same memory architecture. Therefore, as future research becomes more fine-grained in comparing comprehension and production, we should expect to find an overlap in terms of which sentence processing account best explains comprehension and production.

In addition to improving the experimental paradigms used to assess comprehension and production (see 5.4.1, 5.4.2), one way to follow this issue up is to use a more intensive battery of measures for individual differences in memory and vocabulary skills (Language and Reading Research Consortium, 2015). Ideally, working memory tasks should measure the storage *and* manipulation (processing) of information, thereby tapping the two critical functions of working memory (Oberauer & Lewandowsky, 2010). However, as noted, Gathercole et al. (2004) have reported that 5-year-olds find it too difficult to perform such complex span tasks. Specifically, the sentence processing demands of a listening recall task (Daneman & Carpenter, 1980), resequencing demands of numbers in a backward digit recall task (Morra, 1994), and dot counting in counting recall tasks (Case, Kurland, & Goldberg, 1982) led to floor performance by 4- and 5-year-olds. Nevertheless, to build upon the single measure of working memory storage in the present thesis (forward digit recall), a future study of older children (i.e., 6 years onward) should include these as multiple measures of working memory in order to have a more comprehensive assessment of this construct (Kidd, 2013).

It is also important to note that Daneman and Blennerhassett (1984) have previously demonstrated a version of the listening recall task that can be used with 3-year-olds. In their task, children are presented with sets that contain between 1 to 5 sentences, and must recall those sentences in verbatim. However, as with other listening recall tasks (e.g., Daneman & Carpenter, 1980), the verbal nature of this task means that if it displays any influence on comprehension, that can be attributed to the notion that it is simply tapping into the language knowledge that underlies language processing (Kidd, 2013; MacDonald & Christiansen, 2002). Therefore, digit based tasks can be advantageous because they are less strongly related to language processing ability in young children than outright verbal tasks (Cain, 2006;

Seigneuric, Ehrlich, Oakhill, & Yuill, 2000). For these reasons, the forward digit recall task was the most suitable single measure for the experimental work of the present thesis.

To build upon the measure of the breadth of vocabulary that was used in the present thesis, the depth of vocabulary knowledge should also be measured in order to take into account the richness of knowledge for particular words (Cain & Oakhill, 2014; Ouellette, 2006). Breadth and depth of word knowledge is important for testing the assumption of a language-based account that lexically rich representations support a more accurate representation of information in working memory (e.g., Van Dyke et al., 2014). Similarly, tests of grammatical knowledge (e.g., TROG; Bishop, 1983) should be included as a measurement of language knowledge, as rich grammar representations should influence the role of working memory on language processing in the same way that vocabulary is proposed to (Van Dyke et al., 2014). Note, however, that the measures used in the present thesis were strongly predictive of performance, and that these suggestions are for developing a more fine-grained picture of the influence of memory and language knowledge.

5.4.4. Corpus studies of spoken language

Of course, children's ability to understand and use *before* and *after* in two-clause sentences cannot be entirely explained by the inherent memory and language demands of the sentence structure itself (e.g., Just & Carpenter, 1992; Van Dyke et al., 2014). An alternative but not mutually exclusive explanation for poor performance would be a usage-based account (e.g., Tomasello, 2003). In its simplest form, this posits that the understanding and processing ease for specific sentence structures is influenced by their distributional frequencies in the input. The more children hear a specific sentence structure, the more able they are to understand an utterance that takes the form of that sentence structure.

The role of input in language acquisition has been well supported in a host of other sentence constructions that essentially provide another test bed for investigating the interplay of factors that influence complex sentence acquisition (e.g., Kidd, Brandt, Lieven & Tomasello, 2007; Goldberg, Casenhiser, & Sethuraman, 2004, Huttenlocher, Vasilyeva, Cymerman, & Levin, 2002; Theakston, Lieven, Pine, & Rowland, 2004). A notable example is the relative clause sentence, in which children sometimes find it more difficult to comprehend and produce object relative clauses (e.g., *This is the dog that the cat chased*) compared to subject relative clauses (e.g., *This is the dog that chased the cat*). This difference in performance has previously been explained using theories that focus on the inherent memory (e.g., Just & Carpenter, 1992) and language demands (e.g., Van Dyke et al., 2014) of the sentence structure itself (see Finney, Montgomery, Gillam, & Evans, 2014). However, Kidd et al. (2007) have shown that children's difficulty with relative clauses is also predicted by their frequency of occurrence in naturalistic speech.

Despite the previous research on the role of input frequency on the acquisition of various complex sentence structures, there has not yet been an empirical demonstration of its influence on the acquisition of two-clause sentences containing *before* and *after*. Therefore, an aim for future research should be to use corpus methods to analyse the frequency with which children hear these different sentence structures in various outlets, such as adult speech, children's books, and children's television programs. Only a comprehensive account which considers the role of input frequency in conjunction with the inherent memory and language demands of the sentence structure, can fully reveal how children learn to construct a coherent representation of the information provided in the discourse.

5.5. Educational implications

Skilled language users benefit from connectives by using them to understand and signal the relations between sentences (Cozijn, Noordman, & Vonk, 2011; Sanders & Noordman, 2000; Traxler, Bybee, & Pickering, 1997). The National Curriculum in England (DfE, 2014) outlines that teachers should first introduce temporal connectives (including *before* and *after*) at 7- to 8-years-old (Year 3). These are taught explicitly as grammatical concepts, and children are encouraged to note and explore their use in speech, writing, and book reading. However, the findings of this thesis show that children display competence with *before* and *after* earlier than previously thought, for example 3- to 4-year-olds can perform at above chance overall and 5- to 6-year-olds can perform at ceiling (Experiment 3). Therefore, given that children who understand connectives find it easier to integrate two-clause sentences when the clauses are linked by a connective (e.g., Cain & Nash, 2011), an earlier focus of connectives such as *before* and *after* in the curriculum may be advantageous for improving early educational attainment.

Teachers should also be made aware of the findings in this thesis that connectives such as *before* and *after* may be less easily understood when they are used in particular sentence structures. The first reason for this is that teachers may need to allocate extra time and instruction for teaching children to understand and use these connectives in more difficult sentence structures, such as reverse order sentences. The second reason is that classroom instructions which provide information about the order that things will happen, are more likely to be understood by 3- to 7-year-olds when the events are presented in easier sentence structures such as chronological order: '*You will have to finish your work before you go out to play*'.

