Interactions between givenness and clause order in children’s processing of complex sentences

Laura E. de Ruiter\(^{a,b,*}\), Elena V.M. Lieven\(^{b}\), Silke Brandt\(^{c}\), Anna L. Theakston\(^{b}\)

\(^{a}\) Tufts University, USA
\(^{b}\) University of Manchester, United Kingdom
\(^{c}\) Lancaster University, United Kingdom

**ARTICLE INFO**

**Keywords:**
Language development
Comprehension
Complex sentences
Context
Information structure

**ABSTRACT**

Understanding complex sentences that contain multiple clauses referring to events in the world and the relations between them is an important development in children’s language learning. A number of theoretical positions have suggested that factors like syntactic structure (clause order), iconicity (whether the order of clauses reflects the order of events), and givenness (whether information is shared between speakers) affect ease of comprehension. We tested these accounts by investigating how these factors interact in British English-speaking children’s comprehension of complex sentences with adverbial clauses (after, before, because, if), while controlling for language level, working memory and inhibitory control. 92 children in three age groups (4, 5 and 8 years) and 17 adults completed a picture selection task. Participants heard an initial context sentence, followed by a two-clause sentence which varied in: (1) the order of the main and subordinate clause; (2) the order of given and new information; and (3) whether the given information occurred in the main or subordinate clause. Accuracy and response times were measured. Our results showed that given-before-new improves comprehension for four- and five-year-olds, but only when the given information is in the initial subordinate clause (e.g., “Sue crawls on the floor. Before she crawls on the floor, she hops up and down”). Temporal adverbials (after, before) were processed faster than causal adverbials (because, if). These effects were not found for the eight-year-olds, whose performance was more similar to that of the adults. Providing a context sentence also improved performance compared to presenting the test sentences in isolation. We conclude that existing accounts based on ease of processing or information structure cannot fully account for these findings, and suggest a more integrated explanation which reflects children’s developing language and literacy skills.

1. Introduction

One of the most important differentiating features of our species is the ability to communicate complex, hypothetical and even counterfactual situations to one another. As we acquire language, we learn how to use the information in the sentences we hear to construct a mental representation of events they describe. This construction/interpretation process is influenced by a host of factors: what the listener already knows (background context), in what order the information is presented, and the processing load associated with the particular syntactic structure of the sentence. It is also affected by general cognitive factors, such as working memory and inhibitory control. Researchers have put forward different theories to explain how specific properties of the sentence, the discourse context, or the listener influence this process, but we do not yet have a model that brings these strands together. As most of these theories relate to only a subset of the factors known to influence sentence processing, the goal of the present study is to establish how these different factors interact.

We test how different linguistic factors (lexical item-, sentence- and discourse-level) and general language and cognitive skills (memory, inhibitory control) influence children’s ability to understand a complex situation from sentences containing adverbials like “Before he eats a green pear, he drinks some water” or “He sees the snowman because he opens the door”. Complex sentences like these are an ideal testbed for understanding how these processes interact, because they allow the simultaneous manipulation of several factors and help us tease apart their contributions to sentence processing. Complex sentences containing adverbials describe different relationships between two events (here: temporal or causal), they can occur in different structural configurations (clause orders), and they can be embedded into a discourse context (e.g., ”Tom eats a green pear. Before he eats a green pear, he drinks some water”), allowing us to test how previous knowledge (or

\(^{*}\) Corresponding author at: Tufts University, Eliot-Pearson Department of Child Study and Human Development, 105 College Ave., 02155 Medford, MA, USA.

E-mail address: laura.deruiter@tufts.edu (L.E. de Ruiter).

https://doi.org/10.1016/j.cognition.2019.104130

Received 2 October 2018; Received in revised form 30 October 2019; Accepted 5 November 2019

Available online 04 February 2020

0010-0277/ © 2020 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY license (http://creativecommons.org/licenses/BY/4.0/).
the lack thereof) may influence how easily a sentence can be processed.

In what follows, we will lay out what we know about how different factors affect (complex) sentence processing in adults and children and summarize specific hypotheses that have been put forward to explain these findings. We first take each major hypothesis separately before considering how they might relate to each other.

1.1. Syntactic structure

Different syntactic structures tax listeners’ processing capacities in different ways. A common assumption is that sentences are harder to process if they require the listener to keep information in working memory (Frazier & Clifton Jr, 1989; Klouender & Kutas, 1995; O’Grady, 2005) when two syntactically related but separate elements need to be combined. The longer the distance between two syntactically related elements in a sentence, the more difficult the sentence is to parse. Hawkins (1990, 1992, 1994) developed a theory of complexity that predicts the processing difficulty associated with a given linguistic structure. According to Hawkins, the parser recognizes the type of structure it is parsing and of a verb phrase. This is termed by Hawkins the ‘Early Immediate Constituent recognition domain”; that is, the string of words that the parser must process to recognize all the ‘daughter nodes (immediate constituents) of a verb phrase. This is termed by Hawkins the Early Immediate Constituents’ principle. Sentences are easier to process if the parser can quickly (based on fewer words) recognize the type of structure it is parsing and build a hierarchical representation of that structure. Diessel (2005) applied Hawkins’ theory specifically to complex sentences such as the ones that we are using in this study. The sentences in (1) illustrate Hawkins’ Early Immediate Constituents principle.

1. a. [[He drinks some water][before he eats a green pear]]
1. b. [[Before he eats a green pear][he drinks some water]]

At the top level of the parse, the complex sentences (1a) and (1b) have two immediate constituents: the main clause (main) and the subordinate clause (sub). Diessel (2005) suggested that listeners find these kinds of complex sentences easier to process if the main clause comes first. In (1a), there is initially no indication that this is a complex sentence. The parser can process the main clause fully. Only when it encounters the subordinating conjunction before, does it recognize a subordinate clause, and as a consequence realizes that it is dealing with a complex sentence, and not a simple one. The parser can construct the sentence and combine the subordinate clause with the already processed main clause right away. In (1b), on the other hand, the parser first encounters the connective before which immediately signals that the structure will need a main clause to be complete. Thus, the parser needs to keep the subordinate clause in memory, until the main clause can be accessed. The constituent recognition domain is long, which makes the structure harder to process. In short, complex sentences in main-subordinate orders are said to be easier to process, because they have a shorter recognition domain.

While it is beyond the scope of this article to go into more detail, it should be noted that Hawkins’ larger goal is to explain how language is shaped by processing more generally. In a similar vein, more recent versions of the generative paradigm have sought to incorporate performance constraints to account for grammatical choices (e.g., Trotzke, Bader, & Frazier, 2013).

On the basis of this theoretical perspective the prediction is that listeners – both children and adults – should find main-subordinate structures easier to process.

1.2. Givenness and presupposition

The notion of givenness is a core concept in the field of information structure, and has been extensively discussed and studied in the linguistic and psycholinguistic literature (e.g., Arnold, 2016; Gundel, Hedberg, & Zacharski, 1993; Lambrecht, 1996). Givenness has been equated with different concepts, such as activation, accessibility, familiarity, or salience, but is generally used to refer to entities or events that are, by one means or another, already established in the minds of the speaker and/or listener. In the context of this paper, we view givenness largely as activation in the mind of the listener. Givenness affects word order: in language production, both adults and children tend to place given referents earlier in the sentence than new referents (Ferreira & Yoshita, 2003; Levelt, 1989; Stephens, 2015), and it has been suggested that this given-before-new preference holds also in comprehension. For example, Haviland and Clark suggested that sentences are easier to process when given information precedes new information (H. H. Clark & Haviland, 1977; Haviland & Clark, 1974). The idea is that given information provides the listener with a “point of departure”, which has already been established in discourse, (Dahl, 1976: 38), and from which they can then add new information to update their mental model of the world. Confirming this, Birner and Ward (2009) analyse a variety of discourse and sentence structures including preposing, postponing and argument reversal in terms of the information structure constraints that are involved. They conclude that the default for argument ordering is given (old information) before new. Only when both arguments are old or both new, is either ordering felicitous.

Givenness has been shown to also interact with other factors such as weight of constituents (e.g., length), animacy, or the types of syntactic constructions involved. Arnold, Losoncgo, Wasow, and Ginstrom (2000) compared the relative importance of constituent length and newness on production in adult speakers. They concluded that there was an interaction such that the influence of either factor on the ordering of the constituents depended on their relative weights. If there was a big difference in length between the two constituents, this appeared to be the over-riding determinant of order, with the shorter constituent being placed before the longer (and heavier). But when there was relatively little difference in length, speakers preferred to place given information before new (see also Bresnan, Cueni, Nikitina, & Baayen, 2007).

For children, the picture is less clear. On the one hand, we know that very young children are sensitive to the given-new distinction in a variety of tasks. Weist (1983) found in an act-out task with Polish two- and three-year-olds that children made more errors when the referents occurred in new-given order. Similarly, Otsu (1994) found that Japanese-speaking four-year-olds understood non-canonical sentence structures (here: Object-Subject-Verb, OSV) better when the object was given in an act-out task. However, Smolik (2015) studied the comprehension by children aged between 2;9 and 4;5 of case-marked SVO and OVS sentences in Czech for which either the subject or the object was given. He found that all the children found SVO sentences easier than OVS sentences and that this improved with age and language proficiency. However, the information structure cues were not found to have a significant effect. These findings indicate that the information status (given vs. new) of discourse referents can affect children’s sentence comprehension but the accessibility and salience of other cues in the language may be more important at younger ages.

It is important to note that these studies were concerned with the givenness of individual referents (i.e., noun phrases that referred to concrete objects or persons), not with the givenness of whole events encoded by propositions. The only two studies that we are aware of that examined the effect of entire propositions (clauses) being given in adverbial sentences (i.e., propositional givenness) were conducted by Gorrell, Crain, and Fodor (1989), who looked at sentences with after and before, and Junge, Theakston, and Lieven (2015), who investigated the comprehension of when-sentences. Junge et al. tested three- and five-year-olds and adults with an act-out task in which participants had to perform the actions in both clauses one after the other. The authors manipulated the givenness of the clauses (i.e., whether the subordinate when-clause or the main clause was given) as well as the clause order
When the given information occurred second, irrespective of the clause order, all groups tended to perform the action in the given clause before that in the new clause. Clause order did not affect how children responded. The authors concluded that children are sensitive to discourse context, but not to syntax.

Gorrell et al. (1989) studied how givenness affected children’s comprehension of sentences with after and before. Their study was a follow-up to an earlier study by Amidon and Carey (1972), in which five-year-olds were asked to act-out commands like in (25).

(2) a. Move the red plane before you move the blue plane.
   b. Before you move the blue plane, move the red plane.

Amidon and Carey observed that children’s most frequent error was to fail to act out the action described by the subordinate clause; they would often only act out the main clause. They suggested that children simply tend to ignore the information in the subordinate clause. However, Gorrell et al. provided an alternative explanation: The sentences were infelicitous (i.e., inappropriate in the context). Gorrell et al. argued that subordinate clauses trigger presupposition. That is, the information contained in them will be treated as part of the shared common ground between speaker and listener (Geurts, 2017; Kriika, 2008). If the information is indeed part of the common ground, then the presupposition is satisfied, and the utterance is appropriate. If it is not part of the common ground, the presupposition is not satisfied, leading to presupposition failure and rendering the utterance inappropriate. Regarding the stimuli used by Amidon and Carey (1972), Gorrell et al. argued that using before or after would only be inappropriate when the presuppositions triggered by the subordinate clause were satisfied. For example, one would typically only utter a sentence like “Before you move the blue plane, move the red plane” to the child after it has been established that the child intends to move the blue plane. In other words, the clause “before you move the blue plane” presupposes “you will move the blue plane”.

They tested this hypothesis with five-year-olds using a modified version of the Amidon and Carey (1972) task. The crucial added manipulation was whether or not children were first asked which toy they wanted to push (e.g., the truck), thus making this action part of the common ground. Confirming their hypothesis, Gorrell et al. found that when the established action (e.g., pushing the truck) occurred in the subordinate clause (e.g., “After you push the truck, push the bus”), children performed better than when it occurred in the main (e.g., “Push the truck after you push the bus”). The authors conclude that already at an early age, children are “sensitive to proper contextual embedding of utterances” (p. 628), and that “[t]here is no evidence that the children (...) relied upon a structure-independent strategy of ‘old information precedes new information’” (p. 629). It is thus argued that what is crucial to children’s sentence processing is not the order of given and new information, but whether the given information is contained in the subordinate clause, in order to satisfy the presupposition. Indeed, the assumption that in adverbial sentences the subordinate clause typically expresses a presupposition while the main clause makes an assertion (i.e., constitutes new information) has been made by other linguists before (e.g., Bever, 1970; Haiman, 1978).

