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

Liam P. Blything[ab]* Kate Cain[a]

[a] Department of Psychology, Lancaster University, LA1 4YF, UK.

[b] Liam P. Blything is now at the Department of Linguistics, University of Alberta, Edmonton, T6G 2E7, Canada.

*corresponding author.

Email addresses: blything@alberta.ca (Liam P. Blything), k.cain@lancaster.ac.uk (Kate Cain).

Accepted for publication in Journal of Experimental Child Psychology, 18 Jan 2019

Acknowledgements

This work was supported by a studentship from the Economic and Social Research Council awarded to Liam P. Blything. We would like to thank the nurseries and schools who supported this programme of research and Michael Ratajczak for providing second rater reliability coding. For further details on materials or analysis methods please contact the corresponding author.

Abstract

We assessed 3- to 6-year-olds' production of two-clause sentences linked by *before* or *after*. In two experiments, children viewed an animated sequence of two actions, and were asked to describe the order of events in specific target sentence structures. We manipulated whether the target sentence structure matched the chronological order of events, for example: '*He finished his homework, before he played in the garden*' (chronological order) or not, for example: '*Before he played in the garden, he finished his homework*' (reverse order). Children produced fewer accurate 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*. Independent measures of vocabulary and memory were both related to performance, but vocabulary was the stronger predictor. We conclude that developmental improvements in children's ability to produce two-clause sentences linked by a sequential temporal connective is driven primarily by language ability, rather than memory capacity per se. The work also highlights the advantages of using both sentence repetition (Experiment 1) and blocked elicited production (Experiment 2) paradigms to elicit sentence production in young children.

Keywords: temporal connectives, language production, sentence repetition, memory, language acquisition.

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

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 knowledge 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 has a stronger influence on performance than working memory capacity per se.

Successful production of language draws on an integrated and coherent mental representation of the state of affairs being described, also known as a pre-linguistic message (Bock, 1987; Levelt, 1989). When a speaker narrates events in reverse order, as in *She put on her gloves, after she had combed her hair,* the language used deviates from the speaker's mental representation of the actual sequence of events. When adult speakers choose to do this, they draw on greater processing resources than when planning and producing chronological order sentences (Habets, Jansma, & Münte, 2008; Ye, Habets, Jansma, & Münte, 2011). As noted above, children's understanding and production of *before* and *after* continues to develop for several years after these temporal expressions first appear in their speech. What is not known is whether or not other linguistic and structural features of sentences with temporal connectives, such as reverse order narration, contribute to these developmental differences. We present the first systematic study of how sentences expressing different temporal orders of events affect children's sentence production accuracy.

A speaker's choice to narrate events in their chronological order using either *before* or *after* influences whether the temporal connective occurs in the 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 research 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? Studies of children's production and comprehension of temporal connectives show that *before* is acquired earlier than *after* (Blything & Cain, 2016; Clark, 1971; Pyykkönen & Järvikivi, 2012). Clark (1971) attributed the difference in age of acquisition for *before* and *after* to the semantic features of each term: *before* indicates the prior event, whereas *after* does not, making the latter more semantically complex. In addition, b*efore* 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). Thus, *after* has a less consistent form-meaning relationship and is theoretically more complex than *before.* This literature suggests the use of *after* may involve greater planning and processing effort than the use of *before*, which may influence the accuracy of sentence production.

Another feature that might influence children's sentence production accuracy is the position of the connective. Corpus studies of spoken language have reported that children and adults use connectives in an initial position infrequently (Diessel, 2004; 2008). This finding has been related to the memory load involved in maintaining the information signalled by the connective from the beginning of the sentence while processing the meaning of the first clause (Diessel, 2004; 2008). Conversely, the preference to use a medially placed connective is associated with processing ease because it provides the linguistic information about temporal order at a point close to when the events can be integrated during the incremental processing of language.

Our second research question is to identify which framework best explains variation in performance between different sentence structures and across development. A traditional memory capacity-constrained account attributes performance on language processing tasks to the availability of resources within an independent system of working memory, which limits the amount of information that can be maintained during planning and production (e.g., Carpenter, Miyake, & Just, 1994). A more nuanced language-based perspective of working memory shifts emphasis from the 'quantity' of information that can be represented to the 'quality' (i.e., the content) of the representation of that information in long term memory (McElree, 2006), which in turn frees up shared processing resources so that they are allocated to the representation of information in active working memory (e.g., MacDonald, 2016). We examined both of these accounts in our study.

A classic theory concerning the role of working memory in sentence processing is the memory capacity-constrained account (e.g., Carpenter et al., 1994). According to this viewpoint, an effect or interacting effect of the aforementioned features - order of events, connective, position of connective - is driven by working memory capacity alone. The primary emphasis is that some sentence structures are more difficult to process than others, because they require more information to be held within the limited-capacity working memory system. The account builds on a framework that assumes that working memory is a separate system from long term memory (Baddeley, 2003; Baddeley & Hitch, 1974), to argue that the accurate representation of information is driven by the availability of processing resources specific to the working memory system. The availability of processing resources determines how many individual language units can be accurately represented (but note that the constitution of a 'unit' is undefined; see McElree, 2006, for a full review of limitations). It follows that production accuracy is expected to be weaker in individuals with a low working memory capacity because they have fewer resources available for maintaining information in working memory. Under such circumstances, the representation of the language form and structure may decay and be forgotten.

There is empirical support for the memory capacity-constrained account of sentence processing from a variety of studies. First, there are studies suggesting that difficulties in the production of more complex utterances can be attributed to the availability of resources within an independent working memory system. Patients with working memory capacity deficits display a substantially longer speech onset than controls when producing various utterances, and this difference is more pronounced for utterances with more complex structures (e.g., Martin, Miller, & Vu, 2004; Martin & Freedman, 2001). In addition, healthy speakers produce an increased proportion of double object datives (a more complex dative structure: e.g., *the pirate is giving the monk the book*) relative to prepositional datives (a simpler dative structure: e.g., *the pirate is giving a book to the monk*) when they are *not* required to maintain a verbal memory load (Slevc, 2011). Specifically in relation to the production of two clause sentences containing temporal connectives, fMRI and EEG studies with adults have attributed the extra processing effort for reverse order sentences to the maintenance of additional concepts within working memory (Habets et al., 2008; Ye et al., 2011). However, these latter studies did not include an independent measure of working memory. Complementary work from studies of sentence comprehension, show that both adults' (Münte, Schiltz, & Kutas, 1998) and children's (Blything, Davies & Cain, 2015; Blything & Cain, 2016) weaker performance for reverse order sentences is related to an independent measure of working memory (but see de Ruiter, Theakston, Brandt, & Lieven, 2018, for a study with children that did not demonstrate a relationship between memory and sentence comprehension).

Alternatively, 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 the pattern of performance seen by children or adults in sentence processing tasks (MacDonald, 2016; McElree, 2006). A language-based account of sentence processing proposes that the effects of working memory are indirect via language knowledge (e.g., MacDonald, 2016). 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). From this viewpoint, language knowledge influences sentence processing because good language skills free up shared resources within the proposed unitary architecture to support the accurate representation of information in active working memory.

There is empirical support for this position. The specificity or distinctness of words in the target utterance has been shown to influence adults' sentence production (Gennari, Mirković & MacDonald, 2012; Montag & MacDonald, 2014, 2015; Smith & Wheeldon, 2004). In these studies, speakers are less accurate when a task involves the activation of competitors that carry a similar meaning to target items. In a picture description task, Gennari et al. (2012) contrasted conditions in which the pictured agent and patient were highly similar (e.g., *builder*, *miner*) or not (e.g., *builder, astronaut*). Speakers were more likely to avoid more complex structures and omit optional words in the highly similar condition that permitted potential competition between the agent and patient meanings (e.g., producing *The builder who's being slapped*, rather than *The builder who's being slapped by the miner*). It follows that robust language knowledge will ease the accessibility and retrieval of target items over competitors that are also partially activated in memory. In relation to the production of reverse order sentences containing temporal connectives, this account would posit that an accurate transformation of the mental representation of the order of events is determined by the availability of processing resources that are shared with language retrieval operations. Crucially, a weak lexical representation of a target connective, or other words in the sentence, will lead to less differentiation in activation compared to competitors with similar meaning (i.e., different temporal connectives to the target). This would disrupt sentence planning and production. On this basis, young language users may experience difficulties with complex sentences (i.e., reverse order) because the quality of their lexical representations (i.e., connectives or other words in the sentence) is weaker.