5.6. Final conclusions

The research in this thesis provides an insight into how early children acquire an understanding of two-clause sentences containing *before* and *after*, so provides a fundamental insight into how young children construct temporal representations of a text's meaning. My findings indicate earlier competence than previously reported: children were above chance level from as young as 3- to 4-years-old (Experiment 1) and performed at ceiling as young as 5- to 6-years-old (Experiment 3). However, the 3- to 4-year-olds typically demonstrated poor knowledge of the distinction between the meanings of *before* and *after*, and tended to interpret the event order as the order of mention of events. Older children displayed an adequate understanding of these connectives, but the reason for their failure shifted to sentence processing limitations. The explanations for why children displayed these processing difficulties differed for comprehension versus production, but were replicated within each respective domain. Specifically, working memory capacity best predicted the processing difficulties in comprehension (Experiments 1 and 2) and language knowledge best predicted the processing difficulties in production (Experiments 3 and 4). This highlights the importance for language researchers to seek converging evidence across a variety of measurements before defining full competence and reaching final conclusions. A key aim for future theoretical and experimental work is to examine the interplay of the role of working memory capacity and language retrieval processes, along with other factors such as the distributional frequencies in the input, on sentence processing in order to elucidate the commonalities and differences in their influence on language comprehension and production.

6. 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

Ambridge, B. & Lieven, E.V.M. (2011) *Child language acquisition: Contrasting theoretical approaches*. Cambridge, UK: Cambridge University Press.

Ambridge, B., & Rowland, C. F. (2013). Experimental methods in studying child language acquisition. *Wiley Interdisciplinary Reviews: Cognitive Science, 4,* 149–168. doi: 10.1002/wcs.1215

Amidon, A., & Carey, P. (1972). Why five-year-olds cannot understand before and after. *Journal of Verbal Learning and Verbal Behaviour, 11*, 417-423. doi: 10.1016/S0022-5371(72)80022-7

Arnold, J. E., Eisenband, J., Brown–Schmidt, S., & Trueswell, J. C. (2000). The rapid use of gender information: Evidence of the time course of pronoun resolution from eyetracking. *Cognition, 76*, 13–26. doi: 10.1016/S0010-0277(00)00073-1

Atassanova, M. (2001). On the acquisition of temporal conjunctions in Finnish. *Journal of Psycholinguistic Research, 30*, 115-134. doi: 10.1023/A:1010321611031

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

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., & 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.

Bannard, C., & Matthews, D. (2008). Stored word sequences in language learning: The effect of familiarity on children's repetition of four-word combinations. *Psychological Science, 19*, 241-248. doi:10.1111/j.1467-9280.2008.02075.x

Barnes, M.A., Ahmed, Y., Barth, A., & Francis, D.J. (2015). The relation of knowledge-text integration processes and reading comprehension in 7th- to 12th-grade students. *Scientific Studies of Reading, 19*, 253-272, doi: 10.1080/10888438.2015.1022650

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.

Bates, D.M., Kliegl, R., Vasishth, S., & Baayen, H. (2016). Parsimonious mixed models. Manuscript submitted to *Journal of Memory and Language* for publication. doi: arXiv:1506.04967

Benedict, H. (1979). Early lexical development: Comprehension and production. *Journal of Child Language, 6*, 183–200. doi: 10.1017/S0305000900002245

Bereiter, C., & Scardamalia, M. (1987). *The psychology of written composition*. Hillsdale, NJ: Lawrence Erlbaum Associates.

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.

Bishop, D. V. M. (1983). *Test for the Reception of Grammar*. Manchester, UK: Chapel Press.

Bloom, L., Lahey, M., Hood, L., Lifter, K., & Fiess, K. (1980). Complex sentences: acquisition of syntactic connectives and the semantic relations they encode. *Journal of Child Language, 7*, 235-61. doi: 10.1017/S0305000900002610

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

Bock, J. K. (1986). Syntactic persistence in language production. *Cognitive Psychology*, *18***,** 355–387. doi: 10.1016/0010-0285(86)90004-6

Bos, L. T., de Koning, B. B., van Wesel, F., Boonstra, A.M., & van der Schoot, M. (2015). What can measures of text comprehension tell us about creative text production? *Reading and Writing, 28*, 829–849. doi: 10.1007/s11145-015-9551-6

Boyle, W., Lindell, A.K., & Kidd, E. (2013). Investigating the role of verbal working memory in young children's sentence comprehension. *Language Learning, 63*, 211-242. doi: 10.1111/lang.12003

Bransford, J. D., Barclay, J. R., & Franks, J. J. (1972). Sentence memory: a constructive versus interpretive approach. *Cognitive Psychology, 3*, 193–209. doi: 10.1016/0010-0285(72)90003-5

Brown, A. S. (1991). A review of the tip-of-the-tongue experience. *Psychological Bulletin, 109*, 204–223. doi: 10.1037/0033-2909.109.2.204

Brown, A. L. & French, L. A. (1976). Construction and regeneration of logical sequences using causes or consequences as the point of departure. *Child Development, 47*, 930-940. doi: 10.1016/0022-0965(75)90106-X

Brown, A. L. & Murphy, M. D. (1975). Reconstruction of arbitrary versus logical sequences by preschool children. *Journal of Experimental Child Psychology, 20,* 307-326. doi: 10.1016/0022-0965(75)90106-X

Brown-Schmidt, S., & Konopka, A. E. (2015). Processes of incremental message planning during conversation. *Psychonomic Bulletin & Review, 22*, 833-843. doi: 10.3758/s13423-014-0714-2

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. (2007). Syntactic awareness and reading ability: Is there any evidence for a special relationship? *Applied Psycholinguistics, 28*, 679-694. doi: 10.1017/S0142716407070361

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

Cain, K., Oakhill, J. V., & Bryant, P. E. (2004). Children's reading comprehension ability: Concurrent prediction by working memory, verbal ability, and component skills. *Journal of Educational Psychology, 96*, 671-681. doi: 10.1037/0022-0663.96.1.31

Cain, K., Patson, N., & Andrews, L. (2005). Age- and ability-related differences in young readers' use of conjunctions. *Journal of Child Language, 32*, 877-892. doi: 10.1017/S0305000905007014

Carretti, B., Borella, E., Cornoldi, C., & De Beni, R. (2009). Role of working memory in explaining the performance of individuals with specific reading comprehension difficulties: A meta-analysis. *Learning and Individual Differences, 19*, 246-251. doi:10.1016/j.lindif.2008.10.002.

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

Case, R., Kurland, D. M., & Goldberg, J. (1982). Operational efficiency and the growth of short-term memory span. *Journal of Experimental Child Psychology, 33*, 386–404.

Chafe, W. (1984). *How people use adverbial clauses*. *Berkeley Linguistics Society, 10*, 437-449. doi: 10.3765/bls.v10i0.1936

Chater, N., McCauley, S. M., & Christiansen, M. H. (2016). Language as skill: Intertwining comprehension and production. *Journal of Memory and Language, 89,* 244–254. *doi:10.1016/j.jml.2015.11.004*

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

Cohen, J. (1992). A power primer. Psychological Bulletin, 112, 155-159.