There are thus two competing hypotheses about how givenness and clause order affect children’s complex-sentence comprehension: According to Haviland and Clark’s (1974) given-before-new strategy, children, like adults, should find sentences easier to process when given information precedes new information. Junge et al.’s (2015) study on children and adults acting out when-sentences supports this hypothesis. We will call this the given-before-new hypothesis. In contrast, Gorrell et al. (1989) suggested that children find sentences easier to process when the given information is expressed by the subordinate clause, independent of clause order, and the findings of their own study with after- and before-sentences supported this assumption. We will refer to this as the given-in-subordinate hypothesis.

In the present study we test both hypotheses and investigate whether they equally apply to different types of adverbial clauses. Foreshadowing our results, it will become clear that both hypotheses are partially correct: We will see that both the clause order and the order of given and new information play a role in processing. This is almost certainly because syntactic order and information structure are not fully separate factors. Indeed Verstraete (2004) suggests that rather than seeing the order of main and subordinate clauses as a purely syntactic issue, the role of subordinate clauses should be analysed in terms of their informational function and the extent to which they are integrated into the scope of the main clause. He argues that initial subordinate clauses actually form a distinct construction, in that they are presupposed (‘discourse organising’) and are not within the local scope of the main clause. They cannot therefore be ‘at-issue’ (i.e. questioned or denied) whereas subordinate clauses in final position can hold new information and can also be questioned.

1.3. Iconicity

Language allows us to describe events in an order that is different from the order in which they occur in the real world. For example, in the sentence “Before she left the house, she ate a piece of cake” the action that happens later (leaving the house) is mentioned first. In contrast, in the sentence “She ate a piece of cake before she left the house”, the order in which the two actions are mentioned reflects the order of events in the real world. Sentences that map the event order directly are called iconic, while sentences that reverse it are called non-iconic.

Evidence is converging that children find non-iconic sentences more difficult to process than iconic ones. Clark (1971) was the first to suggest that young children assume that what they hear first happens first. She found that when she asked three- to five-year-olds to act out sentences like “Before she patted the horse, the girl jumped the gate”, error rates were much higher for non-iconic sentences than for iconic ones. Children would erroneously act out the putting action before the jumping action. Since Clark’s initial study, many more studies on after and before with different methodologies have been conducted, most of which confirmed its results (see De Ruiter, Theakston, Brandt, & Lieven, 2018, for an overview and references). De Ruiter et al. (2018) tested whether children assume iconic mapping also with other conjunctions, specifically with because and if. Note that when because- and if-sentences describe a causal relation between two events (e.g., falling and grazing one’s knee), the order can be either iconic (i.e., cause precedes effect, e.g., “Because/If he falls he grazes his knee”) or non-iconic (i.e., effect precedes cause, e.g., “He grazes his knee because/if he falls”).

Testing four- and five-year-old English-speaking children with a forced-choice picture selection task, they found this to be the case. While the four-year-olds’ performance was just above chance for all sentence types, five-year-olds performed better with iconic sentences. However, findings regarding the impact of iconic ordering on the comprehension of causal and conditional sentences are varied. Some studies have found an advantage for iconic ordering (e.g., De Ruiter et al., 2018), some have not (e.g., Corrigan, 1975) and another study found it to be task-dependent (Emerson, 1979).

Why children have difficulty with non-iconic sentences is not entirely clear. De Ruiter et al. (2018) suggest that this can be linked to children’s development in non-linguistic temporal-causal reasoning. Flexible temporal-causal reasoning develops around the age of five or six years (Lohse, Kalitschke, Ruthmann, & Rakoczy, 2015; McCormack & Hanley, 2011). Younger children’s mental representations of events are fragile and may be not be robust enough to reason about them in order to solve the task at hand (e.g., acting out, selecting a matching image). Another potential explanation lies in what has been called the attention-grammar interface. In language production, it has been shown that speakers’ attention influences how they conceptualize an event and the syntactic structure they use to describe it (e.g., whether to use an active or a passive form, Myachykov, Tomlin & Posner, 2005, Ibotson,
Lieven & Tomasello 2013). Similar processes may be at work in comprehension, such that the order in which children see events unfold biases their attention in ways that affect their success in comprehending non- iconic sentences.

While the present study is not primarily concerned with iconicity, clearly this may interact with information structure. As outlined above, the argument for the significance of given information preceding new is that the established background is already present from prior discourse and therefore the listener’s attention can proceed more easily to processing the information that is new. Iconic orders could also be seen as supporting the easier flow of attention. And, in turn, this may well interact with the order of main and subordinate clauses since, for after, if and because, the iconic order is for the subordinate clause to come first while the reverse is the case for before. However, it is not clear which of these potentially interacting factors might carry more weight, nor if this changes developmentally.

1.4. Adverbial type

The adverb before also appears to be easier to process in general. Clark first observed better performance with before compared to after in her (1971) original study, and the finding has since been replicated several times (Blything & Cain, 2016; Blything, Davies, & Cain, 2015; De Ruiter et al., 2018; Feagans, 1980; Stevenson & Pollitt, 1987), although some studies obtained different results (Amidon, 1976; Gorrell et al., 1989; Keller-Cohen, 1987). Clark explained this effect using a semantic feature account, which assumes that before is semantically less complex than after (see Clark, 1971, for details). De Ruiter et al. (2018) offer an alternative explanation, suggesting that before may be easier to process than after because, in English, it has a more consistent form-meaning relationship than other adverbials, making it easier for children to access the correct mapping. For example, after occurs more often in phrasal verbs like “looking after the children”. While there is evidence that consistent form-meaning mappings generally facilitate processing (e.g., MacWhinney & Bates, 1989), this specific hypothesis still needs to be tested in other languages that have different form-meaning distributions of adverbial conjunctions.

Irrespective of the underlying cause, we hypothesize that children will perform better with before-sentences in the present study as well.

1.5. Age and cognitive differences

Like other aspects of language comprehension, adverbial sentence comprehension improves as children get older (Amidon & Carey, 1972; Blything & Cain, 2016; Blything et al., 2015; E. V. Clark, 1971; De Ruiter et al., 2018). One would expect similar improvements when those sentences are presented with context, although context may boost levels of performance above those observed in the absence of context. Even within a given age group, children differ with respect to their general language skills, memory, and inhibitory control. Several language comprehension studies have found that these individual differences affect performance, in that children with better inhibitory control, better working memory, or larger vocabulary perform better (e.g., Blything et al., 2015; Brown-Schmidt, 2009; Kidd, Donnelly, & Christiansen, 2018; Nation, Marshall, & Altmann, 2003). While De Ruiter et al. (2018) did not find any evidence for an independent contribution of these skills to complex sentence comprehension, it is possible that children with better language and cognitive skills show an advantage when they have to process more speech material (as introduced by a context sentence, for example), or can take better advantage of contextual information to assist sentence processing.

1.6. The present study

The present study tests how the different linguistic and cognitive factors discussed in the introduction affect the processing of complex sentences in children of different ages and in an adult control group, when a prior context is included. The study is an extension of De Ruiter et al.’s (2018) study with complex sentences in isolation, using the same methodology but, crucially, adding a context sentence to allow us to determine if and how the different factors known to affect sentence interpretation interact. This is an important addition because only by assessing different factors simultaneously will it be possible to develop comprehensive theories of language processing. As in De Ruiter et al. (2018), the study focuses on the four adverbials after, before, because and if. The two temporal adverbials after and before were chosen because of their semantics (clearly indicating the temporal order of two events) and because they have received considerable attention in the experimental literature (see De Ruiter et al., 2018 for a review). The adverbials because and if were selected both because they indicate causal relationships, which allows testing the hypotheses beyond the frequently studied temporal adverbials, and because they are very frequent in child-directed speech (De Ruiter, Lemen, Brandt, Theakston, & Lieven, submitted).

We also took measures of children’s general language ability, memory, and executive control (inhibition) to assess the influence of individual differences on complex sentence comprehension. Based on our literature review, we formulated the following main hypotheses:

1. Main-subordinate orders are easier to process (greater accuracy and faster response times) than subordinate-main orders.
2. Given-before-new hypothesis: Sentences containing given information before new information are easier to process than sentences containing new information before given information.
3. Given-in-subordinate hypothesis: Sentences that contain the given information in the subordinate clause are easier to process.
4. Before-sentences are easier to process than other adverbial sentences.
5. Performance improves with age.
6. Individual differences (in general language ability, memory, and inhibitory control) influence children’s processing of adverbial sentences above and beyond the other factors.

Given that this is the first systematic investigation of this kind that tests all of these hypotheses together, a possible outcome is that they turn out to be only partially correct, because the various factors interact. For example, the type of adverbial could interact with information-structural factors (syntactic structure and givenness) to render some kinds of sentences easier to process than others. Indeed, we found that before-sentences behaved differently from other adverbial types. The goal of the current study was therefore to determine the relations between the factors that have been proposed to influence complex sentence comprehension, and identify what kind of model can explain our findings.

This study has been pre-registered on the OSF framework. The registration form, which specifies research questions, hypotheses, planned sample size and statistical analyses can be viewed at: https://osf.io/7pw5j/?view_only=55be7a0667c540938dc7cd768762e03.

2. Materials and methods

2.1. Participants

We tested 92 children and 17 adult controls. The children were recruited through nurseries and primary schools in the Manchester area (North-West of England), and at the Manchester Museum. Prior informed consent was obtained from caregivers/parents. All children were monolingual, native speakers of English without any known history of speech or language problems or developmental delays. Of the 92 child participants, 40 were between 3.6 to 4.5 years old (M = 48 months, SD = 3 months, 17 girls), 40 were between 4.6 and 5.5 years old (M = 60 months, SD = 2.8 months, 22 girls), and 12 were
between 7;1 and 9;6 years old (M = 8;3, SD = 0;1, 7 girls). We will refer to the first group as the four-year-olds, the second group as the five-year-olds, and the third group as the eight-year-olds. Nine additional children were tested, but their data had to be excluded because they didn’t understand the task (five children), turned later out to be older than the targeted age range (three children), or because they did not want to continue after the warm-up (one child). In addition, one child chose not to do the second session, and one child did not do the digit-span (memory) task. For one child, the data for the dimensional change card sort (inhibition) were lost due to experimenter error. As mixed-effects models deal well with missing data, the data of these three participants were retained in the final data set. The adult participants (N = 17, M = 35 years, 13 women) were visitors to the Manchester Museum and students or staff members from the University of Manchester, and native speakers of English. One additional adult participant’s data were excluded because he was a non-native speaker of English. Due to a technical error, for two adult participants, the final trials were not recorded, resulting in the loss of two trials. Note that the initial study plan (see pre-registration) contained only four- and five-year-olds and adults. Because we later found that five-year-olds were far from adult-like, we tested an additional, smaller sample of eight-year-olds to get a more comprehensive picture of the developmental trend. For reasons of interpretation we present the eight-year-olds’ data together with the other data.

2.2. General procedure

The four- and five-year-old children were tested in a quiet area in their nurseries and primary schools. In addition to the sentence comprehension test, the children completed six tasks on general language ability and vocabulary, working and short-term memory, and inhibitory control, which are detailed below. The tasks were spread over two sessions on two separate (and typically consecutive) days. Each session lasted between 25 and 40 min. Children completed half of all items of the sentence comprehension task in session one, and the other half in session two. In each session, the children also completed one inhibition task, one general language task, and one memory task. With the exception of the first inhibition task (Flanker task), children always first completed the sentence comprehension task before doing the other tasks (see Appendix 1 for details). The eight-year-olds and the adults were tested in the Study area at the Manchester Museum and at the Child Study Centre at the University of Manchester. They only did the sentence comprehension task and the digit span (memory) task. They completed all items in one session, with a short break between two blocks. The allocation of trials across sessions and the experimental lists are described in Experimental lists (Section 2.3.5) below.

2.3. Sentence comprehension

We tested participants’ comprehension of adverbial sentences using a forced-choice picture-sequence selection task on a touch-screen (child participants) or using a gamepad (adult participants). The task was to select out of two picture sequences the one that matched an aurally presented sentence. We measured both response accuracy and response time. The materials and procedure are described in detail also De Ruiter et al. (2018), who investigated children’s comprehension of the same adverbial sentences in isolation.