We do not yet know precisely how children's sentence production differs for sentences expressing different temporal orders of events. As noted, studies of adult's sentence production have reported processing difficulties for reverse order sentences (Habets et al., 2008; Ye et al., 2011). However, these studies have used stimuli in which the connective was presented only in the sentence initial position. As a result, the effects of connective (*before, after*) and event order (chronological, reverse) cannot be disentangled. From a developmental perspective, a fully factorial design that includes all permutations of these factors (before-chronological, before-reverse, after-chronological, after-reverse) is essential because children display developmental differences in their understanding of *before* and *after* (Clark, 1971).

Overview of study aims, methods, and hypotheses

We conducted two experiments designed to determine whether a memory capacity-constrained account (e.g., Carpenter et al., 1994) or a language-based account (e.g., MacDonald, 2016) of sentence processing best explains young children's production of two-clause sentences containing *before* and *after*. Each clause related a single event. We manipulated the connective (*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 connective and event order in our materials. We also examined the extent to which independent measures of language (receptive vocabulary) and working memory explained variance in performance.

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 contain features that would not be expected to increase the amount of information that must be held in working memory (chronological order, medial position, less complex connective). In contrast, the other structures each have two factors that increase the amount of information maintained in working memory: before-reverse (reverse order, initial position), after-chronological (initial position, more complex connective), and after-reverse (reverse order, more complex connective). Further, if memory capacity is the critical influence on accurate production, our independent measure of memory should predict the effect or interacting effect of these features, and also serve as a proxy for age, when both age and memory are included in the model.

A language-based account predicts that the influence of working memory is indirect and modulated by language knowledge (MacDonald, 2016). As a result, any difficulties with the production of reverse order sentences should be more pronounced when they are linked by the connective *after*, because the planning of a reverse order sentence should be disrupted more easily when it contains *after* than *before*. Also, the independent measure of vocabulary should modulate performance because it indicates the quality of an individual's language knowledge. According to this account, the inclusion of an independent measure of vocabulary should improve model fit over and above the inclusion of an independent measure of memory. Furthermore, because vocabulary knowledge grows with age, vocabulary effects should supersede any age effects, once included in the model.

In summary, there were three aims. First, to establish which features – order of events, connective, position of connective – influence the accuracy of children's production of two-clause sentences containing the temporal connectives *before* and *after.* Second, we examine which account best explains the pattern of performance: a traditional memory capacity-constrained account or a more nuanced perspective of working memory which argues that language knowledge influences memory processing and storage. Third, we ask whether the same pattern of performance is reproduced across our two paradigms designed to elicit sentence production.

Experiment 1

We assessed sentence production using a sentence repetition task, in which participants heard a target sentence and were asked to repeat it back to the experimenter. Sentence repetition is a sensitive measure of processing ease 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, Lindell, & Kidd, 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 sentences containing temporal connectives using sentence repetition have contrasted sequential (e.g., *then, before*) and simultaneous (e.g., *whilst, when*) connectives (Keller-Cohen, 1981; Winskel, 2003), so these do not speak to the issues addressed in this paper.

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. 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 split between two sessions. In addition, one session included an assessment of receptive vocabulary; the other, an assessment of memory. Each session lasted no longer than twenty minutes.

Sentence repetition. Thirty-two two-clause sequences containing *before* and *after* were constructed. 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 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*,* rather than arbitrary (e.g., *He put on the socks, before he put on the shoes.*). 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). We included these sentences to enhance the likelihood that children would produce full sentences in the task and to maintain their confidence (proportion accuracy reported in Appendix: Table A.1).

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 increasing accuracy (Ambridge & Lieven, 2011). Each animation segment explicitly showed an object (e.g., shoes) from one of the clauses; 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. Presentation of items was fully randomised.

Table 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.

Practice trials emphasised the importance of producing an exact copy of the narrated sentence. Children practiced each of the four sentence constructions used in the experimental items (i.e., two clause sentences linked by *before* and *after*) (see Table 1 for examples). 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. The instruction to prompt production 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. Responses were recorded using a digital voice recorder (Olympus VN-5500), and later transcribed and scored. 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, each child was able to copy at least one sentence.

For each experimental trial, an exact repetition was scored as a target response. Based on recommendations by Lust, Lynn and Foley (1995), a response was also marked as a target response if a change was only minor such as a change to the label for 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 because marking such changes as non-target responses 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 connective. The time taken between the beep and the start of a child's response was extracted using Audacity (Mazzoni, 2014). There were no significant differences in response times between age groups or sentence constructions for target responses, so response time data are not reported.

Non-target responses were first categorised into three broad types: *sense maintained, sense changed*, and *incomplete*. We categorised responses as a *sense maintained* response if the child inaccurately repeated the target sentence, but successfully communicated the order of events by using a temporal connective. The sense maintained responses were counted as non-target responses because at least one critical feature of the target sentence was missing (connective, order of mention, or position, see Table A.2 in Appendix). Responses were categorised as *sense changed* when a non-target order of events was communicated. 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) were categorised as incomplete because *and* does not explicitly specify order (Peterson & McCabe, 1987), so we were unable to categorise whether the response maintained or changed the sense or order. Within each of the three broad non-target response categories, we coded the specific change or combination of changes that the child had made. Our Appendix materials include examples and frequency counts of each specific non-target response type (see Table A.2).

A second coder blind to the hypotheses coded at least 10% of the data (randomly selected) from each year group. Agreement between coders was good for both accuracy (target vs non-target responses, agreement = 99%; Cohen's κ = .96) and also for the categories of non-target responses (Agreement = 96%; Cohen's κ = .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 recall the order of a string of digits read aloud by the assessor. The number of digits in a string increases 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). By using this measure, we could capture variance in memory performance across the entire age range.

Raw scores were used in 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 has been made. Raw scores were used in the analysis.

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 and order. Two analyses were conducted: one with number of target responses as the dependent variable; the other with non-target response types as the dependent variable.

Results

Method of analysis

The main analysis of the number of target responses was completed using Generalised Linear Mixed-effects Models (GLMMs) (Baayen, Davidson, & Bates, 2008; Barr, Levy, Scheepers & Tilly, 2013). Significant interactions were explored by additional analyses to identify the source of the interaction, and are reported below. These were conducted using the lme4 package from the R statistics environment (R Core Team, 2014) (Bates, Maechler & Bolker, 2014). A binomial link function was specified because the outcome variable was binary (i.e., target/non-target). We followed the recommendations of Barr et al. (2013) for obtaining an optimal model. Our maximum random effects models did not converge, so the decision to incorporate random intercepts and slopes for participants and items was determined by the result of incremental likelihood ratio tests that demonstrated whether each specific random effect significantly improved model fit (Barr et al., 2013). We describe the optimum models for each respective dataset in Tables 2 and 3, in which the first column provides the coefficient estimates of effects (*b*) due to experimental conditions, the change in the log odds accuracy of responses associated with each fixed effect. A positive coefficient indicates that the effect of a factor is to increase the odds of a target response whilst a negative coefficient indicates that the factor decreases the odds of a target response. Age in months (continuous), order (chronological, reverse), and connective (*before*, *after*) were entered as fixed effects. We used the scale function to scale and centre the age, memory, and vocabulary predictors.

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); and 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 available for 3- to 4-year-olds.