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

Coltheart, M., Rastle, K., Perry, C., Langdon, R., & Ziegler, J. (2001). DRC: A dual route cascaded model of visual word recognition and reading aloud. *Psychological Review, 108*, 204–256. doi: 10.1037/0033-295X.108.1.204

Costermans, J., & Fayol, M. (Eds.). (1997). *Processing interclausal relationships: Studies in the production and comprehension of text.* Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

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

Crosson, A. C., & Lesaux, N. K. (2013). Does knowledge of connectives play a special role in the reading comprehension of English language learners and English–only students? *Journal of Research in Reading, 36*, 241–260. doi: 10.1111/j.1467-9817.2011.01501.x

Crosson, A.C., Lesaux, N.K., & Martiniello, M. (2008). Factors that influence comprehension of connectives among language minority children from Spanish-speaking backgrounds. *Applied Psycholinguistics, 29*, 603–625. doi: 10.1017/S0142716408080260

Currie, N., & Cain, K. (2015). Children’s inference generation: the role of vocabulary and working memory. *Journal of Experimental Child Psychology, 137*, 57-75. doi: 10.1016/j.jecp.2015.03.005

Daneman, M., & Blennerhassett, A. (1984). How to assess the listening comprehension skills of prereaders. *Journal of Educational Psychology, 76*, 1372–1381.

Daneman, M., & Carpenter, P. A. (1980). Individual differences in working memory and reading comprehension. *Journal of Verbal Learning and Verbal Behavior, 19,* 450–466.

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

Dehaene-Lambertz, G., Dehaene, S., & Hertz-Pannier, L. (2002). Functional neuroimaging of speech perception in infants. *Science, 298*, 2013–15. doi: 10.1126/science.1077066

DeKeyser, R. (2003). Implicit and explicit learning. In C. Doughty & M. Long (Eds.), *The Handbook of Second Language Acquisition* (pp. 313-349). Oxford: Blackwell.

DfE. (2014). *National curriculum in England. English programmes of study - Key stages 1 and 2* Retrieved from: https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/335186/PRIMARY_national_curriculum_-_English_220714.pdf

Diessel, H. (2004). *The acquisition of complex sentences in English*. Cambridge: Cambridge University Press.

Diessel, H. (2005). Competing motivations for the ordering of main and adverbial clauses. *Linguistics, 43*, 449–470. doi: 10.1515/ling.2005.43.3.449

Diessel, H. (2008). Iconicity of sequence: A corpus-based analysis of the positioning of temporal adverbial clauses in English. *Cognitive Linguistics, 19*, 465–490. doi: 10.1515/COGL.2008.018

Ditman, T., & Kuperberg, G. R. (2010). Building coherence: a framework for exploring the breakdown of links across clause boundaries in schizophrenia. *Journal of Neurolinguistics, "Language in Schizophrenia" Special Issue, 23,* 254-269.

Dunn, L. M., Dunn, D. M., Styles, B., & Sewell, J. (2009). *The British Picture Vocabulary Scale III – 3rd Edition*. London: GL Assessment.

Ericsson, K. A., & Kintsch, W. (1995). Long-term working memory. *Psychological Review*, *102,* 211-245. doi: 10.1037/0033-295X.102.2.211

Erlam, R. (2006). Elicited imitation as a measure of L2 implicit knowledge: An empirical validation study. *Applied Linguistics, 27*, 464-491. doi: 10.1093/applin/aml001

Evers-Vermeul, J., & Sanders, T. (2009). The emergence of Dutch connectives: How cumulative cognitive complexity explains the order of acquisition. *Journal of Child Language, 36*, 829-854. doi: 10.1017/S0305000908009227

Ferreira, V. S. (1996). Is it better to give than to donate? Syntactic flexibility in language production. *Journal of Memory and Language, 35*, 724–755. doi:10.1006/jmla.1996.0038

Ferreira, F., Bailey, K. G., & Ferraro, V. (2002). Good-enough representations in language comprehension. *Current Directions in Psychological Science, 11*, 11-15. doi: 10.1111/1467-8721.00158

Field, A. P. (2009*). Discovering statistics using SPSS: and sex and drugs and rock 'n' roll (third edition).* London: Sage publications.

Finney, M. C., Montgomery, J. W., Gillam, R. B., & Evans, J. L. (2014). Role of working memory storage and attention focus switching in children’s comprehension of spoken object relative sentences. *Child Development Research, 2014*, 1-11. http://dx.doi.org/10.1155/2014/450734

Fraser, C., Bellugi, U., & Brown, R. (1963). Control of grammar in imitation, comprehension, and production. *Journal of Verbal Learning and Verbal Behavior, 2*, 121–35. doi: 10.1016/S0022-5371(63)80076-6

French, L. A., & Brown, A. L. (1977). Comprehension of before and after in logical and arbitrary sequences. *Journal of Child Language, 4,* 247–256. doi: 10.1017/S0305000900001641

Friend, M., & Keplinger, M. (2008). Reliability and validity of the Computerized Comprehension Task (CCT): data from American English and Mexican Spanish infants. *Journal of Child Language, 35*, 77–98. doi: 10.1017/S0305000907008264

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

Garrett, M. F. (1980). Levels of processing in sentence production. In B. Butterworth (Ed.), *Language production: Vol. 1. Speech and talk* (pp. 177–220). London: Academic Press.

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

Gennari, S., & MacDonald, M. (2009). Linking production and comprehension processes: The case of relative clauses. *Cognition, 111*, 1-23. doi: 10.1016/j.cognition.2008.12.006

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

Gernsbacher, M. A. (1997). Coherence cues mapping during comprehension. In J. Costermans & M. Fayol (Eds.), *Processing interclausal relationship: Studies in the production and comprehension of text* (pp. 3–21). Mahwah, NJ: Lawrence Erlbaum Associates.

Gernsbacher, M. A., & Hargreaves, D. (1988). Accessing sentence participants: The advantage of first mention. *Journal of Memory and Language, 27*, 699-717. doi:10.1016/0749-596X(88)90016-2

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

Goldberg, A. E., Casenhiser, D. M., & Sethuraman, N. (2004). Learning argument structure generalizations. *Cognitive Linguistics, 15*, 289–316. doi:10.1515/cogl.2004.011

Gorrell, P., Crain, S., & Fodor, J. D. (1989). Contextual information and temporal terms. *Journal of Child Language, 16,* 623-632. doi:10.1017/S0305000900010758

Gough, P. B., & Tunmer, W. (1986). Decoding, reading and reading disability. *Remedial and Special Education, 7*, 6–10. doi: 10.1177/074193258600700104

Graesser, A. C., Singer, M., & Trabasso, T. (1994). Constructing inferences during narrative text comprehension. *Psychological Review, 101*, 371–395. doi: 10.1037/0033-295X.101.3.371

Grice, H. P. (1975). Logic and conversation. In P. Cole & J. Morgan (Eds.), *Syntax and semantics Vol. 3* (pp. 41–58). New York: Academic Press.