2.3.1. Design

The planned design (as detailed in the pre-registration of this study, see link above) had four factors: one between-subjects factor (AgeGroup), and three within-subjects factors (adverbial Type, ClauseOrder, ClauseGiven), each with the following levels:

- **AgeGroup:** 4 years, 5 years
- **Type:** after, before, because, if
- **ClauseOrder:** main-subordinate, subordinate-main
- **ClauseGiven:** given main clause, given subordinate clause

The adult control data and the eight-year-olds’ data were analysed separately from the younger children’s data and thus did not include AgeGroup as a between-subjects factor. This is because we did expect adults’ performance to be at ceiling. Note that we will still present all the data together for ease of comparison.

2.3.2. Materials

2.3.2.1. Audio stimuli. First, 24 adverbial sentences were constructed, each containing a main and subordinate clause (see Appendix 1). There were six sentences for each of the four adverbials after, before, because, and if. All sentences occurred in both clause orders (main-subordinate and subordinate-main), resulting in 48 sentences. These were the same sentences that were used in De Ruiter et al. (2018). The two clauses described two actions performed by a single actor (a boy in half of the sentences, and a girl in the other half), such as “Before he eats a green pear, he drinks some water”. The because- and if-sentences always expressed a physical causal relationship between the two events (e.g., opening a door and seeing a snowman outside). The subject of the sentence was always expressed as a pronoun (i.e., he or she), and all verbs were in present tense. All sentences were between 11 and 13 syllables long. In addition, we manipulated the givenness of the individual clauses by adding a context sentence that preceded the adverbial sentence. The context sentence expressed the proposition of either the main (given main clause) or the subordinate clause (given subordinate clause), but without the conjunction, giving 96 stimuli overall. Thus, the design was a $2 \times 2$ design crossing ClauseGiven (main clause given: “GivenMain” vs. subordinate clause given: “GivenSub”) and ClauseOrder (“Main-sub” vs. “Sub-main”). Table 1 shows all four versions for the sentence “Before he eats a green pear, he drinks some water” to illustrate temporal sentences, and for the sentence “She hears the doorbell, because she presses the button” to illustrate causal sentences. A list of all sentences is provided in Appendix 1. The sentences were spoken by a female native speaker of British English. There were always 1000 ms between the context sentence and the adverbial sentence, and 250 ms between the two clauses of the adverbial sentence.

2.3.2.2. Visual stimuli. For each audio stimulus, two picture sequences were created (for an example, see Table 2, showing the two actions expressed by the sentence in both orders, in left-to-right orientation, which is the convention in English picture books). For the sentences containing before and after, the second picture sequence was the reversal of the pictures of the first picture sequence. This was not possible for the sentences containing because and if, since the semantics of these sentences require there to be some change of state involved. For example, in the sequence matching the sentence “Because she presses the button, she hears the doorbell”, the actor is outside a house, where she first presses a button (the doorbell) and then hears the doorbell ring. The other sequence has to offer a plausible scenario for the opposite order of events (i.e., first hearing, then pressing) in order to be an acceptable distractor. In this case, the actor was depicted inside a house, first hearing the doorbell ring, and then pressing the button (at the intercom in order to open the door). The stimuli were created using the software Anime Pro (version 9.1).

2.3.3. Presentation

All stimuli were presented using the software E-Prime (versions 1.2 and 2). For children, the stimuli were presented on a laptop with a 14-inch resistive touch-screen, E-Prime version 1.2, and the sound was presented via loudspeakers. For the eight-year-olds, the stimuli were presented on a laptop with a 13.3-inch resistive touch-screen, E-Prime Version 2; the sound was presented via loudspeakers. The adults did the experiment on a 14-inch laptop using a game pad controller, E-Prime version 2, with sound presented via headphones.
the experiment when not selecting a sequence. The children were told that they were going to play a game, in which a lady was telling them stories about two characters, Sue and Tom, and about some animals (these were the filler trials, see below), and that they had to select from two picture stories the one that matched the story that they had heard. The children were instructed to listen carefully and touch the matching sequence after they hear a beep. The children heard each sentence twice, followed by a beep. Once the children had selected a sequence, the screen showed a blue circle to indicate that the trial had been successfully completed. The structure of one experimental trial is shown in Table 2. Response time was measured from the offset of the beep. After every three experimental trials, there was a filler trial, in which the participant had to select only one of two pictures, in which an animal was performing different actions (e.g., “Lion is drying his hair”, where one picture shows a lion drying his hair and the other one the lion counting money).

Before the start of the actual experiment, there was a warm-up phase to familiarise the children with the task and the left-to-right reading of the picture sequences. In the warm-up, the experimenter controlled the second presentation of the target sentence (see below for details of the trial structure). This allowed her to explain the layout of the screen before playing the sentence again (e.g., “Here we see that Tom is doing two things in this story. First, he is watering his plants. And then he switches the light on”, while pointing to the appropriate picture). The first two warm-up trials were like the filler trials (i.e., simple sentences with only two pictures). The other warm-up trials (six) were like the experimental trials, except that the sentences were of the structure “First, …, then…” If in any of the warm-up trials a child did not choose the correct picture or picture sequence, feedback was given, and the trial was repeated up to two times. If the child still made the wrong selection, the experimenter proceeded to the experimental trials, but noted that the child had failed to complete the warm-up successfully. This was the case with five children, who were later excluded (see 1.9 above).

Adults followed the same procedure but used a gamepad instead of a touch-screen to select the sequence, and with a shorter warm-up phase with only two items (with picture sequences).

### 2.3.5. Experimental lists

Eight different experimental lists were constructed. Each list contained 62 trials (24 test sentences and seven fillers in each session). Overall, each participant heard 12 sentences with each of the four conjunctions (after, before, because, if). In addition, each participant heard half of the 48 test sentences in main-subordinate order, and half in subordinate-main order. Furthermore, for half of the sentences, the main clause was given, and for the other half the subordinate clause was given. Recall that each sentence had four possible versions. In order to avoid boredom and carry-over effects from the same item (e.g., “Before he eats a green pear, he drinks some water”), each participant saw only one version of an item in a given session, and another version in the other session. For example, a participant might have heard the version “Tom drinks some water. He drinks some water, after he eats a green pear” (GivenMain, Main-subordinate) in session 1, and in session 2 “Tom eats a green pear. Before he eats a green pear, he drinks some water” (GivenSub, Subordinate-main). Note that while in this example, the given information precedes the new information, for other items, the new information came first (e.g., “Sue draws a picture. She takes a bath after she draws a picture”).

Different lists were created by (a) distributing different versions of the same sentence as described above (e.g., in List 1 and 2), (b) swapping sessions (e.g., List 3 and 4 were the same as List 1 and 2, but with the order of sessions reversed), and (c) turning all after-sentences in Lists 1–4 into before-sentences for Lists 5–8 and vice versa, and all if-sentences into because-sentences and vice versa.

The order of the trials within each session was pseudo-randomised. There was a maximum of two consecutive trials in the same condition. The position of the correct picture sequence in session 1 was
counterbalanced, so that in half of the trials the correct picture sequence was at the top and in the other half of the trials at the bottom. In addition, the position of the correct picture sequence across sessions was counterbalanced, so that for any given item, when the correct picture was at the top in session 1, it was at the bottom in session 2, and vice versa.

Participants were randomly assigned to one of the eight lists.

2.3.6. Internal reliability

We measured the internal reliability (consistency) of the accuracy and response time measures using Cronbach’s alpha (Cronbach, 1951). It measures whether all items (which are supposed to measure the same underlying construct) produce similar scores. The higher the coefficient, the higher the internal consistency of the instrument (i.e., the experiment). For accuracy, Cronbach’s alpha (both raw and standardised) was 0.77, which is considered acceptable to good (Hair, Black, Babin, & Anderson, 2010; Mallory & George, 2000). For response times, Cronbach’s alpha (raw and standardised) was 0.93, which is considered excellent (ibid).

2.4. Language ability

We measured children’s receptive language ability using the “Linguistic Concepts” sub-test of the Clinical Evaluation of Language Fundamentals®-Preschool-2 (CELF; Wiig, Secord, & Semel, 2004), and the British Picture Vocabulary Scale III (BPVS; Dunn, Dunn, Styles, & Sewell, 2009). The sub-test “Linguistic Concepts” requires the child to follow directions of increasing length and complexity (e.g., “Point to either of the monkeys and all of the tigers.”). The BPVS tests children’s receptive vocabulary through a forced-choice picture selection from four illustrations (e.g., “Point to the castle”). The CELF sub-test took about 5 min, the BPVS test between 5 and 10 min. Cronbach’s alpha for the CELF sub-test is 0.70. Traditional reliability measures are not available for the BPVS; the manual instead reports standardised score uncertainty (Dunn et al., 2009).

2.5. Memory

Verbal short-term memory was tested with the “Sentence Imitation Test” (SIT) from the Early Repetition Battery® (ERB; Seeff-Gabriel, Chiat, & Roy, 2008). Short-term and working memory were tested using the forward and backward digit span task from the Wechsler Intelligence Scale for Children (Wechsler, 2003). Each task (SIT and digit span) took about 5 min. Cronbach’s alpha for the SIT has been reported to be 0.92, and the average split-half correlation for the digit span 0.85.

2.6. Inhibitory control

Children’s inhibitory control was tested using two tasks: the computer-based Flanker task from the National Institutes of Health (NIH) Examiner testing battery (Kramer et al., 2014, version 3.2.0.1) and the dimensional change card sort (DCCS) task (Zelazo, 2006). The Flanker task assesses a participant’s ability to inhibit the predominant response in the face of interfering stimuli. Children were asked to respond to the central target on the screen, ignoring the flanking distractor items by pressing one key on the keyboard for one type of stimulus (e.g., a left-facing fish), and another key for another type of stimulus (a right-facing fish). Cronbach’s alpha for the entire testing battery has been reported to be 0.77 (Heaton et al., 2014). In the DCCS task, children are required to sort a series of bivalent test cards, first (pre-switch phase) according to one dimension (colour), and then (post-switch phase) according to the other (shape). The task taps into children’s flexibility to switch their attention to a different dimension. The flanker task took between 5 and 10 min (48 trials), the DCCS task about 5 min (6 pre-switch and 6 post-switch trials). Cronbach’s alpha for the DCCS task has been reported to be between 0.90 and 0.93 (Moriguchi, Chevalier, & Zelazo, 2016).

2.7. Statistical analysis strategy

All analyses were done in R (version 3.3.2; R Core Team, 2016). We ran generalized linear mixed effect models (glmer; for analyses of accuracy, a binary outcome variable with the levels “incorrect” and “correct”, using the logit-link function) and linear mixed-effects models (lmer; for analyses of response times, a continuous outcome variable), using the lme4 package and the BOBYQA algorithm for optimization (Bates, Maechler, Bolker, & Walker, 2015). We used the R packages lmerTest (Kuznetsova, Brockhoff, & Bojesen Christensen, 2016) and pbkrtest (Halekoh & Hojsgaard, 2014) for the calculation of p-values for (g)lmer models. Confidence intervals for the coefficients were obtained using the confint.merMod function of the lme4 package. We added fixed and random effects incrementally to a minimal model, and tested if the inclusion of an additional term was justified using the likelihood ratio test for model comparisons (Pinheiro & Bates, 2000), and pruned non-significant effects, unless they were part of a significant interaction. By-participant random slopes were added using the same procedure. The binary factors AgeGroup, ClauseOrder, and ClauseGiven were coded using treatment contrast (the default coding in R). The reference level for AgeGroup was “four-year-olds”, for ClauseOrder it was “main-sub”, and for ClauseGiven “given main”. Type was coded using sum contrasts. In sum contrasts, each level is compared to the overall average of all levels (as opposed to one base level). We also ran simple correlations to test if performance was correlated with measures of general language ability, memory, and inhibitory control. When an individual measure was correlated with overall performance, we added it to the model and tested whether it explained any variance over and beyond the main factors listed above.

Furthermore, we used Bayesian statistical tests to complement the traditional analyses: Bayesian correlations, using the BayesFactor package (Morey, Rouder, & Jamil, 2015), and Bayesian mixed effects models using the rstan package with the rstanarm extension (Gabry & Goodrich, 2016). Bayesian mixed models are more appropriate than the originally planned Bayesian regression using the BayesFactor package. Bayesian mixed models use logistic regression, like the traditional frequentist (g)lmer models, but they were not readily available in R at the time of the pre-registration. For completeness, we report the results for the Bayesian regression (using the BayesFactor package), as originally planned in the pre-registration (https://osf.io/7pw5j/?view_only=55be7a0667c540938dc7cd7687622e03) in Appendix 2. We opted for these additional Bayesian analyses because of the problems involved with traditional null-hypothesis significance testing (NHST). NHST is designed to either reject the null hypothesis, or fail to do so. However, if the null hypothesis is true, p-values do not converge to any limit value, and all p-values are all equally likely (Rouder, Speckman, Sun, Morey, & Iverson, 2009). This means that we cannot infer from non-significant results that the null hypothesis is true (see e.g., Dienes, 2014). Bayesian analyses, in contrast, provide information about the relative strength of the statistical evidence for either the null or the alternative hypothesis.