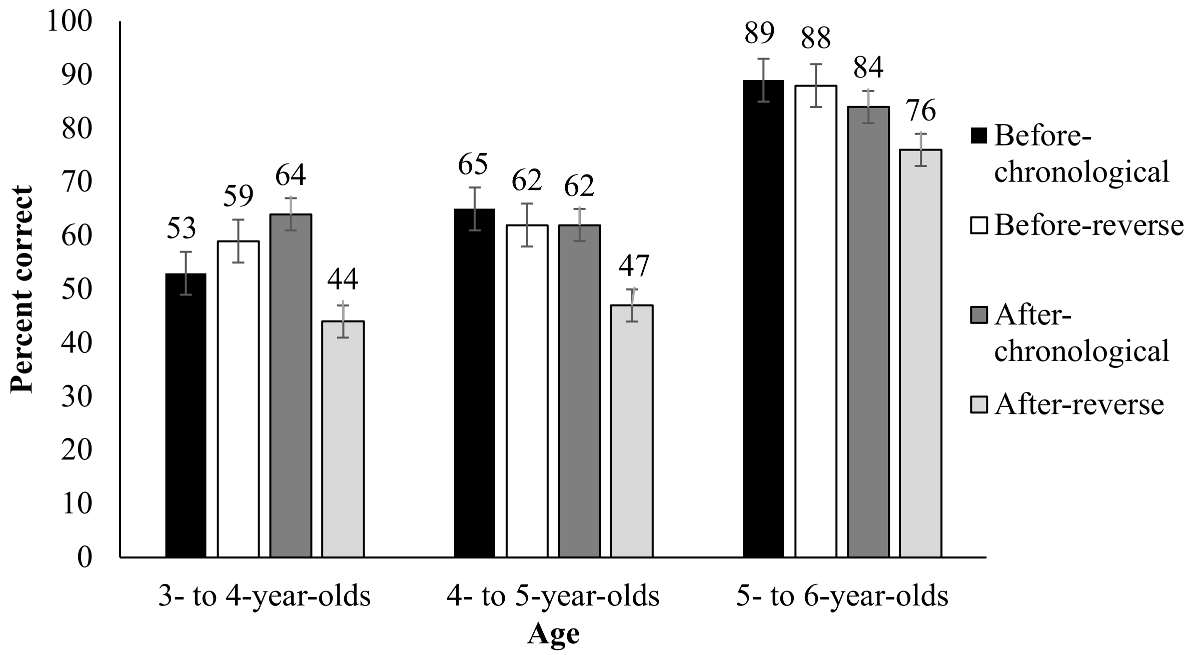
Vocabulary

The raw vocabulary scores [mean (SD)] demonstrated age-related improvements: 3- to 4-year-olds = 72.65 (26.16); 4- to 5-year-olds = 78.26 (9.76); and 5- to 6-year-olds = 102.30 (8.59). All children had a standardised score above 85 and the mean scores indicated 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); and 5- to 6-year-olds = 101.54 (9.10).

Analysis of accuracy data

A total of 2144 responses were recorded. Figure 1 shows the means for each sentence structure by age in years for ease of comparison (note that the analyses were conducted using age in months as a continuous variable). 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 (nonsense or no response), indicating that children understood the purpose of the task.

Figure 1. Mean percentage of target responses (with standard error bars) for each experimental condition in the sentence repetition task by age group.



The initial model included the main predictors of age, order and connective. The inferential statistics, main effects and interactions are summarised in Table A.3 (see Appendix). Response accuracy was significantly affected by age, indicating that performance improved between 3 and 6 years. There was also a significant effect of order, such that children were more likely to repeat chronological sentences accurately than reverse sentences. A significant effect of connective was also found: children were more likely to repeat sentences containing *before* accurately than those containing *after*. Order and connective were involved in a significant two-way interaction, which was examined by conducting simple interaction analyses of the effects of order for each connective separately.

A main effect of order was evident for *after* sentences, but not for *before* sentences. 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.

The final model incorporated memory and vocabulary as additional factors to age, order and connective (see Table 2). In comparison to the initial model, log-likelihood tests indicated that the fit of the data was significantly improved when we incorporated memory alone [$\chi 2(4) = 20.01$, $p < .01$], vocabulary alone [$\chi 2(4) = 12.67$, $p = .01$], and when memory and vocabulary were incorporated together [$\chi 2(8) = 29.21$, $p < .01$]. Memory and vocabulary both significantly influenced performance, such that stronger sets of skills in both domains improved performance. The main effects of order and connective remained significant and were again involved in a significant two-way interaction. There was no three-way interaction with either memory or with vocabulary.

Table 2

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

Main model	***M* (*b*)**	***SE***	***t***	**CI**		***p***
				2.5%	**97.5%**	
(Intercept)	.35	.22	1.56	-.09	.79	.12
Age	.07	.31	.24	-.54	.68	.81
Memory	**1.07**	**.28**	**3.82**	**.52**	**1.62**	**<.01**
Vocabulary	**.64**	**.31**	**2.10**	**.04**	**1.25**	**.04**
Order	**1.19**	**.23**	**5.28**	**.75**	**1.63**	**<.01**
Connective	**.79**	**.24**	**3.30**	**.32**	**1.26**	**<.01**
Order:Connective	**-1.04**	**.26**	**-3.93**	**-1.55**	**-.52**	**<.01**
Age:Order	.15	.30	.49	-.44	.73	.63
Age:Connective	-.14	.33	-.42	-.78	.51	.67
Memory:Order	-.16	.28	-.56	-.71	.39	.58
Memory:Connective	-.51	.30	-1.71	-1.09	.07	.09
Vocabulary:Order	-.20	.29	-.69	-.77	.37	.49
Vocabulary:Connective	.18	.32	.54	-.46	.81	.59
Age:Order:Connective	.33	.35	.93	-.37	1.02	.35
Memory:Order:Connective	.33	.33	1.01	-.31	.97	.31
Vocabulary:Order:Connective	.01	.35	.04	-.66	.69	.97
Random Effects					Variance	*SD*
Subject: (intercept)					1.14	1.07
Subject: (slope) connective					.71	.84
Subject: (slope) order					.74	.86

Note: 1. Bold = predictor is significant at $p<.05$ or lower.2. Number of observations = 2124; groups = 67 participants.

Analysis of non-target responses

The frequency of different types of non-target responses was investigated to determine whether particular types were associated with specific experimental conditions (sentence constructions) and/or age group. This provided an opportunity to examine additional support for either the memory or language accounts, outlined in the Introduction. We excluded responses from the oldest age group, because their high accuracy scores left too few non-target responses for meaningful analysis (120 out of all 702 non-target responses, 17%). The two youngest age groups made 582 non-target responses. The majority of non-target responses involved a change of sense to the meaning of the target sentence (*sense changed* = 358: 61% of all non-target responses analysed). Fewer non-target responses maintained the sentence meaning (*sense maintained* = 131: 23%) or were incomplete (*incomplete responses* = 93: 16%). These three categories of non-target responses did not vary substantially by experimental condition, although it is worth noting that sense changed errors made up a higher proportion of 3- to 4-year-olds' (66%) non-target responses compared with those of 4- to 5-year-olds (57%).

To further examine non-target responses, we calculated the percentage of *sense changed* responses that involved a change of connective, order, or position. A change of connective was the most common (252; 70% of all 358 *sense-changed* responses); there were far fewer position (109; 30%) and order (78; 22%) changes. These values add up to more than 100% because the non-target response types are not mutually exclusive; children could include more than one of these changes in their response. We conducted a further analysis to examine the most common type of *sense changed* response: those involving a change to the target connective (252). Of these, we excluded 34 responses involving a change to the target connective other than *before* and *after* (e.g., *then, and then, when*). This was because children's use of other connectives to communicate a non-target order of events was not a clear indicator for a weak representation of *before* or *after*. The 218 remaining changes to the connective were explicit demonstrations of producing *before* or *after* to communicate a non-target event order: *before* instead of *after* (e.g., *He put on the socks, before he ate the burger* instead of the target *He put on the socks, after he ate the burger*), or *after* instead of *before* (e.g., *He put on the socks, after he ate the burger* instead of the target *He put on the socks, before he ate the burger*). The most obvious reason for why children make these changes is that they have a weak representation of the precise meaning of *before* or *after*.