Grimm, A., Müller, A., Hamann, C., & Ruigendijk, E. (2011). *Production-comprehension asymmetries in child language.* Berlin: De Gruyter Mouton.

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

Habets, B., Jansma, B. M., & Münte, T. F. (2008). Neurophysiological correlates of linearization in language production. *BMC Neuroscience, 9*, 77. doi: 10.1186/1471-2202-9-77

Halliday, M. A. K., & Hasan, R. (1976). *Cohesion in English*. London: Longman.

Hamburger, H., & Crain, S. (1982). Relative acquisition. In S. Kuczaj (Ed.), *Language development, Vol.1: Syntax and semantics* (pp. 245- 274). Hillsdale: NJ: Erlbaum.

Hartsuiker, R. J., & Barkhuysen, P. N. (2006). Language production and working memory: the case of subject-verb agreement. *Language and Cognitive Processes, 21*, 181–204. doi: 10.1080/01690960400002117

Hawkins, J. A. (2004). *Efficiency and Complexity in Grammars*. Oxford: Oxford University Press.

Huttenlocher, J., Vasilyeva, M., & Shimpi, P. (2004). Syntactic priming in young children. *Journal of Memory and Language, 50*, 182–195. doi: 10.1016/j.jml.2003.09.003

Huttenlocher, J., Vasilyeva, M., Cymerman, E., & Levine, S. (2002). Language input and child syntax. *Cognitive Psychology, 45*, 337–374. doi:10.1016/S0010-0285(02)00500-5

Indefrey, P., & Levelt, W. J. M. (2004). The spatial and temporal signature of word production components. *Cognition, 92*, 101–144. doi: 10.1016/j.cognition.2002.06.001

Johnson, H. I. (1975). The meaning of before and after for preschool children. *Journal of Experimental Child Psychology, 19*, 88–99. doi: 10.1016/0022-0965(75)90151-4

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

Juel, C., Griffith, P. L., & Gough, P. B. (1986). Acquisition of literacy: A longitudinal study of children in first and second grade. *Journal of Educational Psychology, 78,* 243-255. doi: 10.1037/0022-0663.78.4.243

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

Junge, B., Theakston, A. L., & Lieven, E. V. M. (2013). Given-new/new-given? Children's sensitivity to the ordering of information in complex sentences. *Applied Psycholinguistics*, *36*, 589-612. doi: 10.1017/S0142716413000350

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

Just, M., & Carpenter, P. (1992). A capacity theory of comprehension: Individual differences in working memory. *Psychological Review, 99*, 122–149. doi: 10.1037//0033-295X.99.1.122

Karimi, H., & Ferreira, F. (2016). Good-enough linguistic representations and online cognitive equilibrium in language processing, *The Quarterly Journal of Experimental Psychology, 69*, 1013-1040. doi: 10.1080/17470218.2015.1053951

Keller-Cohen, D. (1981). Elicited imitation in lexical development: evidence from a study of temporal reference. *Journal of Psycholinguistic Research, 10*, 273–88. doi: 10.1007/BF01067508

Keller-Cohen, D. (1987). Context and strategy in acquiring temporal connectives. *Journal of Psycholinguistic Research, 16*, 165–85. doi: 10.1007/BF01072000

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

Kidd, E., Brandt, S., Lieven, E., & Tomasello, M. (2007). Object relatives made easy: A cross-linguistic comparison of the constraints influencing young children's processing of relative clauses. *Language and Cognitive Processes, 22*, 860-897. doi: 10.1080/01690960601155284

King, J., & Just, M. A. (1991). Individual differences in syntactic processing: The role of working memory. *Journal of Memory and Language, 30*, 580–602. doi: 10.1016/0749-596X(91)90027-H

Kintsch, W. (1998). *Comprehension: A paradigm for cognition*. New York, NY: Cambridge University Press.

Kintsch, W., Welsch, D.M., Schmalhofer, F., & Zimny, S. (1990). Sentence memory: A theoretical analysis. *Journal of Memory and Language, 29*, 133-159. doi: 10.1016/0749-596X(90)90069-C

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

Kucian, K., von Aster, M., Loenneker, T., Dietrich, T., Mast, F.W., & Martin, E. (2007). Brain activation during mental rotation in school children and adults. *Journal of Neural Transmission, 114,* 675-686. doi: 10.1007/s00702-006-0604-5

Kuhl, P., & Rivera-Gaxiola, M. (2008). Neural substrates of language acquisition. *Annual Review of Neuroscience, 31*, 511-534. doi: 10.1146/annurev.neuro.30.051606.094321

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.

Lust, B., Flynn, S., & Foley, C. (1995). What children know about what they say: Elicited imitation as a research method for assessing children's syntax. In D. McDaniel, C. McKee & H. Smith-Cairns (Eds.), *Methods for assessing children's syntax* (pp. 55-76). Cambridge, MA: MIT Press.

MacDonald, M. C. (2013). How language production shapes language form and comprehension. *Frontiers in Psychology, 4,* 1-16. doi: 10.3389/fpsyg.2013.00226

MacDonald, M., & Christiansen, M. (2002). Reassessing working memory: Comment on Just and Carpenter (1992) and Waters and Caplan (1996). *Psychological Review, 109,* 35–54.

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.

Mazzoni, D. (2014). Audacity (Version 2.0.5) [Computer software]. Retrieved June 10, 2014. Available from http://audacity.sourceforge.net

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.

Metusalem, R., Kutas, M., Urbach, T. P., Hare, M., McRae, K., & Elman, J. L. (2012). Generalized event knowledge activation during online sentence comprehension. *Journal of Memory and Language, 66*, 545-567. doi: 10.1016/j.jml.2012.01.001

Millis, K. K., & Just, M. A. (1994). The influence of connectives on sentence comprehension. *Journal of Memory and Language, 33*, 128- 147. doi: 10.1006/jmla.1994.1007

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

Morra, S. (1994). Issues in working memory measurement: Testing for M capacity. *International Journal of Behavioural Development, 17*, 143–159.