For the Bayesian generalized linear mixed models we used weakly informative priors following a suggestion by Sorensen and Vasishth (2015) (Student t-distribution with two degrees of freedom, centred on zero, for intercepts and slopes; a so-called LKJ prior for random effects). We ran 2000 iterations (1000 as warm-up, 1000 to sample from the posterior distribution). To determine the optimal model, we used an approximation of the leave-one-out (loo) cross-validation provided in the rstanarm package, and compared different models with each other using the compare_models function (for details, see http://mc-stan.org/rstantarm/reference/loo.stanreg.html). For this type of analysis, we report the estimate of the mean, and the lower and upper limits of the 95% credible interval. Note that the Bayesian credible interval provides an interval in which the true value of the parameter lies with probability 1-α% (see Morey, Hoekstra, Rouder, Lee, & Wagenmakers, 2016 for a discussion of confidence intervals and credible intervals). When the credible interval does not contain zero, we interpret this as reliable evidence for an effect.
In order to determine if the individual language, memory, and inhibition scores in children were correlated with performance (i.e., accuracy and response times), we calculated traditional and Bayesian correlations between the individual differences measures (two language measures, two memory measures, two inhibition measures) and mean accuracy and mean response time, resulting in 12 (6 IVs × 2 DVs) correlations. For the verbal interpretation of Bayes factors, we used the evidence categories adapted from Jeffreys (1961), cited in Wetzel et al. (2011). For the Bayesian correlations, we used a prior of \( r = 0.5 \) (i.e., a prior with minimal assumption). Those measures that were significantly correlated with the outcome variables were then successively (by decreasing strength of correlation) entered into the optimal models (traditional mixed models, Bayesian mixed models) in order to test if they remain significant predictors in the children’s performance after the experimental factors, including age, have been accounted for. This approach was chosen, because we wanted to determine if the optimal model explaining performance variance due to the experimental factors would be improved if we also added measures of children’s individual differences in general language ability, memory, and inhibitory control.

For the adult control group, we tested the correlation of memory score (digit span) with accuracy and response times, respectively, and entered the scores into the models if there was a significant correlation.

3. Results and discussion

This study contains both confirmatory analyses and follow-up analyses. The confirmatory analyses are those specified in the pre-registration. The follow-up analyses were not pre-planned. Specifically, the study was originally planned with four-year-olds, five-year-olds, and adults only. We later also tested older children (eight-year-olds). For ease of reading, we present the results of the older age group together with the other data. In another follow-up analysis, we directly compared the results of the present study with those of De Ruiter et al. (2018) in order to assess the effect of presenting complex sentences with context compared to no context at all.

We first present the results for hypotheses 1–5, starting with the accuracy data (first traditional analysis, then Bayesian analysis), followed by the response time data (again first traditional analysis, then Bayesian analysis). We then provide an interim discussion (Section 3.3), before moving on to the results for the individual differences hypothesis (hypothesis 6).

3.1. Accuracy

We analysed 5160 responses. Fig. 1 shows the mean accuracy for the four-year-olds, the five-year-olds, the eight-year-olds, and the adult control group, averaged across conditions. The mean accuracy in the four-year-old group was 59%, and the five-year-olds’ mean accuracy was at 70%. The eight-year-olds’ mean accuracy was at 83%. Adults responded correctly in 98% of all trials.

Fig. 2 shows the mean accuracy all for all age groups for sentences in both clause orders (ClauseOrder: main-sub and sub-main) and both ClauseGiven conditions (given-main and given-sub).

3.1.1. Traditional (frequentist) analysis

Table 3 shows for the four- and five-year-olds the \( \chi^2 \) statistics representing the difference in deviance between successive models, as well as the \( p \)-values based on likelihood ratio test comparisons. The final glmer model contained the main factors AgeGroup, Type, ClauseOrder, and ClauseGiven, as well as the interaction of ClauseOrder and ClauseGiven, and the three-way interaction of Type, ClauseOrder, and ClauseGiven. The model included random intercepts for subjects and items, and by-subject slopes for Type and ClauseOrder. There was no significant interaction involving AgeGroup. The results of the final model are shown in Table 4. Model estimates of the random effects can be found in Appendix 3. Fig. 3 shows the mean accuracy for the four different types (after, before, because, if) by ClauseOrder and ClauseGiven, collapsed across both younger age groups.

For the eight-year-olds, Fig. 1 suggests that their accuracy pattern lies in between that of the five-year-olds and the adults. In particular, the difference in performance between main-sub orders with the subordinate clause given, and sub-main orders with the subordinate clause given (right panel of Fig. 2) seems to be smaller than in the two younger age groups, becoming more similar to the adults. Note that traditional analyses are not suitable for exploratory analyses, because they require a pre-set sample size; when the sampling plan is not set in advance – as is the case with the eight-year-old group –, \( p \)-values cannot be interpreted (Wagenmakers, 2007). We therefore analyze their data using a Bayesian mixed-effects model in the next section. For the adult control group, none of the factors (Type, ClauseOrder, ClauseGiven) were significant.

3.1.2. Bayesian analysis

The final Bayesian mixed-effects model for the four- and five-year-olds contained the fixed main factors AgeGroup, Type, ClauseOrder, and ClauseGiven, the interaction of ClauseOrder and ClauseGiven, and the three-way interaction of Type, ClauseOrder, and ClauseGiven. Random effects included in the model were random intercepts for Subject and Item, by-subject slopes for Type and ClauseOrder, and by-item slopes for ClauseOrder and ClauseGiven. Table 5 shows the output of this model. The estimates for the random effects can be found in the Appendix.

The results for the four-and five-year-olds are in line with the results of the frequentist analysis. The only difference is that the Bayesian analysis did not find evidence that performance with after-sentences in main-subordinate order was also worse when the main clause was given.

The models for the eight-year-olds and the adults were the same as for the two younger age groups: They contained the fixed factors Type, ClauseOrder, ClauseGiven, the interaction of ClauseOrder and ClauseGiven, and the interaction of Type, ClauseOrder, and ClauseGiven. The models included random intercepts for subjects and
items, as well as by-subject slopes for Type and ClauseOrder, and by-item slopes for ClauseOrder and ClauseGiven. There was no strong evidence for an effect of any of the factors (see Table A6 in Appendix 4).

For the adult control group, the Bayesian analysis also confirmed the frequentist analysis: There was no evidence that any of the factors had an effect on accuracy.

### 3.1.3. Interpretation of results

We now interpret the accuracy results for the four- and five-year-olds in the context of our hypotheses, outlined above, before interpreting the results for the older children and the adults. As shown in Table 4 (traditional analysis) and Table 5 (Bayesian analysis), there was no overall advantage for sentences in main-subordinate order (e.g., “He drinks some water, after he eats a green pear”), disconfirming hypothesis (1). There was a significant interaction between ClauseOrder and ClauseGiven, in that sentences in subordinate-main orders were easier to process when the subordinate clause contained the given information (e.g., “He eats a green pear. He drinks some water. He drinks some water, before he eats a green pear”). The data thus confirmed the hypothesis only partially. The competing hypothesis, which posits that sentences should be easier to process when the given information is contained in the subordinate clause (hypothesis 3), irrespective of clause order, was not fully confirmed, either, since ClauseGiven interacted with ClauseOrder. If the given-in-subordinate hypothesis was correct, ClauseOrder should not matter.

Against our prediction (hypothesis 4), performance with before-sentences overall was not better than with the other types (after, because, if) (see also Fig. 2). However, the results suggest that information structure affects different sentence types differently: There was an advantage for before-sentences in main-subordinate orders, irrespective of whether the main clause or the subordinate clause was given (main given: “He eats some water. He drinks some water, before he eats a green pear”; subordinate given: “He eats a green pear. He drinks some water. He drinks some water, before he eats a green pear”). For after- and because-sentences, in contrast, the reverse was true: performance with these types of sentences was significantly worse when they were in main-subordinate order, with the main clause given (e.g., “He drinks some water. He drinks some water, after he eats a green pear”). In addition, performance for after-sentences in main-subordinate order was also lower than in subordinate-main order when the subordinate clause was given (“He eats a green pear. He drinks some water, after he eats a green pear”).

In line with our predictions, five-year-olds’ performance was better than that of the four-year-olds (hypothesis 5). However, there was no indication that the two age groups processed information structure differently. There was no interaction between AgeGroup, ClauseOrder and ClauseGiven.

Taken together, the converging evidence from both analyses of the accuracy data confirmed the age effect (hypothesis 5) and information structure effects (hypotheses 2 and 3), although the data supported neither the given-before-new hypothesis (2) nor the given-in-subordinate hypothesis (3) unequivocally. They rather suggest that both hypotheses are partially correct. We will return to this issue in the discussion. In addition, there was evidence for type-dependent effects of information structure.

While results for the eight-year-olds are only exploratory and need to be confirmed with a larger sample, they indicate that children’s adverbial sentence comprehension develops substantially between five and eight years of age. Specifically, they suggest that their processing is not affected by information structure in the same way as the younger children, making the eight-year-olds’ data more similar, albeit still with a lower overall accuracy, to that of the adult control group, who did not show evidence of any effect.

### 3.2. Response times

For the analyses of RTs, only correct responses were analysed (N = 3692). From the correct responses, we removed outliers using the following criteria: For children, we excluded all responses that were shorter than 300 ms and longer than 20,000 ms (152 responses, 5.2% of the data), as it is unlikely that shorter or longer RTs reflect processing of the target stimuli. For adults, we excluded all responses that were shorter than 150 ms and longer than 6000 ms (106 responses, 14.2% of the data). Overall, 66.6% of the data from the full data set were included (56.6% of the 4-year-olds’ data, 67.1% of the 5-year-olds’ data, 75.5% of the eight-year-olds’ data, and 82.9% of the adult data). The RT data are visualised in Fig. 4.

#### 3.2.1. Traditional (frequentist) analysis

The four-year-olds took an average (mean) of 4960 ms to respond, the five-year-olds took 3533 ms on average, the eight-year-olds...
1974 ms, and the adults 819 ms. The final lmer model for four- and five-year-olds contained only the fixed main factors AgeGroup and Type. Neither ClauseOrder nor ClauseGiven produced significant effects, nor were there any significant interactions. In addition to random intercepts for subjects and items the model included by-subject slopes for Type. The model is shown in Table 6. The model estimates for the random effects can be found in Appendix 3.

3.2.2. Bayesian analysis

The final Bayesian mixed-effects model for the four- and five-year-olds contained the fixed main factors AgeGroup, Type, ClauseOrder, and ClauseGiven, the interaction of ClauseOrder and ClauseGiven, and the three-way interaction of Type, ClauseOrder, and ClauseGiven. Random effects included in the model were random intercepts for subjects and items, as well as by-subject slopes for Type and ClauseOrder, by-item slopes for ClauseOrder and ClauseGiven.

The Bayesian model showed the same effects as the traditional analysis: Response times were shorter for five-year-olds, and for after- and before-sentences compared to because- and if-sentences. Again, there was no evidence for effects of either ClauseOrder or ClauseGiven. The results from the Bayesian model are shown in Table 7.

The models for the eight-year-olds and the eight-year-olds contained the fixed factors Type, ClauseOrder, ClauseGiven, the interaction of ClauseOrder and ClauseGiven, and the interaction of Type, ClauseOrder, and ClauseGiven. The models included random intercepts for subjects and items, as well as by-subject slopes for Type and ClauseOrder, and by-item slopes for ClauseOrder and ClauseGiven. In the eight-year-olds, there was no strong evidence for an effect of any of the factors or their interactions (see Table A7). For the adult control group, the Bayesian analysis aligned with the frequentist analysis. As with the eight-year-old group, there was no evidence for any main effects or interactions.