We examined the percentage of the total non-target responses (582) in each experimental condition (i.e., age, order, connective) that was caused by a *sense changed* response involving a change of *before* instead of *after* or *after* instead of *before* (218: 37% of all non-target responses). The Appendix materials provide descriptive statistics (Table A.4) and a summary of the GLMM (Baayen et al., 2008) for this analysis (Table A.5). The changes were involved in a significantly larger percentage of the non-target responses for reverse order sentences (44%) compared to chronological sentences (30%). The changes were more frequent for the younger than older age group (3- to 4-year-olds = 40%; 4- to 5-year-olds = 35%), but the difference was not significant. Similarly, although these changes were less common for *before* than for *after* sentences (*before* = 33%; *after* = 41%), the difference was not significant. Finally, although these changes were most common for after-reverse sentences, the interaction between connective and order was not significant (before-chronological = 29%; after-chronological = 31%; before-reverse = 37%; after-reverse = 48%). Note that Table A.4 also provides descriptive statistics to show that a similar pattern is present for the percentage of the total non-target responses (582) in each experimental condition that was caused by *all* changes of connective (i.e., *sense maintained* and *sense changed* responses, and also inclusive of changes to *then, and then, when*). This pattern (described above) was not evident for non-target responses that involved a change to order, or a change to position.

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 (no more than 8% of all responses in any age group). Our experimental manipulations demonstrated an influence of event order and connective on production accuracy. In addition, performance on independent measures of memory and vocabulary improved the overall fit of the model. These results do not provide unequivocal support for either the memory capacity-constrained account (Carpenter et al., 1994) or the language-based account (e.g., MacDonald, 2016) of sentence processing. When considered together with the analysis of the non-target responses, the results lend greater support for the language-based account, for the reasons discussed below.

Reverse order sentences are proposed to incur a greater memory load than chronological order sentences, because the speaker must produce the first occurring event as the second clause, which requires this information to be maintained in working memory during planning and production (Habets et al., 2008). Our participants were less accurate in producing reverse order sentences linked by *after* than those linked by *before*, demonstrating that this effect was specific to the connective. Independent measures of both memory and vocabulary improved the fit of the model. These findings lend greater support for the language-based account of sentence processing, namely that there is an indirect relation between memory and sentence processing which is modulated by language. The explanation for this is that young children's lexical representations for *after* are less precise and secure than those for *before,* because *after* is acquired later and 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* during language production, particularly when the event order is reversed. In that way, variation in language knowledge may lead to difficulties with sentence production, particularly for sentence structures that have a high processing load, such as those relating events in reverse chronological order.

Our findings do not rule out the alternative memory capacity-constrained account, because the independent measure of memory made a significant and independent contribution to the fit of our statistical model. However, our analysis of non-target response types provides additional support for the language-based account. Changes to the target connective that used *before* and *after* to communicate a non-target event order (*sense changed* responses) were more likely for reverse order than for chronological order sentences. These made up the majority of sense changed responses that involved a connective change (218 out of 252, 87%). The most obvious reason for why children make these changes is that they have a weak representation of the connective itself. This is supported by the main effect of connective type in our main analysis: children were less accurate at producing sentences containing *after* than *before,* in general. This analysis of non-target responses indicates that an inaccurate representation of the connective itself (as measured by a change in connective) does not provide the support needed for the planning and production of reverse order sentences. Also note that, whilst not significant, the descriptive statistics by sentence are in line with a language-based account, because an inaccurate representation of the connective influenced a greater percentage of the non-target responses to target after-reverse sentences (52%) than to the other target sentence constructions (ranging from 31% to 40%).

Experiment 2

A limitation with the sentence repetition paradigm used in Experiment 1 is that it places additional demands on memory compared with speech production, because the child has to store the just-heard sentence prior to production. For that reason, sentence repetition may not be the most sensitive task to differentiate the memory capacity and language-based accounts of children's and adults' sentence processing. Experiment 2 sought to test further these accounts using a different method to elicit sentence productions. We used a blocked design task comprising four blocked sets of items, each assessing the ability to produce one of the four target sentence constructions (e.g., Huttenlocher, Vasilyeva, & Shimpi, 2004). These blocked conditions were designed to complement Experiment 1 by minimising the contributions of sentence comprehension and memory associated with sentence repetition and maximising spontaneous production of sentences.

Method

Participants

A new sample of participants was recruited (N = 67): 23 3- to 4-year-olds (aged 3;8 to 4;11, 10 boys), 23 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 1. 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 1 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 performed first, 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 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 if children failed to produce a single target sentence. 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 accurately produce any of the target structures.

As in Experiment 1, the order in which 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 used the same criteria for scoring accuracy of responses and for categorising non-target responses as in Experiment 1. We did not analyse the time taken to start a response because this measure was found not to be sensitive in Experiment 1. Agreement between the coders was good for both scoring accuracy of responses (target vs non-target responses agreement = 99%; Cohen's $\kappa = .97$) and also for categorising non-target responses (Agreement = 96%; Cohen's $\kappa = .96$).

Results

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); and 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 available for 3- to 4-year-olds.

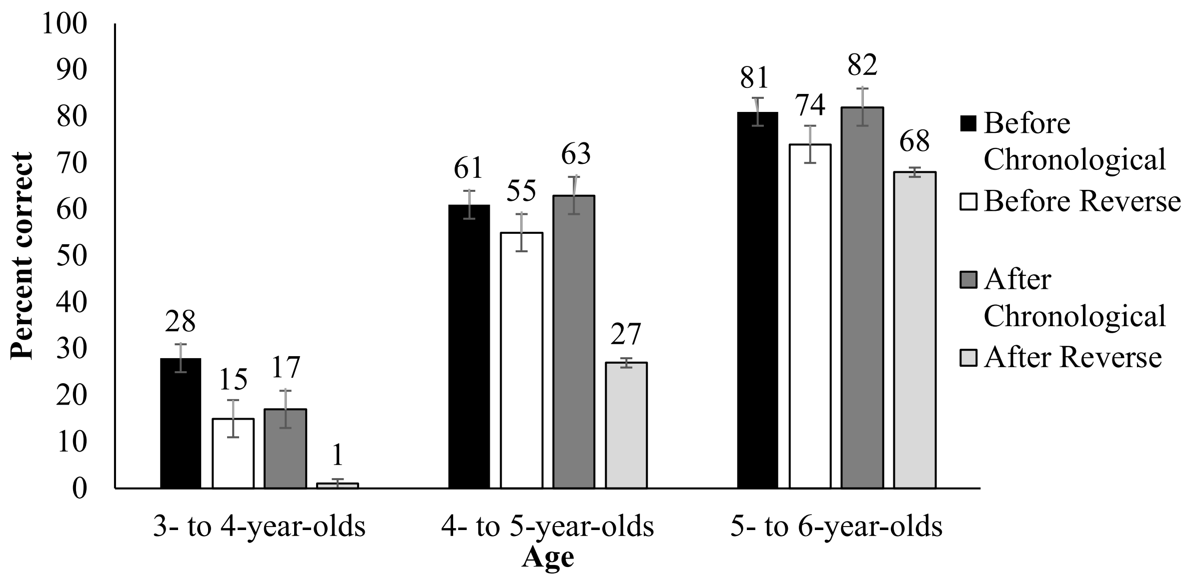
Vocabulary

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); and 5- to 6-year-olds = 90.95 (9.90). All children had a standardised score above 85 and the mean scores (SD) indicated 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 main analysis of the number of target responses was completed using the same procedures of model fitting described for Experiment 1. A total of 45 responses (2%) were excluded because they were inaudible or interrupted, leaving 1345 responses. Figure 2 reports the mean accuracy scores for each experimental condition by age group. Of note, performance for each age group was poorer than in Experiment 1, with the most marked difference in scores for the youngest age group.

Figure 2. Mean percentage of target responses (with standard error bars) for each experimental condition in the blocked elicited production paradigm by age group.



We report the initial model with age, order, and connective entered as fixed effects in the Appendix (Table A.6). Table 3 shows the final model that incorporates memory and vocabulary as additional factors to age, order and connective. For the initial model we found the main effects of age, order, and connective that were reported in Experiment 1. As predicted, older children produced a greater proportion of accurate responses than younger children, chronological order sentences were, in general, easier than reverse order sentences, and sentences containing *before* were easier than those containing *after*. There were two significant two-way interactions. The first between age and order was also apparent in Experiment 1; the other between age and connective, was not found for the sentence repetition task. These effects were qualified by a significant three-way interaction between age, order, and connective.