Mouchon, S., Fayol, M., & Gaonac'h, D. (1995). On-line processing of links between events in narratives: Study of children and adults. *Current Psychology of Cognition, 14*, 171–193. doi: 10.3951/996.07667.001

Müller, H. J., Heller, D., & Ziegler, J. (1995). Visual search for singleton feature targets within and across feature dimensions. *Perception & Psychophysics, 57*, 1-17. doi: 10.3758/BF03211845

Münte, T. F., Schiltz, K., & Kutas, M. (1998). When temporal terms belie conceptual order. *Nature, 395*, 71–73. doi: 10.1038/25731

Murray, J. D. (1997). Connectives and narrative text: The role of continuity. *Memory and Cognition, 25*, 227-236. doi: 10.3758/BF03201114

Myachykov, A., Garrod, S., & Scheepers, C. (2012). Determinants of structural choice in visually situated sentence production. *Acta Psychologica, 141*, 304-315. doi:10.1016/j.actpsy.2012.09.006

Nation, K., Marshall, C., & Altmann, G. T. M. (2003). Investigating individual differences in children's real-time sentence comprehension using language-mediated eye movements. *Journal of Experimental Child Psychology*, *86*, 314-329. doi:10.1016/j.jecp.2003.09.001

Oberauer, K., & Lewandowsky, S. (2010). Modeling working memory: A computational implementation of the Time-Based Resource Sharing theory. *Psychonomic Bulletin & Review, 18*, 10–45.

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

Perfetti, C. A. (2007). Reading ability: Lexical quality to comprehension. *Scientific Studies of Reading, 11*, 357-383. doi: 10.1080/10888430701773926

Peterson, C., & McCabe, A. (1983). *Developmental psycholinguistics: Three ways of looking at a child's narrative*. New York: Plenum.

Peterson, C., & McCabe, A. (1987). The connective 'and': Do older children use it less as they learn other connectives? *Journal of Child Language, 14,* 375–381. doi: 10.1017/S0305000900012988

Peterson, C., & McCabe, A. (1991). On the threshold of the story realm: Semantic versus pragmatic use of connectives in narratives. *Merrill Palmer Quarterly, 37*, 445 464. doi: 10.1017/S0305000900006905

Pickering, M. J., & Branigan, H. P. (1999). Syntactic priming in language production. *Trends in Cognitive Sciences, 3*, 136–141. doi:10.1016/S1364-6613(99)01293-0

Pickering, M. J., & Garrod, S. (2013). An integrated theory of language production and comprehension. *Behavioural and Brain Sciences, 36*, 329-47. doi:10.1017/S0140525X12001495

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.

Reali, F., Spivey, M. J., Tyler, M. J., & Terranova, J. (2006). Inefficient conjunction-search made efficient by concurrent spoken delivery of target identity. *Perception and Psychophysics, 68*, 959. doi: 10.3758/BF03193358

Reichle, E. D., Pollatsek, A., Fisher, D. L., & Rayner, K. (1998). Toward a model of eye movement control in reading. *Psychological Review, 105,* 125–157. doi:10.1016/S0042-6989(99)00152-2

Rochester, S., & Martin, J. R. (1979). *Crazy Talk: A Study of Discourse of Schizophrenic Speakers*. New York: Plenum.

Rowland, C. F., Chang F., Ambridge, B., Pine, J. M., Lieven, E. V. (2012). The development of abstract syntax: Evidence from structural priming and the lexical boost. *Cognition. 125*, 49-63. doi:10.1016/j.cognition.2012.06.008

Rowland, C. F., & Noble, C. L. (2010). The role of syntactic structure in children's sentence comprehension: *Evidence from the dative. Language Learning and Development, 7*, 55-75. doi: 10.1080/15475441003769411

Sanders, T., & Noordman, L. (2000). The role of coherence relations and their linguistic markers in text processing. *Discourse Processes, 29*, 37-60. doi: 10.1207/S15326950dp2901_3

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

Shapiro, L., & Hudson, J.A. (1991). Tell me a make-believe story: Coherence and cohesion in young children’s picture elicited narratives. *Developmental Psychology, 27*, 960-974. doi: 10.1037/0012-1649.27.6.960

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

Silva, M. N. (1991). Simultaneity in children's narratives: the case of when, while and as. *Journal of Child Language, 18*, 641-662. doi: 10.1017/S0305000900011296

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

Spooren, W., & Sanders, T. (2008). The acquisition order of coherence relations: On cognitive complexity in discourse. *Journal of Pragmatics, 40*, 2003-2026. doi:10.1016/j.pragma.2008.04.021

Stevenson, R. J., & Pollitt, C. (1987). The acquisition of temporal terms. *Journal of Child Language, 14*, 533-545. doi: 10.1017/S0305000900010278

Theakston, A. L., Lieven, E. V. M., Pine, J. M., & Rowland, C. F. (2004). Semantic generality, input frequency and the acquisition of syntax. *Journal of Child Language, 31*, 61–99. doi:10.1017/ S0305000903005956

Tomasello, M. (2003). *Constructing a language: A usage-based theory of language acquisition.* Cambridge, MA: Harvard University Press.

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

Traxler, M. J., & Gernsbacher, M. A. (1995). Improving coherence in written communication. In M. A. Gernsbacher & T. Givón (Eds.), *Coherence in spontaneous text* (pp. 216-237). Philadelphia, PA: John Benjamins.

Treiman, R., Clifton, C., Jr, Meyer, A. S., & Wurm, L. H. (2003). Language comprehension and production. In A. F. Healy & R. W. Proctor (Eds.), *Comprehensive Handbook of Psychology*, *Vol. 4. Experimental Psychology* (pp. 527-548). New York: John Wiley & Sons.

Trosborg, A. (1982). Children's comprehension of "before" and "after" reinvestigated. *Journal of Child Language, 9,* 381-402. doi: 10.1017/S0305000900004773

Van Dijk, T.A., & Kintsch, W. (1983). *Strategies of discourse comprehension*. New York: Academic.

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

Van Dyke, J.A., & Shankweiler, D. (2013). From verbal efficiency theory to lexical quality: The role of memory processes in reading comprehension. In M. A. Britt, S. R. Goldman & J. F Rouet (Eds.), *Reading: From Words to Multiple Texts* (pp. 115-131). Routledge: Taylor & Francis Group.

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

Vigliocco, G., & Hartsuiker, R. J. (2002). The interplay of meaning, sound, and syntax in sentence production. *Psychological Bulletin, 128*, 442-472. doi: 10.1037/0033-2909.128.3.442.