3.3. Interim discussion

In the Introduction, we presented hypotheses about how syntactic structure, givenness, and sentence type affect complex sentence processing. Specifically, we presented two different hypotheses regarding the impact of information structure on young children’s processing of adverbial sentences. The given-before-new hypothesis predicts that children’s performance should be better with sentences in which given information precedes new information. The given-in-subordinate hypothesis predicts that children’s performance should be better with sentences in which the given information is contained in the subordinate clause. The accuracy data support neither of these two hypotheses. Rather, it appears that while given-before-new improves comprehension, it does so only when the given information is in the subordinate clause (e.g., “He eats a green pear. After he eats a green pear, he drinks some water”). In other words, sentences with initial given adverbial clauses seem to be easier to understand. In addition, children showed better understanding of before-sentences in main-subordinate order in general (e.g., “He drinks some water before he eats a green pear”). The response time data did not produce any evidence that processing speed (for correct responses) is affected by clause order or givenness in any way. Sentence type, on the other hand, had an effect: Children took longer to respond to because- and if-sentences than to after- and before-sentences. Our exploratory analysis of older children suggests that, by age eight, all of the effects (both on accuracy and on response times) seem to disappear. It appears that eight-year-olds are already similar to adults, in that neither information structure nor adverbial type affects their comprehension systematically in this construction. At the same time, their accuracy is still lower than that of adults, and they take longer to respond.

In the next section, we turn to the possible role of individual differences in children’s comprehension of adverbial sentences. We test if the measures of general language ability, memory, and inhibition are
3.4. Individual differences

We first present descriptive statistics for all other tests that were administered. We then test if any of the scores in the language, memory, and inhibitory control tasks are significantly correlated with mean accuracy and/or mean RTs. Those scores that are significantly correlated with these overall measures are then entered into the optimal statistical models obtained in the analyses above. Recall that the eight-year-olds and the adults only did the memory task (digit span), and we will discuss their data separately.

3.4.1. Descriptive statistics

The means and standard deviations of the (standardised) scores for the CELF “Linguistic Concepts” sub-test, the BPVS, the ERB Sentence Imitation task, the digit span task, the Flanker task, and the DCCS task (post-switch) for both age groups (are presented in Table 8. As the WISC Digit Span task is normed only for children six years and older, we only present the raw scores. The scores for the Flanker task include only those children who successfully completed it (24 four-year-olds, 38 five-year-olds). The Flanker score is calculated on the basis of both accuracy and speed. The maximum score is 10.

The means and standard deviations indicate that each group was performing at an age-appropriate level in all tasks, although both age groups scored slightly above average in the BPVS (the mean standard score is 100). The mean digitspan score for the eight-year-olds was 13.9 (SD = 5.1), for the adults it was 19.2 (SD = 3.2).

Inhibitory control was measured using the dimensional change card sort (DCCS) task. In the post-switch phase of the DCCS task, where a maximum of six correct trials are possible, four-year-olds achieved on average 3.6 correct (SD = 2.1), and five-year-olds 4.8 (SD = 2.3). It should be noted, however, that the means are not necessarily informative, because the distribution tends to be bi-modal – children get all trials either wrong or right – which was also the case here. While the four-year-olds were approximately split between 0 and 6 correct

---

**Table 4**

Results of the final linear mixed effect model for accuracy in the four- and five-year-olds. Significant effects are highlighted in bold. Confidence intervals were obtained using the confint.merMod function of the lme4 package. Note that because sum contrasts only allow n-1 contrasts to be determined at one time (i.e., three comparisons for the four-level variable Type), for readability we combined the output of two models in order to show all contrasts in one table. Number of observations: 3816; Subjects: 80; Items: 24.

| Estimate | 2.5% CI | 97.5% CI | Std. error | z-value | Pr(>|z|) |
|----------|---------|----------|------------|---------|---------|
| (Intercept) | 0.526 | 0.277 | 0.775 | 0.127 | 4.145 | 0.000 |
| AgeGroup5years | 0.351 | 0.105 | 0.597 | 0.126 | 2.794 | 0.005 |
| Type.sumafter | 0.234 | -0.067 | 0.535 | 0.154 | 1.522 | 0.128 |
| Type.sumbefore | -0.163 | -0.467 | 0.141 | 0.155 | -1.049 | 0.294 |
| Type.sumbecause | -0.031 | -0.326 | 0.264 | 0.151 | -0.206 | 0.837 |
| Type.sumif | -0.040 | -0.328 | 0.248 | 0.147 | -0.292 | 0.787 |
| ClauseOrderSub-main | 0.052 | 0.192 | 0.296 | 0.125 | 0.416 | 0.677 |
| ClauseGivenGivenSub | -0.345 | 0.549 | -0.141 | 0.104 | -3.315 | 0.001 |
| ClauseOrderSub-main:ClauseGivenGivenSub | 0.663 | 0.368 | 0.959 | 0.151 | 4.398 | 0.000 |
| Type.sumafter:ClauseOrderMain-sub:ClauseGivenGivenMain | -0.365 | -0.728 | -0.002 | 0.185 | -1.971 | 0.049 |
| Type.sumbefore:ClauseOrderMain-sub:ClauseGivenGivenMain | 0.804 | 0.433 | 1.176 | 0.190 | 4.242 | 0.000 |
| Type.sumbecause:ClauseOrderMain-sub:ClauseGivenGivenMain | -0.459 | -0.808 | -0.110 | 0.178 | -2.580 | 0.010 |
| Type.sumif:ClauseOrderMain-sub:ClauseGivenGivenMain | 0.019 | -0.418 | 0.572 | 0.180 | 0.104 | 0.914 |
| Type.sumbefore:ClauseOrderSub-main:ClauseGivenGivenMain | -0.161 | -0.525 | 0.203 | 0.186 | -0.865 | 0.387 |
| Type.sumbecause:ClauseOrderSub-main:ClauseGivenGivenMain | 0.283 | -0.079 | 0.645 | 0.185 | 1.532 | 0.126 |
| Type.sumif:ClauseOrderSub-main:ClauseGivenGivenMain | -0.056 | -0.409 | 0.298 | 0.180 | -0.309 | 0.757 |
| Type.sumafter:ClauseOrderSub-main:ClauseGivenGivenMain | -0.067 | -0.418 | 0.284 | 0.179 | -0.372 | 0.710 |
| Type.sumbefore:ClauseOrderMain-sub:ClauseGivenGivenSub | -0.382 | 0.753 | -0.033 | 0.184 | -2.138 | 0.033 |
| Type.sumbecause:ClauseOrderMain-sub:ClauseGivenGivenSub | 0.939 | 0.573 | 1.304 | 0.186 | 5.035 | 0.000 |
| Type.sumif:ClauseOrderMain-sub:ClauseGivenGivenSub | -0.254 | -0.602 | 0.094 | 0.178 | -1.428 | 0.153 |
| Type.sumbefore:ClauseOrderMain-sub:ClauseGivenGivenSub | -0.292 | -0.639 | 0.054 | 0.177 | -1.654 | 0.098 |

---

Fig. 3. Mean accuracy for the four different sub-ordinators by ClauseOrder and ClauseGiven, collapsed across four- and five-year-olds. The red dashed line indicates chance level. Error bars indicate standard error. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

---

L.E. de Ruiter, et al.
Cognition 198 (2020) 104130

11
Table 5

Results from the Bayesian linear mixed-effect model for accuracy showing estimates of the mean, and the lower and upper bounds of the credible interval. Credible intervals that do not contain zero are highlighted in bold. Note that because sum contrasts only allow n-1 contrasts to be determined at one time (i.e., three comparisons for the four-level variable Type), for readability we combined the output of two models in order to show all contrasts in one table. Number of observations: 3816; Subjects: 80; Items: 24.

| Mean (estimate) | 2.5% CI lower bound | 97.5% CI upper bound |
|-----------------|----------------------|----------------------|
| Intercept       | 0.4807               | 0.215                | 0.7651 |
| AgeGroup5years  | 0.4191               | 0.0202               | 0.8392 |
| Type.sumafter   | 0.2087               | -0.1075              | 0.5142 |
| Type.sumbefore  | -0.1999              | -0.4988              | 0.0858 |
| Type.sumbecause | -0.0105              | -0.2938              | 0.2759 |
| Type.sumif      | -0.0029              | -0.2779              | 0.2974 |
| ClauseOrderSub-main | 0.0278          | -0.2946              | 0.3653 |
| ClauseGivenGivenSub | -0.2772        | -0.6669              | -0.0689 |
| ClauseOrderSub-main:ClauseGivenGivenSub | 0.2725          | 0.3172               | 1.1411 |
| Type.sumafter:ClauseOrderMain-sub:ClauseGivenGivenMain | -0.3364 | -0.6814              | 0.0322 |
| Type.sumbefore:ClauseOrderMain-sub:ClauseGivenGivenMain | 0.8008          | 0.4319               | 1.1886 |
| Type.sumbecause:ClauseOrderMain-sub:ClauseGivenGivenMain | -0.4656 | -0.8122              | -0.1179 |
| Type.sumif:ClauseOrderMain-sub:ClauseGivenGivenMain | -0.001          | -0.3486              | 0.3504 |
| Type.sumbefore:ClauseOrderMain-sub:ClauseGivenGivenMain | -0.1487 | -0.5247              | 0.2031 |
| Type.sumbecause:ClauseOrderMain-sub:ClauseGivenGivenMain | 0.2694          | 0.091                | 0.6471 |
| Type.sumif:ClauseOrderMain-sub:ClauseGivenGivenMain | -0.0488         | -0.4138              | 0.3091 |
| Type.sumafter:ClauseOrderMain-sub:ClauseGivenGivenMain | -0.0693          | -0.4142              | 0.2898 |
| Type.sumbefore:ClauseOrderMain-sub:ClauseGivenGivenMain | 0.0381           | 0.3712               | -0.0092 |
| Type.sumbecause:ClauseOrderMain-sub:ClauseGivenGivenMain | -0.2595          | -0.6118              | 0.0876 |
| Type.sumif:ClauseOrderMain-sub:ClauseGivenGivenMain | -0.0309         | -0.6678              | 0.025 |
| AgeGroup5years:ClauseOrderMain-sub:ClauseGivenGivenMain | -0.0162          | -0.5012              | 0.4889 |

responses, the majority of the five-year-olds got all trials correct. Still, the mode for both age groups was 6 correct responses.

3.4.2. Correlation of individual measures with accuracy and RT

3.4.2.1. Correlation with accuracy. Table 9 shows the standard correlations and Bayesian correlations of the six measures with mean accuracy in the sentence-comprehension task for the four- and the five-year-olds.

For the standard correlations, of the six predictors, five were significantly positively correlated with mean accuracy: the CELF Linguistic Concepts score, the BPVS score, the ERB Sentence Imitation Score, the Flanker task, and the Digit Span score. The Bayesian analyses suggest ‘decisive evidence’ for a correlation between accuracy and the Flanker score, and ‘strong’ evidence for the correlations with the ERB Sentence Imitation Score and the Digit Span score, and ‘substantial’ evidence for the correlation with the CELF Linguistic Concepts Score. With a Bayes factor of 0.12, there was no significant correlation with mean accuracy. The correlation results for accuracy thus indicate that for the younger children, general language ability and memory are significantly correlated with performance, but not inhibitory abilities.

In the eight-year-olds’ data, the mean accuracy was not significantly correlated with mean accuracy (r = 0.413, t = 1.422, df = 10, p = .1824), and the Bayes factor of 0.52 indicates that there is moderate evidence for the absence of a correlation. Similarly, in the adult control group, there was no significant correlation between the Digit Span score and mean accuracy (r = -0.06533). The Bayesian analysis supported this: With a BF of 0.35, there was moderate evidence for no correlation. In the adult control group, there was no significant correlation between the Digit Span score and mean response times (r = -0.4173, t = 1.966, df = 14, p = .06533). The Bayesian analysis showed that there was not enough data to find evidence in favour of either the presence or the absence of a correlation (BF = 1.02).

3.4.3. Influence on mean accuracy and mean RT

On the basis of the results of the correlation tests, we entered the CELF Linguistic Concept Score, the BPVS score, the ERB Sentence Imitation Score, the Flanker score, and the Digit Span score successively (one by one, by decreasing strength of correlation) as predictors into the optimal model for the prediction of accuracy in the sentence comprehension task (see Section 3.1 above). Similarly, the CELF Linguistic Concepts score and the BPVS score were added to the optimal model for the prediction of response times in the sentence comprehension task (see Section 3.2 above).

Recall that hypothesis (6) states that these scores should make an independent contribution to performance. In other words, the predictors should remain significant even after all experimental factors, including AgeGroup, have been accounted for.

Of the five predictors added to the optimal accuracy model, only the BPVS score remained a significant predictor (χ² = 4.0783, df = 34, p = .043). The Bayesian analysis supported this (lower bound of credible interval: 0.010, upper bound of credible interval: 0.2812). Neither of the two predictors entered into the optimal response times model (CELF Linguistic Concepts score, BPVS score) remained significant, and there was no evidence for an effect in the Bayesian analysis, either.