We examined the significant 3-way interaction by conducting simple interaction analyses for the effects of age and order for each connective separately. For *before* sentences, only the main effect of age reached statistical significance (see Table A.6 for a full breakdown of results and Figure 2 for graphs of these effects by sentence construction): Accuracy was equivalent for before-chronological and before-reverse sentences. For *after* sentences, there were main effects of age and order and these were also involved in a significant two-way interaction. Children found it more difficult to produce after-reverse sentences accurately than after-chronological sentences, and this difficulty with after-reverse sentences was more pronounced for the younger children.

We tested three additional models. The addition of memory to the original model significantly improved the fit of the data, $\chi2(4) = 20.11$, $p = .01$, and resulted in a significant three-way interaction between memory, order and connective. This suggests that memory modulated the interaction between connective and order. The memory alone model is reported in the Appendix (Table A.7). In another model, we added vocabulary to the original model and also found improved fit compared with the original model, $\chi2(8) = 33.57$, $p = .01$. In the final reported model (see Table 3), we included both vocabulary and memory. This resulted in improved fit compared with the memory alone model, $\chi2(4) = 12.08$, $p = .02$, and 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.

Analysis of non-target responses

Responses by the 5- to 6-year-olds were excluded because their high accuracy scores resulted in too few non-target responses for meaningful analysis (15% of all non-target responses by the three age groups; 152 out of 1019). We analysed the 867 non-target responses made by 3- to 4-year-olds and 4- to 5-year-olds. The *sense maintained* responses made up the highest percentage of responses (410; 47%), followed by incomplete responses (305, 35%), and then *sense changed* responses (152; 18%). These findings contrast with Experiment 1, in which *sense changed* responses made up the highest percentage of non-target responses. The different types of non-target responses did not vary significantly by experimental conditions, although 3- to 4-year-olds made a substantially greater number of incomplete responses than 4- to 5-year-olds (220; 42% vs 85; 25%, by age group respectively).

To further examine non-target response type, we calculated the percentage of *sense maintained* responses that involved a change to connective, order, or position. As noted in Experiment 1, these response types do not add up to 100% because more than one non-target change can be used in a single non-target response. Change in connective was the most common type of *sense maintained* response (313; 76% of *sense maintained* responses). In addition, position (237; 58%) and order (256; 62%) changes were also evident in over half of the total responses that maintained the sentence meaning. Of the 313 *sense maintained* responses involving a change of connective, only 129 were a change to the connective that involved the replacement of *before* for *after*, or *after* for *before*. Therefore, unlike Experiment 1, there were too few responses of this type for further analysis.

Table 3

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

Main model	***M* (*b*)**	***SE***	***t***	**CI**		***p***
				2.5%	**97.5%**	
(Intercept)	-3.49	.63	-5.58	-4.72	-2.27	<.01
Age	**3.48**	**.80**	**4.34**	**1.91**	**5.05**	**<.01**
Memory	-.48	.66	-.73	-1.78	.81	.47
Vocabulary	**2.08**	**.76**	**2.75**	**.60**	**3.57**	**<.01**
Order	**3.84**	**.63**	**6.12**	**2.61**	**5.07**	**<.01**
Connective	**2.80**	**.38**	**7.34**	**2.05**	**3.55**	**<.01**
Order:Connective	**-2.48**	**.44**	**-5.65**	**-3.34**	**-1.62**	**<.01**
Age:Order	**-1.77**	**.81**	**-2.18**	**-3.35**	**-.18**	**.03**
Age:Connective	**-2.63**	**.46**	**-5.70**	**-3.54**	**-1.73**	**<.01**
Memory:Order	.47	.67	.70	-.84	1.78	.48
Memory:Connective	**1.40**	**.38**	**3.71**	**.66**	**2.14**	**<.01**
Vocabulary:Order	-.76	.77	-.99	-2.26	.74	.32
Vocabulary:Connective	-.26	.37	-.70	-.99	.47	.49
Age:Order:Connective	**2.35**	**.54**	**4.33**	**1.28**	**3.41**	**<.01**
Memory:Order:Connective	-.76	.47	-1.60	-1.69	.17	.11
Vocabulary:Order:Connective	-.40	.47	-.85	-1.33	.52	.39
Random effects					Variance	*SD*
Subject: (intercept)					10.73	3.28
Subject: (slope) order					10.00	3.16

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

2. Number of observations = 1962; groups = 63 participants.

Discussion

The elicited production task complements the sentence repetition task used in Experiment 1, yielding complete two-clause sentences linked by an appropriate temporal connective from young children. As in Experiment 1, there was a main effect of connective, because *after* was more difficult than *before,* in general. In addition, there was a main effect of order because reverse order sentences were more difficult than chronological sentences. Also replicating Experiment 1 was the finding that children were least accurate when instructed to produce reverse order sentences linked by the connective *after*. That is, we again found that difficulty with reverse order sentences was modulated by connective: the effect was limited to after-reverse sentences. A critical difference between the two experiments was that production of after-reverse sentences was not modulated by children's working memory capacity in Experiment 2 when vocabulary was entered into our statistical model. Together, these findings suggest that language knowledge, rather than memory, is the stronger determiner of accurate sentence production.

Our analysis of non-target responses revealed a lower proportion of these involving a change of sense, compared with Experiment 1. It is important to note that there were few incomplete responses made by 5- to 6-year-olds (30; 5% of all responses) and 4- to 5-year-olds (85; 12% of all responses), although a third of 3- to 4-year-olds' responses were incomplete (220; 34% of all responses). This highlights the utility of the blocked elicitation paradigm to restrict speaker use to target sentence structures (Ambridge & Lieven, 2011).

General Discussion

These two experiments demonstrate that young children have difficulties producing two-clause sentences containing *before* and *after* in the developmental period that follows their emergence in spontaneous speech. In both experiments, children up to 6 years of age had particular difficulties producing reverse order sentences linked by the connective *after*. Clear developmental improvements were evident within this age range. Our experiments advance our understanding of the factors that influence young children's sentence production, demonstrating that memory capacity-constrained accounts of sentence processing need to factor in the influence of language knowledge rather than attribute difficulties to limited working memory capacity per se. We have also demonstrated that investigations that include more than a single paradigm are important to yield robust conclusions in the study of children's language production.

We tested two memory-based accounts for why some sentence structures are more difficult than others: a traditional memory capacity-constrained account (e.g., Carpenter et al., 1994) versus a more nuanced language-based perspective of working memory (e.g., MacDonald, 2016). Both accounts predict that reverse order sentences would be produced less accurately than chronological order sentences, in general. Our findings support this prediction with main effects of order evident in both experiments. According to the memory capacity-constrained account (Carpenter et al., 1994), this effect arises because these sentences require more units of information to be held active in working memory. Additional support for this account was evident: higher working memory capacity predicted better overall performance in Experiment 1. However, other findings indicate that such an interpretation cannot fully explain our findings, for the reasons discussed below.

Our results are in line with the language-based account of sentence processing proposed by MacDonald (2016) and others, in which the effects of working memory are not direct, but rather the result of its relation with language knowledge. The ability to represent information accurately in short-term memory is a requirement for good performance on a sentence production task. The language-based account proposes that short-term memory performance is influenced by the quality of language knowledge. We found support for this account in several ways. First, the effect of order was not consistent across connective: in both experiments, there was a significant order by connective two-way interaction, which was further qualified by age in a three-way interaction in Experiment 2. Critically, in both experiments, sentences with the connective *after* were less likely to be produced accurately in the reverse condition than in the chronological condition; this effect was not found for sentences with the connective *before*. These findings indicate a role for language, over and above, any memory effects. In addition, an independent measure of language ability explained performance over and above our independent measure of memory. Third, an inferential analysis of non-target responses in Experiment 1 indicated that a weak representation of our target connectives (measured by connective change responses) does not provide the support needed for the planning and production of reverse sentences, and descriptive statistics for the sentence constructions showed that this influence was most pronounced with target reverse sentences linked by *after.* Note that these findings together indicate that the influence of the connective is not explained by features of the language unit placing additional load on working memory per se; rather, findings are in line with a language-based account proposal that more processing resources are required to retrieve the context relevant meaning of *after* compared with *before,* so there are fewer processing resources available to accurately represent reverse sentences in active memory.