Vion, M., & Colas, A. (2004). On the use of the connective and in oral narration: a study of French-speaking elementary school children. *Journal of Child Language, 31*, 399-419. doi: 10.1017/S0305000904006014

Vion, M., & Colas, A. (2005). Using connectives in oral French narratives: Cognitive constraints and development of narrative skills. *First Language, 25*, 39–66. doi: 10.1177/0142723705046899

Wager, T. D., & Smith, E. E. (2003). Neuroimaging studies of working memory: A metaanalysis. *Cognitive, Affective, & Behavioural Neuroscience, 3*, 255–274. doi: 10.3758/CABN.3.4.255

Weist, R. M., Lyytinen, P., Wysocka, J., & Atanassova, M. (1997). The interaction of language and thought in children's language acquisition: A crosslinguistic study. *Journal of Child Language, 24*, 81–121. doi: 10.1017/S0305000996003017

Wells, J. B., Christiansen, M. H., Race, D. S., Acheson, D. J., & MacDonald, M. C. (2009). Experience and sentence comprehension: Statistical learning and relative clause comprehension. *Cognitive Psychology, 58*, 250-271. doi:10.1016/j.cogpsych.2008.08.002

Winskel, H. (2003). An investigation into the acquisition of temporal event sequencing: A cross-linguistic study using an elicited imitation task. *First Language, 23*, 65–95. doi: 10.1023/B:JOPR.0000035105.65440.b8

Winskel, H. (2007). The expression of temporal relations in Thai children's narratives. *First Language, 27*, 133-158. doi: 10.1177/0142723706074794

Ye, Z., Habets, B., Jansma, B. M., & Münte, T. F. (2011). Neural basis of linearization in speech production. *Journal of Cognitive Neuroscience, 23*, 3694-3702. doi: 10.1162/jocn_a_00037

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. (2008). Time in language, situation models, and mental simulations. *Language Learning, 58*, 13–26. doi: 10.1111/j.1467-9922.2008.00458.x

Zwaan, R. A., Langston, M. C., & Graesser, A. C. (1995). The construction of situation models in narrative comprehension: an event-indexing model. *Psychological Science, 6*, 292–297. doi:10.1016/j.cogpsych.2008.08.002

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

7. Appendix

Table A.3.1

Summary of GLMM: 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%**	
(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

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

Table A.3.2

Summary of GLMM: Main effect of 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%**	
Fixed effects:						
(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**

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

Table A.3.3

Summary of GLMM: 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%**
(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**

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

Table A.3.4

Summary of GLMM: 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%**
(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**

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

Table A.4.1.

Frequency counts of each individual error types made by 3- to 5-year-olds in the sentence repetition and blocked elicited production task.

Error type	Example target: Tom ate the burger, after he poured the ketchup	Sentence Repetition	Blocked Elicited Production
Sense maintained		131	410
Connective only	Tom ate the burger, when he poured the ketchup	22	69
Connective and order	Tom poured the ketchup, before he ate the burger	41	104
Connective and position	Before Tom ate the burger, he poured the ketchup	22	88
Connective, order and position	When Tom poured the ketchup, he ate the burger	17	52
Order and position	After Tom poured the ketchup, he ate the burger	29	97
Sense changed		358	152
Connective only	Tom ate the burger, before he poured the ketchup	189	33
Connective and order	Tom poured the ketchup, when he ate the burger	16	3
Connective and position	When Tom ate the burger, he poured the ketchup	29	4
Connective, order and position	Before Tom poured the ketchup, he ate the burger	18	11
Order only	Tom poured the ketchup, after he ate the burger	62	26
Position only	After Tom ate the burger, he poured the ketchup	44	69
Incomplete		93	305
No response	No response made or nonsensical	13	112
Clause omission	Tom ate the burger after he…I’ve forgotten	36	24
Full stop, no connective	Tom ate the burger. He poured the ketchup	2	75
‘And’ used as connective	Tom ate the burger and he poured the ketchup	42	95
Total errors		582	867

Table A.4.2.

Summary of GLMM: Main effect and interactions of age, connective, and order on the percentage of connective substitution errors in relation to the total errors by 3- to 4- and 4- to 5- year-olds in the sentence repetition task

Main model		**M (*b*)**	**SE**	**t**	**p(>|z|)**
	(Intercept)	0.36	0.39	0.94	0.35
	Age	-0.44	0.54	-0.82	0.41
	Order	**-1.14**	**0.50**	**-2.26**	**0.02**
	Connective	-0.68	0.40	-1.71	0.09
	Age:Order	0.60	0.69	0.87	0.38
	Age:Connective	0.61	0.56	1.10	0.27
	Order:Connective	0.60	0.60	1.01	0.31
	Age:Order:Connective	-0.87	0.86	-1.01	0.31

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

Table A.4.3.

Summary of GLMM: Main effect and interactions of age, memory, order and connective on 3- to 6- year-old's accuracy responses in the elicited production task.

Main model	**M (*b*)**	**SE**	**t**	**p(>|z|)**
(Intercept)	-35.51	5.25	-6.76	<0.01
Age	**0.51**	**0.08**	**6.18**	**<0.01**
Memory	0.01	0.18	0.05	0.96
Order	**17.04**	**5.15**	**3.31**	**<0.01**
Connective	**13.42**	**2.58**	**5.21**	**<0.01**
Order:Connective	**-11.69**	**3.23**	**-3.62**	**<0.01**
Age:Order	**-0.25**	**0.08**	**-3.10**	**<0.01**
Age:Connective	**-0.30**	**0.05**	**-5.98**	**<0.01**
Memory:Order	0.09	0.17	0.53	0.59
Memory:Connective	**0.34**	**0.09**	**3.72**	**<0.01**
Age:Order:Connective	**0.24**	**0.06**	**4.11**	**<0.01**
Memory:Order:Connective	**-0.24**	**0.11**	**-2.10**	**0.04**

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

Table A.5

Summary of GLMMs (medial position sentences only, by age group) for the log-odds of accuracy responses to sentences: Effect of order.

Accuracy	**Age 3-4**				**Age 4-5**				**Age 5-6**				**Age 6-7**			
	(*b*)	**SE**	**z**	***p***	**(*b*)**	**SE**	**z**	***p***	**(*b*)**	**SE**	**z**	***p***	**(*b*)**	**SE**	**z**	***p***
Exp. 1																
Intercept	0.24	0.19	1.27	0.21	0.79	0.33	2.42	0.02	1.58	0.33	4.81	<0.01	2.83	0.32	8.88	<0.01
Order	**0.55**	**0.21**	**2.65**	**0.01**	**1.20**	**0.50**	**2.40**	**0.02**	**2.13**	**0.50**	**4.27**	**<0.01**	**2.28**	**0.69**	**3.33**	**<0.01**
Exp. 2																
Intercept	-0.26	0.21	-1.25	0.21	0.45	0.27	1.66	0.10	1.10	0.33	3.35	<0.01	0.97	0.28	3.47	<0.01
Order	**0.82**	**0.30**	**2.76**	**0.01**	**0.79**	**0.34**	**2.34**	**0.02**	0.47	0.34	1.40	0.16	**0.98**	**0.31**	**3.18**	**<0.01**
Exp. 3																
Intercept	-0.36	0.40	-0.90	0.37	-0.57	0.68	-0.84	0.40	-	-	-	-	-	-	-	-
Order	0.50	0.38	1.32	0.19	**1.87**	**0.65**	**2.87**	**<0.01**	-	-	-	-	-	-	-	-
Exp. 4																
Intercept	-5.01	1.00	-4.99	<0.01	-8.67	4.31	-2.01	0.04	1.69	0.93	1.82	0.07	-	-	-	-
Order	-5.61	6.97	-0.80	0.42	**13.68**	**4.62**	**2.96**	**<0.01**	**5.81**	**2.59**	**2.24**	**0.02**	-	-	-	-