Taken together, the data disconfirm hypothesis (6): We did not find evidence supporting the assumption that individual differences in language ability (apart from vocabulary knowledge), memory, or...
inhibition are significant predictors of accuracy and speed in adverbial sentence comprehension, after the experimental factors, including age, have been accounted for.  

3.5. Follow-up analysis: the effect of context

Before we discuss all of our results, we present one follow-up that was not part of the original pre-registered study design. In order to determine whether adding discourse context made a difference to children’s comprehension of adverbial sentences in general, we compared the performance of the four- and five-year-olds in the present

Table 6

| Estimate  | 2.5% CI     | 97.5% CI     | Std. Error | df | t-value | Pr(>|t|) |
|-----------|-------------|--------------|------------|----|----------|---------|
| (Intercept) | 4797.85     | 4190.1993    | 5405.5076  | 92.15 | 15.475 | <2e-16 |
| AgeGroup5years | −1079.02    | −1868.1021   | −289.9456  | 402.60 | 74.06  | 0.009064 |
| Type.sumafter | −607.41     | −908.9727    | −305.8454  | 58.66  | −3.948 | 0.000214 |
| Type.sumbefore | −705.79     | −988.0612    | −423.5109  | 58.39  | −4.901 | 7.95e-06 |
| Type.sumbecause | 724.27      | 428.6703     | 1019.8678  | 54.35  | 4.802  | 1.27e-05 |
| Type.suminf   | 588.93      | 292.9517     | 884.9004   | 54.78  | 3.900  | 0.000265 |

Number of observations: 2362, Subjects, 80; Items: 24.
study with that of the same age children in De Ruiter et al.’s (2018) study. As shown below, the participants in both studies were comparable in terms of their language, memory, and executive control development. This allowed us to compare both data sets directly, since the set-up in our study is identical to that of De Ruiter et al.’s study, with the exception of the added context sentence. Fig. 5 shows mean accuracy for the four- and five-year-olds. Asterisks indicate statistical significance; diamonds indicate at least substantial evidence (for the H₀, that is, no correlation if below 1/3, for the Hₐ if above 3).

Table 8
Means and standard deviations for the standard scores of the CELF Linguistic Concepts task, the BPVS task, the ERB Sentence Imitation task, the WISC Digit Span Task, the Flanker task and the DCCS post-switch phase for four-and five-year-olds.

| Task                      | AgeGroup | 4 years | Mean | SD | 5 years | Mean | SD |
|---------------------------|----------|---------|------|----|---------|------|----|
| CELF linguistic concepts  |          |         | 10   | 2.2|         | 9.2  | 2.9|
| BPVS                      |          |         | 108.5| 12.4| 108.1   | 10.7|
| ERB sentence imitation    |          |         | 100.6| 13.9| 102     | 15.2|
| WISC digit span           |          |         | 5.4  | 2.3|         | 8.1  | 2.7|
| Flanker                   |          |         | 4.5  | 1.7|         | 5.9  | 1.4|
| DCCS post-switch          |          |         | 3.6  | 2.6|         | 4.8  | 2.1|

Table 9
Correlation coefficients, t-values, degrees of freedom (df), probabilities (p), correlation coefficients (r) obtained through Bayesian tests, and Bayes factors (BF) for the correlations between standardised test scores (z-scores) and mean accuracy for the four- and five-year-olds. Asterisks indicate statistical significance; diamonds indicate at least substantial evidence (for the H₀, that is, no correlation if below 1/3, for the Hₐ if above 3).

| Task                      | Standard correlation | Bayesian correlation |
|---------------------------|----------------------|----------------------|
|                           | r        | t      | df     | p       | r       | BF     |
| CELF Linguistic Concepts  | 0.341    | 3.180  | 77     | 0.00212**| 0.334   | 9.732* |
| BPVS                      | 0.431    | 4.215  | 78     | 0.00007***| 0.422   | 241.065*|
| ERB Sentence Imitation    | 0.367    | 3.459  | 77     | 0.00089**| 0.359   | 21.664*|
| WISC Digit Span           | 0.366    | 3.434  | 76     | 0.00097**| 0.358   | 20.109*|
| Flanker                   | 0.278    | 2.243  | 60     | 0.02859* | 0.270   | 1.081  |
| DCCS post-switch          | 0.168    | 1.485  | 76     | 0.14161 | 0.163   | 0.261* |

Table 10
Correlation coefficients, t-values, degrees of freedom (df), probabilities (p), correlation coefficients (r) obtained through Bayesian tests, and Bayes factors (BF) for the correlations between standardised test scores (z-scores) and mean response times for the four- and five-year-olds. Asterisks indicate statistical significance; diamonds indicate at least substantial evidence (for the H₀, that is, no correlation if below 1/3, for the Hₐ if above 3).

| Task                      | Standard correlation | Bayesian correlation |
|---------------------------|----------------------|----------------------|
|                           | r        | t      | df     | p      | r     | BF     |
| CELF Linguistic Concepts  | −0.293   | −2.690 | 77     | 0.00875**| −0.288 | 2.710  |
| BPVS                      | −0.240   | −2.185 | 78     | 0.03189 | −0.236 | 0.871  |
| ERB Sentence Imitation    | −0.187   | −1.668 | 77     | 0.09942 | −0.183 | 0.342  |
| WISC Digit Span           | −0.184   | −1.631 | 76     | 0.10711 | −0.192 | 0.310  |
| Flanker                   | −0.173   | −1.359 | 60     | 0.17935 | −0.168 | 0.245* |
| DCCS post-switch          | −0.018   | −0.157 | 76     | 0.87553 | −0.017 | 0.990* |
In this study we tested how the order of given and new information (given-new vs. new-given) and syntactic structure (clause order: main-subordinate vs. subordinate-main) affect adverbial-sentence processing in four-, five-, and eight-year-old children and adults, using a forced-choice picture selection task. Specifically we tested two competing hypotheses: that sentences are easier to process when given information precedes new information (hypothesis 2) (H. H. Clark & Haviland, 1977; Haviland & Clark, 1974), and that sentences are easier to process when the given information is contained in the subordinate (adverbial) clause (hypothesis 3) (Gorrell et al., 1989).

In addition to these two main hypotheses, we also tested – motivated by previous findings and the literature – whether main-subordinate orders are generally easier to process than subordinate-main orders (hypothesis 1), whether before-sentences are easier than other types of adverbials (hypothesis 4), whether adverbial sentence comprehension improves with age (hypothesis 5), and to what extent adverbial sentence processing is related to general language ability, memory, and inhibitory control (hypothesis 6).

In addition, as the first systematic investigation of these factors, we also assessed whether givenness and clause order interact with adverbial type, or with age (i.e., whether information structure has a different impact on younger children’s processing than on older children’s processing).

We first discuss the results concerning the two competing main hypotheses, the given-before-new hypothesis, and the given-in-subordinate hypothesis.

### 3.6.1. Effects of givenness and clause order

The results suggest that both givenness and clause order influence young children’s comprehension of complex sentences, but the data support neither of the two hypotheses unequivocally. Children of both four and five years (but not eight years) comprehended sentences best when the given information was in the subordinate clause and the sentence was in subordinate-main order information (e.g., “He eats a green pear. After he eats a green pear, he drinks some water”). If the given-before-new hypothesis were correct, we would have expected children to perform better with sentences in which the initial clause is given, both for subordinate-main and for main-subordinate structures. If the given-in-subordinate hypothesis were correct, we would have expected children to perform better with sentences in which the subordinate clause is given, irrespective of clause order. By eight years, children’s processing does not seem to be affected by either givenness or clause order – just like that of adults.

Our results differ both from Gorrell et al.’s (1989) findings for before- and after-sentences, and from Junge et al.’s (2015) findings for when-sentences. Gorrell et al. (1989) found that children performed better when the given information was in the subordinate clause, irrespective of clause order (e.g., when hearing a sentence “Push the bus before/after you push the truck”, after the child had indicated they wanted to push the truck). Junge et al. (2015) found that children tended to change the order of actions to given-before-new (e.g., acting out pushing before washing upon hearing the sentence “The cat is washing when the dog is pushing”, after ‘pushing’ had been mentioned before), but did not pay attention to syntactic structure (i.e., whether the information was in the subordinate or the main clause). In addition to Junge et al. using a different adverbal (when) than the present study, we note that both studies used a different methodology (act-out), and had considerably lower statistical power. Gorrell et al. used a between-subjects design, with 14 children in each of the four conditions (spanning a wide age range from 3;4 to 5;10). Junge et al. tested 16 five-year-olds in their (within-subjects) study. A systematic comparison of the different methods with sufficient statistical power could help clarify to what extent the diverging results are due to task-effects.

With neither of the two main hypotheses being fully supported by the data, we need to look for other explanations for the results. Linguists have noted before that preposed adverbial clauses have special grammatical characteristics that set them apart from final adverbial clauses, and that they also tend to be used to fulfill a particular function in discourse. In fact, Verstraete (2004) argues “that adverbial clauses in initial position constitute a distinct construction type” (p. 820). We will focus here on their discourse function (for discussions of their grammatical characteristics, see e.g., Diessel, 2013; Verstraete, 2004). Diessel (2013) states that their basic function is “to present information that is pragmatically presupposed providing a thematic ground for new information asserted in subsequent clauses” (p. 343). The idea that subordinate clauses are the locus of presuppositions was already discussed in the Introduction. Recall that Gorrell et al. (1989) assumed that the given information ought to be contained in the subordinate (adverbial) clause in order to satisfy the presupposition. However, the claim here is more specific: that only initial adverbial clauses are presupposed (see also Ford, 1993; Givón, 1990), and that new information in subordinate clauses can be presented only when these are final. This claim aligns with other work in the pragmatics literature, which suggests that some types of sentence-final subordinate clauses (including non-restrictive relative clauses) can be at-issue and take the role of advancing the discourse in a way that main clause are typically expected to do (Jasinskaia, 2016; Loock, 2007). Presuppositions thus appear to arise in sentence-initial adverbial clauses, and they are satisfied when the information in the adverbial clause is already part of the common ground, or given. In the present study, this was the case.
when the sentences were in subordinate-main order, and the sub-
ordinate clause was given (e.g., “Tom eats a green pear. After he eats a
green pear, he drinks some water”). In the other condition in which the
subordinate clause was new and initial, the presupposition was thus not
satisfied, because the context sentence made the main clause given, and
not the subordinate clause (e.g., “Tom drinks some water. After he eats a
green pear, he drinks some water”).

When a presupposition is not satisfied, the hearer needs to update
their discourse model in order to make sense of what the sentence
stated (e.g., that it is apparently part of the common ground that Tom
eats a green pear). Speakers are then expected to repair the failure to
make sense of the sentence, by a mechanism called accommodation
(Heim, 1983; Lewis, 1979).

Research in experimental pragmatics has found that in adults, ac-
commodation slows down reading (Domaneschi & Di Paola, 2018;
Schwarz, 2007), and leads to slower response times and higher error
rates for example in verification tasks (e.g., Tian & Breheny, 2016).
Although this research has been concerned with presuppositions trig-
gered by certain lexical items such as focus particles (e.g., again, also),
change of state verbs (e.g., stop, regret), or negation, it suggests that
pragmatic accommodation is associated with processing costs in adults.
This same process may lead to higher error rates in children. Supporting
evidence for this hypothesis comes from research on young children’s
processing of negation. Children often make errors when interpreting
negated sentences in isolation (e.g., Kim, 1985; Nordmeyer & Frank,
2014). In contrast, when sentences are presented in pragmatically fe-
licitous contexts, children do not have problems in interpreting them
(Nordmeyer & Frank, 2018; Reuter, Feiman, & Snedeker, 2018). It has
been suggested that children make more errors with isolated negated
sentences, because these require pragmatic accommodation (Reuter et al.,
2018). If a new subordinate clause appears in initial position, this
may also require accommodation – the listener needs to infer that what
is stated in the subordinate clause is apparently part of the common
ground, although this has not been mentioned before. Thus, for sub-
ordinate-main sentences, we would expect children to perform less well
with sentences in which the main clause is given (when accommodation
is required), and conversely, to perform well when the subordinate
clause is given (when accommodation is not required, because the
presupposition triggered by the initial adverbial clause is satisfied
through the preceding context). This was indeed what we observed.

We conclude from our findings that both hypotheses were partly
correct: Presupposition (satisfaction) plays a role, as suggested by
Gorrell et al. (1989), but only initial adverbial clauses trigger pre-
suppositions; and given-before-new does facilitate comprehension, as
suggested by Haviland and Clark (1974), but only when the preceding
context satisfies the presupposition triggered by the initial adverbial
clauses (e.g., “Tom eats a green pear. After he eats a green pear, he
drinks some water”).