It is also worth noting that if children had displayed a low accuracy for *before* sentences in the reverse condition relative to the chronological condition, such a finding would not necessarily have opposed the proposal by a language-based account that accurate production is influenced by the quality of language knowledge (e.g., MacDonald, 2016). Specifically, even if children have a robust representation of *before*, planning and production of reverse sentences linked by *before* can be disrupted by a weak lexical representation for other words in the sentence. Though not significant, descriptive statistics indicated that accuracy for chronological and reverse sentences linked by *before* was equivalent for Experiment 1, but not for Experiment 2 (lower accuracy for before-reverse). Unlike Experiment 1, Experiment 2 did not provide children with the target words prior to their task to produce a target sentence. Only Experiment 2 reported that vocabulary had a greater influence over memory, which suggests that before-reverse sentences could be more difficult to produce when the task provides less support for words in the sentence. This in line with the proposal that a robust representation of words in the sentence frees up processing resources for the accurate representation of reverse sentences in active memory. Our sentence constructions were counterbalanced across conditions, and the effects were specific to *after*, which has a less consistent form-meaning relationship relative to *before*. Thus, we conclude that the accurate production of reverse sentences linked by *after* is more likely to be influenced by a weak representation for the target connective, rather than representing more general language effects.

Although a difficulty for after-reverse sentences was replicated across both experiments, there are at least two reasons to remain cautious about accepting a language-based explanation as the sole reason for young children's difficulties with multiple event sentence production. First, we must consider the possibility that order effects are modulated by a confounding variable, connective position, rather than connective. Second, we must address why the stronger influence of vocabulary over memory (determined by examining model fit) was apparent only in Experiment 2. These limitations are considered, in turn, below.

A natural consequence of our design was that the interaction between connective and order was influenced by connective position because this also differs across sentence structures. For example, *after* is used in a sentence initial position when the order of events is presented chronologically (*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 the 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; in our case the temporal connective, that act as cues to reduce ambiguity for the listener (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 working memory.

We believe that this account (that connective position rather than connective itself modulates order effects) does not adequately explain our pattern of findings. If position accounts for our results, the difficulty for after-reverse sentences would arise because the late signalling of reverse order places greater demands on memory capacity than early signalling. However, our independent measure of memory was a weaker predictor of performance than our independent measure of vocabulary. 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, 2004, 2008). Clearly, more experimental work is needed to investigate the role of connective position on sentence production.

The second reason for caution in accepting a language-based account over a memory capacity-constrained account was that both memory and vocabulary improved model fit in

Experiment 1. That is, stronger memory and vocabulary were both associated with more accurate performance, and our independent measure of vocabulary did not explain unique variance in children's specific difficulty with after-reverse sentences (Experiment 1). The greater influence of vocabulary over memory in Experiment 2 compared with Experiment 1 may have arisen due to the task demands. Participants in the sentence repetition task used in Experiment 1 were provided with the language form in their input, whereas participants in the elicited production task (Experiment 2) had to use their language knowledge to specify every level of detail of the form themselves (i.e., syntactic, morphological, phonological, and articulatory), so that it could be mapped onto the intended meaning (see Garrett, 1980; Gennari & MacDonald, 2009; Vigliocco & Hartsuiker, 2002). Therefore, there may be greater demands on language knowledge retrieval processes in the production task used in Experiment 2, in which children were not first provided with the input to repeat.

The above explanation may help to understand why the pattern of findings across sentence structures in these production experiments differs to that reported in recent work examining comprehension of the same sentences (Blything et al., 2015; Blything & Cain, 2016). Blything et al. found that reverse order sentences that contained *after* were the most difficult to comprehend, the same pattern reported here for production. However, in contrast to the findings of these production experiments, after-reverse sentences were not statistically more difficult to comprehend; instead an advantage for before-chronological sentences drove the effect. Further, in both of the previous comprehension studies, an independent measure of working memory accounted for significant variance in performance whereas an independent measure of vocabulary did not. This is in contrast to the present findings: our replication across two production studies of a difficulty with after-reverse sentences, in addition to stronger effects of an independent measure of vocabulary than of memory capacity, suggests a different explanation is required for production to that used for comprehension.

The difference between the findings of the present production study and previous comprehension studies can be explained by how comprehension and production draw on memory and language. Comprehension tasks provide the participant with the language form in their input in the same way as described earlier for sentence repetition tasks. Therefore, differences across the domains might be explained in the same way that was proposed above for why the present study provides greater support for the language-based account in the more pure production task (Experiment 2) relative to the production task that carried a comprehension component (Experiment 1).

Limitations, implications, and future research

A strength of this work was the replication of the main finding across two different tasks: children up to 6 years of age had difficulties producing two-clause reverse order sentences linked by the connective *after*. However, the analysis of non-target responses highlighted differences in the nature of our two experiments, which we believe is informative for researchers considering a marriage of the two paradigms. First, incomplete responses comprised 35% of the non-target responses in the elicited production task (Experiment 2), compared with only 16% in the sentence repetition task (Experiment 1). This may be due to the scaffolding provided by the initial input in a sentence repetition task, which supports the child to produce the target response. Another notable difference between the experiments was that *sense maintained* responses made up the largest percentage of non-target responses in elicited production (Experiment 2), whereas *sense changed* responses comprised a large percentage of non-target responses in the sentence repetition (Experiment 1). In *sense-maintained* responses, 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 resulted in a high number of *sense maintained* responses, is more likely to result in children reverting back to a sentence structure that they are familiar with, when required to signal temporal order with a connective. Overall, the difference in non-target response types, along with the differences in the nature of the tasks themselves, illustrates that investigations which include more than a single paradigm are important to yield robust conclusions in the study of children's language production.

Age differences in Experiment 2 persisted when memory and vocabulary were incorporated in the model. Given the high accuracy by 5- to 6-year-olds in Experiment 1, other experimental methods are required to study and understand better developmental and individual differences in the planning and production of complex sentences such as these. Habets et al. (2008) have successfully used ERPs to study processing differences in the production of chronological and reverse order temporal sentences, in adults. Such techniques might be adapted for use with children.

In addition to using different experimental paradigms to assess the time course and difficulty of production, a more comprehensive battery of tasks could be used to measure the constructs of both working memory and language. Ideally, working memory tasks should measure the storage and manipulation of information to tap these two critical functions of working memory (Oberauer & Lewandowsky, 2010). However, as noted, 5-year-olds find such complex span tasks hard to perform (Gathercole et al., 2004). In addition, memory tasks with a low semantic load should be used to determine the relationship between sentence processing and working memory capacity, distinct from language knowledge (Kidd, 2013). Digit based tasks as used here (forward digit recall) can be advantageous in this respect because they have a lower semantic load and so are less strongly related to independent measures of language (Cain, 2006; Seigneuric, Ehrlich, Oakhill, & Yuill, 2000; but see Jones & Macken, 2015). Similarly, additional measures of vocabulary as well as tests of grammatical knowledge could be included to assess more fully the construct of language and children's knowledge of cohesive devices such as connectives (LARRC, 2015; see also Cain

& Nash, 2011, for work with older children demonstrating differences between connective knowledge and use to at least 10 years of age). Note, however, that the measures used in the present study were predictive of performance, and these suggestions do not undermine the current findings; rather they offer ways to develop a more fine-grained picture of the influence of memory and language knowledge on the production of complex sentences.

A critical implication is that a memory capacity-constrained account of sentence processing (Carpenter et al., 1994) is likely too simplistic on its own and we need to factor in the influence of the specificity or distinctness of retrieval cues (i.e., language knowledge). Converging evidence for this viewpoint is has been provided in studies of adult language production (Gennari et al., 2012; Montag & MacDonald, 2014, 2015; Smith & Wheeldon, 2004) and comprehension (for review see Van Dyke & Shankweiler, 2012).