RTs	**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**
Exp. 2															
Intercept	1.19	0.06	19.48	1.07	1.30	1.02	0.06	16.57	0.90	1.14	0.94	0.07	14.37	0.81	1.06
Order	<0.01	0.06	-0.04	-0.12	0.12	**-0.12**	**0.05**	**-2.22**	**-0.22**	**-0.01**	-0.07	0.05	-1.27	-0.17	0.04

Note (1) Fixed effects labels: Order = effect of order, chronological vs. reverse (2) Bold = predictor is significant at $p < .05$ or better.

Table A.6.1
Mean (SD) proportion correct for each sentence type by 3- to 7-year-olds in Experiment 1

	3 to 4 years	4 to 5years	5 to 6 years	6 to 7 years
Before-chronological	0.72 (0.45)	0.81 (0.39)	0.92 (0.27)	0.96 (0.19)
Before-reverse	0.66 (0.47)	0.67 (0.47)	0.87 (0.33)	0.93 (0.24)
After-chronological	0.67 (0.47)	0.71 (0.45)	0.79 (0.41)	0.87 (0.33)
After-reverse	0.57 (0.5)	0.64 (0.48)	0.77 (0.42)	0.91 (0.28)

Table A.6.2

Mean (SD) proportion correct for each sentence type by 3- to 7-year-olds in Experiment 2.

	3 to 4 years	4 to 5 years	5 to 6 years	6 to 7 years
Before-chronological	0.63 (0.49)	0.70 (0.46)	0.82 (0.38)	0.86 (0.34)
Before-reverse	0.42 (0.49)	0.45 (0.50)	0.63 (0.49)	0.66 (0.47)
After-chronological	0.63 (0.49)	0.75 (0.43)	0.80 (0.40)	0.84 (0.37)
After-reverse	0.44 (0.50)	0.58 (0.49)	0.69 (0.46)	0.69 (0.46)

Table A.6.3.
Mean (SD) response times for each sentence type by 4- to 7-year-olds in Experiment 2.

	4- to 5-year-olds	5- to 6-year-olds	6- to 7-year-olds
Before-chronological	1.60 (1.35)	1.02 (1.16)	0.95 (0.96)
Before-reverse	2.11 (1.37)	1.42 (1.34)	1.04 (1.3)
After-chronological	1.76 (1.44)	1.17 (1.13)	1.11(1.23)
After-reverse	1.67 (1.41)	1.17 (1.06)	1.12 (1.33)

Table A.6.4.
Mean (SD) proportion correct for each sentence type by 3- to 6-year-olds in Experiment 3.

	3- to 4-year-olds	4- to 5-year-olds	5- to 6-year-olds
Before-chronological	0.53 (0.50)	0.65 (0.48)	0.89 (0.32)
Before-reverse	0.59 (0.49)	0.62 (0.49)	0.88 (0.33)
After-chronological	0.64 (0.48)	0.62 (0.49)	0.84 (0.36)
After-reverse	0.44 (0.50)	0.47 (0.50)	0.76 (0.43)

Table A.6.5.
Mean (SD) onset times (seconds) for each experimental condition by 5- to 6-year-olds in Experiment 3.

Before-chronological	0.73 (0.55)
Before-reverse	0.79 (0.64)
After-chronological	0.71 (0.42)
After-reverse	0.75 (0.5)

Table A.6.6.
Mean (SD) proportion correct for each experimental condition by 3- to 6-year-olds in a blocked elicited production paradigm.

	3- to 4-years	4- to 5-years	5- to 6-years	All ages
Before-chronological	0.28 (0.45)	0.61 (0.49)	0.81 (0.39)	0.57 (0.50)
Before-reverse	0.15 (0.36)	0.55 (0.50)	0.74 (0.44)	0.48 (0.50)
After-chronological	0.17 (0.38)	0.63 (0.48)	0.82 (0.39)	0.55 (0.50)
After-reverse	0.01 (0.08)	0.27 (0.45)	0.68 (0.47)	0.32 (0.47)

Table A.7.1.

Zero-order correlations between the main effects and interactions of age, order, and connective in Experiment 1 (a follow up for Table 2.2).

	1	2	3	4	5	6	7
1. Age							
2. Order	0.46						
3. Connective	0.45	0.21					
4. Age:Order	-0.47	-0.98	-0.21				
5. Age:Connective	-0.46	-0.21	-0.98	0.22			
6. Order:Connective	-0.22	-0.45	-0.43	0.45	0.43		
7. Age:Order:Connective	0.21	0.42	0.41	-0.44	-0.44	-0.98	

Table A.7.2.

Zero-order correlations between the main effects and interactions of age, memory, order, and connective in Experiment 1 (a follow up for Table 2.4).

	1	2	3	4	5	6	7	8	9	10	11
1.Age											
2.Memory	-0.71										
3.Order	0.25	0.12									
4.Connective	0.23	0.11	0.23								
5.Order:Connective	-0.11	-0.06	-0.45	-0.42							
6.Age:Order	-0.51	0.38	-0.49	-0.11	0.23						
7.Age:Connective	-0.46	0.34	-0.11	-0.52	0.24	0.25					
8.Memory:Order	0.37	-0.52	-0.24	-0.06	0.11	-0.72	-0.18				
9. Memory: Connective	0.34	-0.48	-0.06	-0.21	0.08	-0.19	-0.71	0.27			
10. Age:Order: Connective	0.23	-0.17	0.21	0.22	-0.49	-0.46	-0.44	0.34	0.32		
11. Memory:Order: Connective	-0.17	0.24	0.10	0.07	-0.22	0.34	0.31	-0.46	-0.42	-0.73	

Table A.7.3.
Zero-order correlations between the main effects and interactions of age, order, memory and vocabulary on accuracy in Experiment 2 (a follow up for Table 3.2).