The findings suggest that already at age four, children are not only
sensitive to the given-new distinction, and to syntactic structure (sub-
ordinate vs. main clauses), but also to the discourse-structuring func-
tion of initial adverbial clauses: Initial adverbial clauses are expected to
convey information that is part of the common ground. When this as-
sumption is violated, children have a harder time understanding ad-
verbial sentences correctly.

However, if we assume that the information structural effects in the
present study are due to accommodation processes, some other findings
– or the lack thereof – need more explanation. Neither eight-year-olds
nor adults showed any effects such as slower response times in those
conditions that would require accommodation. We suggest that in these
older age groups this is due to a ceiling effect, which can be explained
by the task and the nature of the presupposition trigger. In the present
study, the task for participants was to select the picture sequence that
showed the correct order of events, whereas other studies that found
accommodation effects measured reading times in self-paced reading or
asked participants to judge the veracity of a sentence (e.g., one
containing focus particles) when presented with an image. The task is
more difficult than that of the present study, and thus more likely to
lead to measurable accommodation effects.

We also observed that young children performed less well with main-
subordinate constructions, with the exception of before-sentences.
According to ‘accommodation’ accounts, final adverbial clauses do not
trigger presuppositions, so for these sentences, it should not matter
whether the given information is contained in the subordinate clause or
not. We suggest that our findings can be explained by taking iconicity
(i.e., the interaction between clause order and adverbial type) into
account.

3.6.2. Effects of clause order and adverbial type

We did not find that children found main-subordinate orders gen-
erally easier (hypothesis 1), as suggested by Diessel (2005). This con-
irms the results of De Ruiter et al.’s (2018) study with isolated ad-
verbial clauses. Rather, we found an interaction between clause order
and type and evidence for effects of iconicity. In the main-subordinate
order, before-sentences, both with the main clause given and with the
subordinate clause given, had higher accuracy scores than any of the
other sentence types. Recall that this is the iconic order for before-
sentences. For the other three subordinators (after, because, and if),
the iconic order is subordinate-main. As described in the introduction,
evidence is converging that iconicity plays an important role in chil-
dren’s adverbial sentence comprehension, with iconic sentences being
easier to comprehend than non-iconic ones. It is therefore not surprising
that the children made more errors with main-subordinate sentences,
given that the majority were non-iconic. The only sentence-type for
which main-subordinate is iconic is before-sentences, and children
performed very well with these.

Against our expectations, we did not find that children performed
better with before-sentences in general (hypothesis 4). As discussed in
the Introduction, several studies found an advantage for before, even in
non-iconic clause orders. In our study, the one before-condition that
stuck out in terms of children not reaching higher accuracy were sub-
ordinate-main sentences with the main clause given (see Fig. 3), such as
“Tom shouts out loudly. Before he drives away fast, he shouts out
loudly”. One possibility is that the combination of non-iconicity and
necessary accommodation dampened children’s accuracy in this con-
dition. Note that for the other three adverbials, accommodation is not
necessary in the non-iconic orders (main-subordinate).

Another finding from De Ruiter et al. (2018) was replicated: Re-
response times for because-sentences were longer than those for
after- and before-sentences. De Ruiter et al. suggested that this is due to
the fact that because and if require inferences regarding the causal
structure of the two events, which takes longer to process.

Notably, and against our predictions, we did not observe any of the
givenness or clause order effects found in the accuracy scores discussed
above reflected in the children’s response times. Response times were
affected only by adverbial type. This is generally in line with Blything
and Cain’s (2016) study on before and after, although they did find that
children responded faster to before-sentences than to after-sentences
(mirroring higher accuracy scores on before), a difference not present in
our study. Why there seems to be a disconnect between accuracy and
response times remains unclear and needs to be investigated in future
studies. One possibility is that the response time measure is not suffi-
ciently sensitive in this particular forced choice paradigm.

3.6.3. Effects of age

As predicted, adverbial sentence comprehension improved with age.
Five-year-olds were both more accurate and faster than four-year-olds
in processing these sentences, but both age groups differed markedly
from adults. Results from the older children suggest that there is sub-
stantial development throughout the early school years, but that even at
eight years of age, comprehension has not reached adult-levels. This is
in line with an earlier study with isolated because-sentences, which
found near-ceiling effects only around eleven years of age (Johnson & Chapman, 1980), although another study on the comprehension of *because* and *if* found even twelve-year-olds to make quite a few mistakes (Emerson & Gekoski, 1980). The development of adverbial sentence comprehension thus appears to be protracted, and full mastery may not be reached before puberty.

Importantly, the effects we found were the same for both four- and five-year-olds, suggesting that while overall accuracy improves with age, there are no qualitative differences in how information structure and iconicity affect complex sentence comprehension in younger children. The eight-year-olds’ data suggest that these effects diminish over time, as they were similar to the adult control group in that their responses were not affected by information structure or iconicity.

We suggest that these developmental changes can be explained by language experience. The first change concerns the decreasing effect of iconicity. One likely source of influence here is schooling. In the time between five and eight years, children learn how to read and write, and are increasingly exposed to written texts. An important feature of the “language of schooling” (Schleppegrell, 2004) is a higher rate of subordination and a focus on precision in expressing relationships (e.g., Snow & Uccelli, 2009; Uccelli et al., 2016). Text exposure has been found to predict eight- and twelve-year-olds’ use of passive relative clauses (Montag & MacDonald, 2015). It is likely that exposure to a range of adverbial sentences also impacts on children’s flexibility to construct mental representations from these sentences, reducing their reliance on heuristics like iconicity. Assuming a cue competition account, children may start out with using broad heuristics such as iconic mapping between language and events. As children are exposed to various kinds of adverbial sentences that express different semantic relationships and occur in different clause orders, this cue would gradually be outweighed by more reliable but less frequent cues, such as the meaning of the adverbial in a specific construction. The second developmental change concerns the decreasing sensitivity to non-canonical information structure. Increasing experience, in this case with conversational practices, could explain this. As children gain experience in conversing with others, they will more often encounter both non-canonical information structural contexts and infelicitous utterances that require repair (accommodation). A similar argument has been made by Aravind, Hackl, and Wexler (2018), who studied children’s comprehension of it-cleft constructions (which are associated with a presupposition). They suggest that learning repair mechanisms to deal with infelicitous utterances is a late development.

3.6.4. Individual differences

We did not find evidence that individual differences in language ability, memory, and inhibition had an influence on sentence comprehension beyond the effects of givenness, clause order, and iconicity. The results diverge from other studies on adverbial sentence comprehension that did observe an independent contribution of memory in particular (Blything et al., 2015; Blything & Cain, 2016), but echo those of De Ruiter et al.’s (2018) study on isolated adverbial sentences: While most measures were positively correlated with accuracy, and some negatively with response times, none of these measures remained significant predictors in the models after the experimental factors, including age, were accounted for. We can thus conclude that they do not make an independent contribution to overall performance. De Ruiter et al. concluded that “the ability to construct a coherent mental model from isolated complex sentences is not just a competence emerging from a combination of general language ability, memory, and executive function, but a distinct construct” (p. 215). Our results extend this to complex sentences in (minimal) context. We do not want to conclude from this that individual differences are irrelevant to complex sentence processing, as numerous studies have found them to play a role (for an overview, see e.g., Kidd et al., 2018). It is possible that they are less pronounced in the face of relatively strong experimental manipulations, or that they only surface under particular conditions. We also note that our sample came largely from a high SES (socio-economic status) background and was relatively homogenous (small standard deviations).

3.6.5. Presenting sentences in context

Finally, we discuss the effect of context on children’s complex sentence comprehension in general. Our exploratory comparison of the original De Ruiter et al. (2018) data and our data showed an overall beneficial effect of context. Both four- and five-year-olds performed better in the present study, even when controlling for memory. This suggests that even slightly odd contexts (such as those in which the presupposition is not satisfied, see above) help children create a mental representation of the events compared to when they are not given any context. Interestingly, this may explain in part why the age at which the children performed above chance (five years) in De Ruiter et al.’s study was older than that found by two other studies by Blything and colleagues (Blything et al., 2015; Blything & Cain, 2016). Blything and colleagues had tested children on *before* and *after* only, and found above-chance performance already in four-year-olds. Both studies used a forced-choice paradigm, but the set-up and the actual task were different. In De Ruiter et al. (2018), as in the present study, children were instructed to “touch the matching story”. They first heard the sentence (e.g., “Before he eats a green pear, he drinks some water”) while looking at a blank screen, and then saw two picture stories that showed both actions in the two orders (eating first/drinking second; drinking first/eating second), while the sentence was played again. In the Blything studies, children watched short animated clips of both actions (e.g., eating a hotdog, putting shoes on) successively next to each other, which ended in a freeze frame. They then heard the prompt “Listen carefully and touch the thing Tom/Sue did first” (Blything et al., 2015) or “Listen carefully and touch the thing Tom/Sue did last” (Blything & Cain, 2016), followed by the critical sentence (e.g., “Before he ate the burger, he put on the sandals”). De Ruiter et al. (2018) argue that the fact that the children were aware that they had to pay attention only to what happened first/last, and that they knew what the two possible actions were before hearing the sentence made the task easier. In contrast, the task in De Ruiter et al.’s study is more challenging, in that it requires listeners to construct a mental representation of the sequence of events from language only, without any initial (visual) support. In other words, children in De Ruiter et al.’s study were not provided with any context (apart from who the protagonist was), while in Blything et al.’s (2015) and Blything and Cain’s (2016) study the events were already known, or at least visually given. This seems to indicate that both providing prior information about the actions visually and providing information about (one of) the actions aurally (through a context sentence) facilitate the construction of the mental representation described by the sentence.

3.6.6. Limitations of the study

Experimental research necessarily involves trade-offs between control on the one hand, and naturalness (ecological validity) on the other. As this is the first study to examine the effects of information structure on the comprehension of four different adverbials, we opted for a higher level of control in our stimuli, in order not to introduce too many other potential sources of variation. We decided to operationalize givenness simply through repeating the verb phrase (e.g., “Tom sweeps the new floor. After he sweeps the new floor...”). This is clearly somewhat marked, but not entirely unnatural (imagine a teacher saying “Go wash your hands. After you’ve washed your hands, line up at the door for the lunch break”). Future investigations could manipulate givenness in different ways, for example by using demonstratives that refer to complex antecedents (“Tom sweeps the new floor. After he does that, ...”). We also did not vary the prosody in the adverbial sentences, depending on which clause was given. The adverbial sentences were identical with the ones used by De Ruiter et al. (2018), spliced together with the context sentences, which had been recorded in isolation. This
allowed for a direct comparison of the two studies. Future investigations could look into how prosody (e.g., relative attenuation of the given clause, focussing of the new clause may improve comprehension of otherwise “odd” constructions, e.g., “Tom eats a green pear. He drinks some water after he eats a green pear”).

4. Conclusions

In this study we have shown that children as young as four and five years of age are sensitive to interactions between form (clause order) and function (information status) and expect sentence-initial adverbial clauses to contain given information. We suggest that sentences with initial adverbial clauses that contain new information require pragmatic accommodation, which leads to higher error rates. Moreover, even a small amount of contextual information to establish one event as given improves children’s ability to interpret complex sentences. We have also found more evidence for the importance of iconicity in children’s comprehension of these adverbial clauses. Purely structure-based explanations (e.g. of main-subordinate order being preferred) cannot account for our results, but nor can explanations positing broad information-structural heuristics such as given-before-new. Instead, we have shown that they interact both with each other and with adverbial semantics. Finally, we have suggested the dependence on iconicity and information structure may reduce as children both move towards adult levels of literacy and gain increasing experience with non-canonical information structural contexts.

Acknowledgements

This work was supported by the International Centre for Language and Communicative Development (LuCiD). The support of the Economic and Social Research Council [ES/L008955/1] is gratefully acknowledged. We would like to thank the editor, Sarah Brown-Schmidt, and two anonymous reviewers for their helpful comments and recommendations on an earlier version of this manuscript. Further, we would like to thank all nurseries and schools for their participation, and the Manchester Museum for allowing us to use their space for this research project. We are also grateful to Eilidh Shaw and Meerab Mazhar, who helped with the data collection. Thanks also to Yu Li for technical support and to Felix Engelmann for helping with using the rstanarm package.

Appendix 1

The four- and five-year-olds completed two sessions on two different days. Below is an overview of the order in which they completed the tasks.