A next question for the language-based account is how language knowledge becomes sufficiently consolidated (precise and robust) to support the comprehension of complex sentences. A straightforward assumption from a developmental perspective is that language representations become stronger through exposure to the language. Thus, differences between vocabulary items, such as *before* and *after* may be due to differences in the frequency of their occurrence in language (Wells, Christiansen, Race, Acheson, & MacDonald, 2009). To explore this possibility, we coded 100 randomly selected occurrences of *before* and *after* from the CHILDES Thomas corpus (age range = 2;07.02–4;11.20; Lieven, Salomo, Tomasello, 2009), which is a corpus of child directed speech. Only 47 (of the 100) instances of *before* and *after* used these terms as a temporal connective within a multi-clause sentence. Of these, there was not a clear bias for either chronological or *before* sentences: there were 17 before-chronological, 4 before-reverse, 10 after-chronological, and 16 after-reverse sentences. This does not provide any evidence that children have less exposure to the more difficult (after-reverse) structure in this study. However, the corpus analysis did show that, of the other 53 occurrences of *before* and *after,* 47 were *after* used as a non-connective (e.g., *every now and then we look after the baby next door*). This finding supports an alternative account that the apparent difficulties with *after* sentences arises because *after* is used less consistently as a connective than *before.* This is consistent with the British National Corpus (Leech et al., 2001). Longitudinal work combining corpus and experimental methodologies could test this hypothesis further.

A final thought for future research is to what extent production accuracy might be enhanced when the sequence of events can be informed by world knowledge. Theoretical models of mental representations of text and discourse (e.g., Zwaan & Radvansky, 1998) suggest that it should be easier to plan and produce sentences when the events follow a typical sequence because world knowledge can inform the order that the events should be mentally represented, for example that *socks* are typically put on prior to putting on *shoes*. World knowledge-present sentences (e.g., *He put on the socks, before he put on the shoes*) served as fillers to scaffold the structure of the sentence, so were not part of our experimental design per se. Nevertheless, children's overall performance was consistent with previous findings in children's comprehension of two-clause sentences containing *before* and *after* (Blything et al., 2015): the filler world knowledge-present sentences were not performed significantly better than test sentences in which event order was arbitrary (e.g., *He put on the socks, before he ate the burger*). Thus, at least for these very simple two-clause sentences, world knowledge does not appear to play a significant role in language production. For more complex language, such as longer texts that require greater processing resources to integrate information across several sentences, world knowledge may have a more powerful influence on performance (Pratt, Tunmer, & Nesdale, 1989).

In conclusion, 3- to 6-year-olds demonstrated an ability to accurately use *before* and *after* as temporal connectives in the production of two-clause sentences, but notably found it difficult to produce reverse order sentences that were linked by *after.* We did not find unequivocal support for either the memory capacity-constrained account or the language-based account, although our findings lend greater support to the latter. These two apparently contrasting accounts have a common core: they seek to explain why memory limitations effect sentence processing. Further experimental work is needed to understand how memory and language knowledge individually and together influence sentence planning and production, and to elucidate the commonalities and differences in their influence on performance in language production and comprehension tasks.

References

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

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., 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

Baddeley, A. D. (2003). Working memory: Looking back and looking forward. *Nature Reviews Neuroscience, 4*, 829-839. doi:10.1038/nrn1201

Baddeley, A. D., & Hitch, G. J. (1974). Working memory. In G. H. Bower (Ed.), *The psychology of learning and motivation Vol. 8* (pp. 47–89). New York: Academic Press.

Barr, D. J., Levy, R., Scheepers, C., & Tilly, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, *68*, 255–278. doi: 10.1016/j.jml.2012.11.001

Bates, D. M., Maechler, M., & Bolker, B. (2014). *lme4: Linear mixed–effects models using S4 classes.* R package version 0.999999–0.

Blything, L. P., & Cain, K. (2016). Children's processing and comprehension of complex sentences containing temporal connectives: the influence of memory on the time course of accurate responses. *Developmental Psychology, 52*, 1517-1529. doi: 10.1037/dev0000201

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. (1987). Coordinating words and syntax in speech plans. In A. Ellis (Ed.), *Progress in the psychology of language Vol. 3* (pp. 337–90). London: Erlbaum.

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

Cain, K. (2006). Individual differences in children's memory and reading comprehension: an investigation of semantic and inhibitory deficits. *Memory, 14,* 553–569. doi: 10.1080/09658210600624481

Cain, K. & Nash, H. (2011). The influence of connectives on young readers' processing and comprehension of text. *Journal of Educational Psychology, 103*, 429-441. doi: 10.1037/a0022824

Carpenter, P. A., Miyake, A., & Just, M. A. (1994). *Working memory constraints in comprehension: Evidence from individual differences, aphasia, and aging*. San Diego, CA: Academic Press.

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

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

De Ruiter, L. E., Theakston, A. L ., Brandt, S., & Lieven, E. V. M. (2018). Iconicity affects children's comprehension of complex sentences: The role of semantics, clause order, input and individual differences. *Cognition, 171*, 202-224. doi: 10.1016/j.cognition.2017.10.015

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

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

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

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

Gennari, S. P., Mirković, J., & MacDonald, M. C. (2012). Animacy and competition in relative clause production: A cross- linguistic investigation. *Cognitive Psychology, 65*, 141–176. doi: 10.1016/j.cogpsych.2012.03.002

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.

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

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

Jones, G., & Macken, B. (2015). Questioning short-term memory and its measurement: Why digit span measures long-term associative learning. *Cognition, 144*, 1-13. doi: 10.1016/j.cognition.2015.07.009

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

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

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.

Levelt, W. J. M. (1989). *Speaking. From intention to articulation*. Cambridge, MA: MIT.

Lust, B., Flynn, S., & Foley, C. (1996). 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. (2016). Speak, Act, Remember: The language production basis of serial order and maintenance in verbal memory. *Current Directions in Psychological Science, 25*, 47-53. doi: 10.1177/0963721415620776

Martin, R. C., & Freedman, M. L. (2001). Short-term retention of lexical–semantic representations: Implications for speech production. *Memory, 9*, 261–280. doi: 10.1080/09658210143000173

Martin, R. C., Miller, M., & Vu, H. (2004). Lexical–semantic retention and speech production: Further evidence from normal and brain-damaged subjects for a phrasal scope of planning. *Cognitive Neuropsychology, 21*, 625–644. doi: 10.1080/02643290342000302

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.

Montag, J. L., & MacDonald, M. C. (2014). Visual salience modulates structure choice in relative clause production. *Language and Speech, 57*, 163–180. doi: 10.1177/0023830913495656

Montag, J. L., & MacDonald, M. C. (2015). Text exposure predicts spoken production of complex sentences in 8- and 12-year-old children and adults. *Journal of Experimental Psychology: General, 144*, 447-468. doi: 10.1037/xge0000054

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

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

Oberauer, K., & Lewandowsky, S. (2010). Modeling working memory: A computational implementation of the time-based resource sharing theory. *Psychonomic Bulletin & Review, 18*, 10–45. doi: 10.3758/s13423-010-0020-6.

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

Pickering, S. J., & Gathercole, S. E. (2001). *Working Memory Test Battery for Children*. Psychological Corporation UK.

Pratt, C., Tunmer, W. E., & Nesdale, A. R. (1989). Young children's evaluations of experience- and non-experience-based oral communications. *British Journal of Developmental Psychology, 7,* 83-91. doi: 10.1111/j.2044-835X.1989.tb00790.x

R Development Core Team (2014). *R: A language and environment for statistical computing.* Austria: R Foundation for Statistical Computing Vienna. ISBN 3-900051-07-0.

Seigneuric, A., Ehrlich, M. F., Oakhill, J. V., & Yuill, N. M. (2000). Working memory resources and children's reading comprehension. *Reading and Writing, 13*, 81–103. doi: 10.1023/A:1008088230941

Silva, M. 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

Slevc, L. R. (2011). Saying what's on your mind: Working memory effects on sentence production. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 37*, 1503-1514. doi: 10.1037/a0024350

Smith Micro Software (2012). Anime Studio Pro (Version 9.1) [Computer software]. Retrieved December 3, 2012. Available from http://anime.smithmicro.com/index.html

Smith, M. C., & Wheeldon, L. R. (2004). Horizontal information flow in spoken sentence production. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 30*, 675-686. doi: 10.1037/0278-7393.30.3.675

Van Dyke, J.A., & Shankweiler, D. (2012). 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.

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.

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

Zwaan, R. A., & Radvansky, G. A. (1998). Situation models in language comprehension and memory. *Psychological Bulletin, 123,* 162–185. doi: 10.1037/0033-2909.123.2.16

Table A.1

Mean (SD) proportion accuracy for filler sentences (world knowledge-present) each sentence type by 3- to 6-year-olds in (i) sentence repetition task and (ii) elicited production task.