	1	2	3	4	5	6	7	8
1.Four-to-Five								
2.Five-to-Six	0.70							
3.Six-to-Seven	0.69	0.80						
4.Order	-0.01	0.00	0.01					
5.Memory	-0.11	-0.17	-0.10	-0.02				
6.Vocabulary	-0.44	-0.59	-0.65	-0.02	-0.41			
7.Memory : Order	-0.02	-0.02	-0.02	0.06	-0.61	0.32		
8.Vocabulary : Order	-0.01	0.00	0.01	0.09	0.41	-0.47	-0.66	

Table A.7.4.

Zero-order correlations between the main effects and interactions of memory, age, order, and connective on response times in Experiment 2 (a follow up for Table 3.3).

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1.Memory															
2.Five-to-Six	-0.36														
3.Six-to-Seven	-0.39	0.58													
4.Order	-0.16	0.39	0.39												
5.Connective	-0.14	0.34	0.34	0.48											
6.Memory:Order	-0.47	0.18	0.19	0.35	0.16										
7.Five-to-Six: Order	0.18	-0.52	-0.31	-0.76	-0.37	-0.35									
8. Six-to-Seven: Order	0.19	-0.31	-0.50	-0.76	-0.37	-0.39	0.59								
9.Memory: Connective	-0.43	0.17	0.17	0.16	0.35	0.46	-0.18	-0.19							
10.Five-to-Six: Connective	0.16	-0.46	-0.28	-0.38	-0.79	-0.18	0.50	0.30	-0.37						
11.Six-to-Seven: Connective	0.17	-0.28	-0.44	-0.38	-0.79	-0.18	0.30	0.48	-0.41	0.64					
12.Order: Connective	0.11	-0.27	-0.26	-0.67	-0.77	-0.24	0.51	0.51	-0.27	0.61	0.61				
13.Memory: Order: Connective	0.32	-0.13	-0.13	-0.25	-0.26	-0.69	0.25	0.27	-0.74	0.27	0.30	0.37			
14.Five-to-Six: Order: Connective	-0.13	0.36	0.21	0.52	0.59	0.25	-0.69	-0.40	0.28	-0.76	-0.48	-0.78	-0.37		
15.Six-to-Seven: Order: Connective	-0.13	0.21	0.34	0.52	0.60	0.27	-0.40	-0.68	0.31	-0.48	-0.76	-0.78	-0.41	0.62	

Table A.7.5
Zero-order correlations between the main effects and interactions of age, order, and connective on accuracy in Experiment 3 (a follow up for Table 4.2).

	1	2	3	4	5	6
1. Age						
2. Order	0.32					
3. Connective	0.41	0.22				
4. Age:Order	-0.33	-0.99	-0.22			
5. Age:Connective	-0.42	-0.22	-0.99	0.22		
6. Order:Connective	-0.16	-0.50	-0.39	0.50	0.39	
7. Age:Order:Connective	0.16	0.50	0.39	-0.50	-0.39	-0.99

Table A.7.6

Zero-order correlations between the main effects and interactions of age, memory, vocabulary, age, order, and connective on accuracy in Experiment 3 (a follow up for Table 4.3).

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1. Age															
2. Memory	-0.34														
3. Vocabulary	-0.27	-0.23													
4. Order	0.18	0.15	0.20												
5. Connective	0.23	0.17	0.24	0.27											
6. Order:Connective	-0.09	-0.09	-0.10	-0.49	-0.39										
7. Age:Order	-0.46	0.18	0.12	-0.42	-0.10	0.20									
8. Age:Connective	-0.57	0.20	0.15	-0.10	-0.43	0.15	0.27								
9. Memory:Order	0.17	-0.48	0.09	-0.30	-0.09	0.16	-0.34	-0.11							
10. Memory:Connective	0.20	-0.59	0.12	-0.10	-0.30	0.14	-0.11	-0.33	0.30						
11. Vocabulary:Order	0.13	0.09	-0.45	-0.42	-0.12	0.20	-0.28	-0.07	-0.21	-0.05					
12. Vocabulary:Connective	0.16	0.12	-0.57	-0.12	-0.42	0.16	-0.07	-0.29	-0.05	-0.21	0.26				
13. Age:Order:Connective	0.26	-0.11	-0.07	0.20	0.16	-0.43	-0.53	-0.42	0.20	0.16	0.15	0.12			
14. Memory:Order:Connective	-0.11	0.29	-0.05	0.17	0.14	-0.35	0.20	0.16	-0.56	-0.43	0.11	0.07	-0.31		
15. Vocabulary:Order:Connective	-0.07	-0.05	0.25	0.21	0.16	-0.41	0.15	0.13	0.11	0.07	-0.52	-0.39	-0.30	-0.17	

Table A.7.7

Zero-order correlations between the main effects and interactions of age, order, and connective on accuracy in Experiment 4 (a follow up for Table 4.4).

	1	2	3	4	5	6
1. Age						
2. Order	0.73					
3. Connective	0.39	0.42				
4. Age:Order	-0.73	-0.99	-0.40			
5. Age:Connective	-0.39	-0.41	-1.00	0.40		
6. Order:Connective	-0.32	-0.47	-0.81	0.46	0.81	
7. Age:Order:Connective	0.31	0.47	0.79	-0.46	-0.79	-0.99

Table A.7.8
Zero-order correlations between the main effects and interactions of age, memory, vocabulary, order, and connective on accuracy in Experiment 4 (a follow up for Table 4.5).

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1. Age															
2. Memory	-0.27														
3. Vocabulary	-0.38	-0.25													
4. Order	0.35	0.17	0.30												
5. Connective	0.28	-0.05	0.19	0.42											
6. Order:Connective	-0.23	0.04	-0.15	-0.47	-0.83										
7. Age:Order	-0.70	0.20	0.22	-0.47	-0.28	0.30									
8. Age:Connective	-0.42	0.22	-0.01	-0.28	-0.56	0.46	0.43								
9. Memory:Order	0.19	-0.69	0.15	-0.25	0.05	-0.01	-0.27	-0.22							
10. Memory:Connective	0.27	-0.34	0.08	0.10	0.10	-0.08	-0.27	-0.59	0.34						
11.Vocabulary:Order	0.22	0.15	-0.69	-0.39	-0.19	0.19	-0.35	0.00	-0.23	-0.08					
12. Vocabulary:Connective	-0.01	0.07	-0.28	-0.23	-0.56	0.46	0.01	-0.13	-0.07	-0.20	0.29				
13. Age:Order:Connective	0.36	-0.18	0.00	0.30	0.47	-0.55	-0.47	-0.84	0.21	0.50	0.03	0.11			
14. Memory:Order:Connective	-0.21	0.27	-0.07	-0.05	-0.08	-0.04	0.24	0.46	-0.42	-0.78	0.11	0.15	-0.44		
15. Vocabulary:Order:Connective	0.00	-0.06	0.22	0.21	0.43	-0.45	0.03	0.11	0.11	0.15	-0.35	-0.78	-0.20	-0.27	