	Sentence repetition			Elicited production		
	3- to 4	4- to 5	5- to 6	3- to 4	4- to 5	5- to 6
Before-chronological	.62 (.49)	.66 (.48)	.86 (.34)	.32 (.47)	.68 (.47)	.84 (.37)
Before-reverse	.58 (.50)	.59 (.49)	.89 (31)	.13 (.34)	.54 (.50)	.73 (.44)
After-chronological	.70 (.46)	.63 (48)	.89 (.31)	.21 (.41)	.64 (.48)	.82 (.38)
After-reverse	.46 (.50)	.42 (.50)	.72 (45)	.00 (.00)	.30 (.46)	.70 (.46)

Table A.2

Frequency counts of each individual non-target response type made by 3- to 5-year-olds in the sentence repetition and blocked elicited production task.

Non-target response 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
Juxtaposition of two clauses, 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		582	867

Table A.3

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

Main model	***M* (*b*)**	***SE***	***t***	**CI**		***p***
				2.5%	**97.5%**	
(Intercept)	.39	.25	1.55	-.10	.89	.12
Age	**1.05**	**.26**	**4.08**	**.55**	**1.56**	**<.01**
Order	**1.18**	**.22**	**5.47**	**.76**	**1.61**	**<.01**
Connective	**.76**	**.23**	**3.26**	**.31**	**1.22**	**<.01**
Age:Order	-.05	.22	-.21	-.47	.38	.83

Age:Connective	-.27	.24	-1.15	-.74	.19	.25
Order:Connective	**-1.03**	**.25**	**-4.08**	**-1.52**	**-.53**	**<.01**
Age:Order:Connective	.47	.26	1.85	-.03	.98	.06
Random effects					Variance	*SD*
Subject (intercept)					1.73	1.32
Subject: (slope) connective					.97	.98
Subject: (slope) order					.74	.86

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

2. Number of observations = 2124; groups = 67 participants.

Table A.4.

Frequency counts (with % contribution to all non-target responses by condition) by age, connective, order and sentence construction for (i.) 218 sense changed responses involving change of before instead of after or after instead of before, (ii.) all 354 changes of connective, (iii) all 165 changes of order, and (iv) all 177 changes of position in the sentence repetition task.

	Age		**Order**		**Connective**		**Sentence construction**			
	3 to 4	**4 to 5**	**Chronological**	**Reverse**	**Before**	**After**	**Before-chronological**	**After-chronological**	**Before-reverse**	**After-reverse**
Sense changed responses involving change of *before* instead of *after* or *after* instead of *before**	114 (40%)	104 (35%)	80 (30%)	138 (44%)	90 (33%)	128 (41%)	40 (29%)	40 (31%)	50 (37%)	88 (48%)
All changes of connective**	186 (65%)	168 (56%)	149 (56%)	205 (65%)	156 (57%)	198 (64%)	76 (55%)	73 (57%)	80 (60%)	125 (69%)

All changes of order**	89 (31%)	76 (26%)	75 (28%)	90 (28%)	81 (30%)	84 (27%)	35 (25%)	40 (31%)	46 (34%)	44 (24%)
All changes of position**	98 (35%)	79 (27%)	94 (35%)	83 (26%)	92 (34%)	85 (28%)	63 (45%)	31 (24%)	29 (22%)	54 (30%)
All non-target responses	284	298	266	316	273	309	139	127	134	182

Notes. 1. * = Non-target response types as reported GLMM, see Table A.5.

2. ** = S*ense maintained* and *sense changed* responses involving the respective change type (either all changes of connective, all changes of order, or all changes of position).

3. Despite low frequency counts, the same pattern of data was present for non-target responses by 5- to 6-year-olds. For example, sense changed responses involving a change of *before* instead of *after* or *after* instead of *before* made up a greater percentage of their total non-target responses for (i) reverse sentences (40/75= 53%) vs. chronological sentences (16/45 = 36%), and (ii) *after* sentences (36/52 = 53%) vs. *before* sentences (20/52 = 38%).

Table A.5

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

Main model	***M* (*b*)**	***SE***	***t***	**CI**		***p***
				2.5%	**97.5%**	
(Intercept)	.06	.30	.19	-.52	.64	.85
Age	.03	.30	.09	-.56	.62	.93
Order	**-.83**	**.36**	**-2.27**	**-1.54**	**-.11**	**.02**
Connective	-.33	.29	-1.14	-.90	.24	.25
Age:Order	-.05	.35	-.15	-.74	.63	.88
Age:Connective	.15	.27	.57	-.38	.69	.57
Order:Connective	.16	.45	.36	-.72	1.04	.72
Age:Order:Connective	<01	.41	-.01	-.82	.81	.99
Random effects					Variance	*SD*
Subject: (intercept)					2.05	1.43
Subject: (slope) order					1.09	1.12

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

2. Number of observations = 582; groups = 43 participants.

Table A.6

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***	**CI**		***p***
				2.5%	**97.5%**	
(Intercept)	-3.20	.60	-5.32	-4.38	-2.02	<.01
Age	**4.21**	**.68**	**6.19**	**2.88**	**5.55**	**<.01**
Order	**3.51**	**.58**	**6.05**	**2.37**	**4.64**	**<.01**
Connective	**2.45**	**.30**	**8.18**	**1.87**	**3.04**	**<.01**
Age:Order	**-1.70**	**.70**	**-2.43**	**-3.06**	**-.33**	**.02**
Age:Connective	**-1.73**	**.33**	**-5.33**	**-2.37**	**-1.10**	**<.01**
Order:Connective	**-2.13**	**.37**	**-5.82**	**-2.85**	**-1.41**	**<.01**
Age:Order:Connective	**1.36**	**.41**	**3.35**	**.56**	**2.16**	**<.01**
Random effects					Variance	SD
Subject: (intercept)					11.68	3.42
Subject: (slope) order					9.51	3.08
Before only						
(Intercept)	-.71	.89	-.79	-2.45	1.04	.43
Age	**4.62**	**1.20**	**3.85**	**2.27**	**6.98**	**<.01**
Order	1.99	1.38	1.44	-.72	4.69	.15
Age:Order	.88	1.85	.48	-2.74	4.50	.63
After only						
(Intercept)	-4.34	1.09	-3.99	-6.46	-2.21	<.01

Age	**4.38**	**1.04**	**4.22**	**2.34**	**6.41**	**<.01**
Order	**5.01**	**1.09**	**4.59**	**2.87**	**7.14**	**<.01**
Age:Order	-1.13	1.10	-1.03	-3.27	1.02	.30

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

2. Number of observations = 1962; groups = 63 participants.

Table A.7

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

	***M* (*b*)**	**SE**	***t***	**CI**		***p***
				2.5%	**97.5%**	
(Intercept)	-3.45	.65	-5.35	-4.72	-2.19	<.01
Age	**4.58**	**.82**	**5.59**	**2.97**	**6.18**	**<.01**
Memory	.01	.69	.01	-1.35	1.35	.99
Order	**3.75**	**.62**	**6.08**	**2.54**	**4.96**	**<.01**
Connective	**2.72**	**.34**	**7.89**	**2.04**	**3.40**	**<.01**
Order:Connective	**-2.34**	**.41**	**-5.77**	**-3.14**	**-1.55**	**<.01**
Age:Order	**-2.27**	**.81**	**-2.80**	**-3.86**	**-.68**	**.01**
Age:Connective	**-2.70**	**.46**	**-5.85**	**-3.61**	**-1.80**	**<.01**
Memory:Order	.38	.66	.58	-.91	1.68	.56
Memory:Connective	**1.38**	**.37**	**3.74**	**.66**	**2.11**	**<.01**
Age:Order:Connective	**2.20**	**.53**	**4.14**	**1.16**	**3.24**	**<.01**
Memory:Order:Connective	**-.99**	**.45**	**-2.21**	**-1.88**	**-.11**	**.03**
Random effects					Variance	*SD*
Subject: (intercept)					12.18	3.49
Subject: (slope) order					10.32	3.21

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

2. Number of observations = 1962; groups = 63 participants.