Table A1 provides an example of all experimental items as used in one session (of one experimental list).

Table A1

| Context sentences | Adverbial sentence |
|-------------------|--------------------|
| Sue hoovers the house. | She hoovers the house, after she paints the old fence. |
| Tom watches TV. | After he watches TV. |
| Tom drinks some water. | He drinks some water, after he eats a green pear. |
| Tom coughs a few times. | After he coughs a few times, he laughs really hard. |
| Sue runs over here. | She hides over there, after she runs over here. |
| Sue dances around | After she dances around, she bounces away. |
| Tom plays his big drum. | He plays his big drum, before he reads his new book. |
| Sue takes a hot bath. | Before she draws a picture, she takes a hot bath. |
| Sue breaks her small train. | She breaks her small train, before she builds a tower. |
| Sue crawls on the floor. | Before she crawls on the floor, she hops up and down. |
| Tom shouts out loudly. | He drives away fast, before he shouts out loudly. |
| Tom waves happily. | Before he waves happily, he swims on his back. |
| Sue closes her eyes. | She closes her eyes, because she bangs her head hard. |
| Tom sees the snowman. | Because he opens the door, he sees the snowman. |
| Tom misses the bus. | He misses the bus, because he rides his old bike. |
| Tom trips suddenly. | Because he trips suddenly, he cries really hard. |
| Sue dives in the pool. | She feels really warm, because she dives in the pool. |
| Sue looks at the sky. | Because she looks at the sky, she slips to the ground. |
| Tom wins a nice cup. | He wins a nice cup, if he sings a happy song. |
| Sue finds her other shoe. | If she cuts the long grass, she finds her other shoe. |
| Sue hears the doorbell. | She hears the doorbell, if she presses the button. |
| Sue talks to herself. | If she talks to herself, she wakes up in the night. |
| Tom sits down in his chair. | He gets very bored, if he sits down in his chair. |
| Tom sneezes lots of times. | If he sneezes lots of times, he falls in the field. |

Appendix 2

In the main analysis (Section 3) we used the Bayesian mixed models using the rstanarm package, which are more appropriate than the originally planned Bayesian regression using the BayesFactor package. Bayesian mixed models use logistic regression, like the traditional frequentist (g)lmer models, but they have not been readily available in R at the time of the pre-registration. For completeness, we here report the results for the Bayesian regression (using the BayesFactor package), as originally planned in the pre-registration (https://osf.io/7pw5j/?view_only=55be7a0667c540938dc7cd768762e03).

For this, all possible models (main factors and interactions) were entered at the same time to identify the model under which the data are most likely. In line with recommendations by Morey and Rouder (2011) we used a Cauchy prior with scale parameter 1/√2 for the standardised effect size. Bayes Factors provide information about the relative strength in evidence in favour of either the alternative or the null hypothesis, which is not possible in the traditional null-hypothesis significance-testing (NHST) paradigm (see De Ruiter et al., 2018: 209). For the verbal interpretation of Bayes factors we used the evidence categories adapted from Jeffreys (1961), cited in Wetzels et al. (2011).
Accuracy

The Bayesian regression (Bayes factor analysis) included the main factors AgeGroup, Type, ClauseOrder, and the interaction of ClauseOrder and ClauseGiven. Unlike the frequentist analysis, it did not contain the main factor ClauseGiven, and the three-way-interaction of Type, ClauseOrder, and ClauseGiven. This model had a Bayes factor of > 2 million (‘extreme evidence’). The data were > 1000 × more likely under this model than under the model that did contain the main factor ClauseGiven. Note that the main factor of ClauseGiven was retained in the frequentist analysis because it was part of a significant interaction. Furthermore, the Bayes factor analysis suggested that the data were 42 times more likely under a model that did not contain the three-way-interaction of Type, ClauseOrder, and ClauseGiven than they were under a model that did. The results of the linear mixed effect model and the Bayes factor model diverged with respect to whether or not there was a three-way interaction between Type, ClauseOrder, and ClauseGiven. As mentioned above, the likely reason for the divergence is that the current version of the BayesFactor package works with linear regression, and not logistic regression, like the glmer models.

For the adult group, the Bayes factor analysis confirmed the frequentist analysis, with extreme evidence for the intercept-only model (BF > 2 million), under which the data were > 3.5 times more likely than the model that included ClauseOrder.

Response times

The Bayes factor analysis aligned with the traditional analysis: The model under which the data were most likely (BF > 1 million, ‘extreme evidence’) contained only the main factors AgeGroup and Type.

In the adult control group, there was no evidence for that any of the factors (Type, ClauseOrder, ClauseGiven) played a role. The Bayes factor analysis showed that there was extreme evidence for the intercept-only model. The data were for example about six times more likely under the intercept-only model than under a model that included ClauseGiven, or about 400 times more likely under the intercept-only model than under the model that included Type.

Individual differences

As in the Bayesian mixed effect models in the main analysis (Section 3.4.3), we added those factors that were significantly correlated with mean accuracy (the CELF Linguistic Concept Score, the BPVS score, the ERB Sentence Imitation Score, the Flanker score, and the Digit Span score) and mean response times (the CELF Linguistic Concepts score and the BPVS score), respectively, to the final models. For accuracy, in the frequentist analysis, only the BPVS score had remained a significant predictor. In contrast, the Bayes factor analysis suggested that there was no discernible difference between the model with and the model without the BPVS score. The Bayes factor for the model without the BPVS score was 1.3, which would be interpreted as anecdotal evidence for the data to be more likely under this (simpler) model than under the other.

Of the two predictors added to the response time model, neither had remained significant in the frequentist analysis. In line with this, the Bayesian analysis showed that the data were 3.5 times more likely under the model without the BPVS score.

Appendix 3

In this section we provide information about the model estimates for the random effects.

Tables A2 and A3 show the estimates for the random effects of the final linear mixed effects model for accuracy in the four- and five-year-olds, for reference level if and because, respectively.

### Table A2
Model estimates of the variances, standard deviations, and correlations between the random-effects terms of the final linear mixed effect model for accuracy in the four- and five-year-olds. Values were calculated using the VarCorr function in the lme4 package. The reference level for Type is because. Number of observations: 3816; Subjects: 80; Items: 24.

| Groups  | Name                  | Std.Dev. | Corr       |
|---------|-----------------------|----------|------------|
| Subject | (Intercept)           | 0.52873  |            |
|         | Type.sumafter         | 0.31659  | 0.726      |
|         | Type.sumbefore        | 0.47902  | 0.571      |
|         | Type.sumbecause       | 0.42019  | −0.697     | −0.943 | −0.899 |
|         | ClauseOrderSub-main   | 0.58874  | −0.377     | −0.179 | 0.369 | 0.666 |
| ItemNo  | (Intercept)           | 0.26965  |            |

### Table A3
Model estimates of the variances, standard deviations, and correlations between the random-effects terms of the final linear mixed effect model for accuracy in the four- and five-year-olds. Values were calculated using the VarCorr function in the lme4 package. The reference level for Type is if. Number of observations: 3816; Subjects: 80; Items: 24.

| Groups  | Name                  | Std.Dev. | Corr       |
|---------|-----------------------|----------|------------|
| Subject | (Intercept)           | 0.52873  |            |
|         | Type.sumif            | 0.32324  | −0.651     |
|         | Type.sumafter         | 0.31658  | 0.726      | −0.607 |
|         | Type.sumbefore        | 0.47901  | 0.571      | −0.995 | 0.576 |
|         | ClauseOrderSub-main   | 0.58873  | −0.377     | −0.294 | −0.170 | 0.369 |
| ItemNo  | (Intercept)           | 0.26964  |            |
Tables A4 and A5 show the estimates for the random effects of the final linear mixed effects model for response times in the four- and five-year-olds, for reference level if and because, respectively.

Table A4
Model estimates of the variances, standard deviations, and correlations between the random-effects terms of the final linear mixed effect model for response times in the four- and five-year-olds. Values were calculated using the VarCorr function in the lme4 package. The reference level for Type is because. Number of observations: 2362; Subjects: 80; Items: 24.

| Groups       | Name                      | Std.Dev. | Corr  |
|--------------|---------------------------|----------|-------|
| Subject      | (Intercept)               | 0.52873  |       |
|              | Type.sumif                | 0.3224   | −0.651|
|              | Type.sumafter             | 0.31658  | 0.726 |
|              | Type.sumbefore            | 0.47901  | −0.995|
|              | ClauseOrderSub-main       | 0.58873  | −0.377|
|              | (Intercept)               | 0.26964  |       |

Table A5
Model estimates of the variances, standard deviations, and correlations between the random-effects terms of the final linear mixed effect model for response times in the four- and five-year-olds. Values were calculated using the VarCorr function in the lme4 package. The reference level for Type is if. Number of observations: 2362; Subjects: 80; Items: 24.

| Groups       | Name                      | Std.Dev. | Corr  |
|--------------|---------------------------|----------|-------|
| Subject      | (Intercept)               | 1840.11  | −0.466|
|              | Type.sumafter             | 541.63   |       |
|              | Type.sumbefore            | 330.89   | −0.771|
|              | Type.sumbecause           | 423.42   | 0.652 |
|              | ClauseOrderSub-main       | 443.96   |       |
|              | (Intercept)               | 3010.14  |       |

Appendix 4
Here we report the full set of results for the analysis with the eight-year-olds. Table A6 shows the results of the Bayesian mixed effects model for accuracy for the eight-year-old group.

Table A6
Results from the Bayesian linear mixed-effect model for accuracy for the eight-year-olds, showing estimates of the mean, and the lower and upper bounds of the credible interval. Credible intervals that do not contain zero are highlighted in bold. Note that because sum contrasts only allow n-1 contrasts to be determined at one time (i.e., three comparisons for the four-level variable Type), for readability we combined the output of two models in order to show all contrasts in one table. Number of observations: 576; Subjects: 12; Items: 24.

|                         | Mean (estimate) | 2.5% CI lower bound | 97.5% CI upper bound |
|-------------------------|-----------------|----------------------|----------------------|
| Intercept               | 2.0969          | 1.1382               | 3.163                |
| Type.sumafter           | 0.9091          | −0.3995              | 2.343                |
| Type.sumbefore          | −0.073          | −1.2214              | 1.1569               |
| Type.sumbecause         | −0.0986         | −1.1966              | 1.0689               |
| Type.sumif              | −0.6748         | −1.7814              | 0.414                |
| ClauseOrderSub-main     | 0.728           | −0.1623              | 1.7093               |
| ClauseGiventGivenSub    | 0.068           | −0.757               | 0.9198               |
| ClauseOrderSub-main:ClauseGiventGivenSub | 0.2972 | −0.8747 | 1.462 |
| Type.sumafter:ClauseOrderMain-sub:ClauseGiventGivenMain | −0.7056 | −2.1976 | 0.7472 |
| Type.sumbefore:ClauseOrderMain-sub:ClauseGiventGivenMain | 1.1646 | −0.3011 | 2.6886 |
| Type.sumbecause:ClauseOrderMain-sub:ClauseGiventGivenMain | −0.5735 | −1.8775 | 0.6685 |
| Type.sumif:ClauseOrderMain-sub:ClauseGiventGivenMain | 0.0404 | −1.2488 | 1.3168 |
| Type.sumafter:ClauseOrderSub-main:ClauseGiventGivenMain | −0.1861 | −1.7941 | 1.4963 |
| Type.sumbefore:ClauseOrderSub-main:ClauseGiventGivenMain | −0.3722 | −1.8425 | 1.0304 |
| Type.sumbecause:ClauseOrderSub-main:ClauseGiventGivenMain | 0.3188 | −1.0817 | 1.7512 |
| Type.sumif:ClauseOrderSub-main:ClauseGiventGivenMain | 0.1188 | −1.2176 | 1.4718 |
| Type.sumafter:ClauseOrderMain-sub:ClauseGiventGivenSub | −0.5505 | −2.0782 | 0.8821 |
| Type.sumbefore:ClauseOrderMain-sub:ClauseGiventGivenSub | 0.2271 | −1.1312 | 1.5628 |
| Type.sumbecause:ClauseOrderMain-sub:ClauseGiventGivenSub | 0.3202 | −0.9823 | 1.645  |
| Type.sumif:ClauseOrderMain-sub:ClauseGiventGivenSub | −0.1144 | −1.4028 | 1.1362 |

Table A7 shows the results for the Bayesian mixed effects model for response times for the eight-year-olds.
### Table A7

Results from the Bayesian linear mixed-effect model for response times for the eight-year-olds, showing estimates of the mean, and the lower and upper bounds of the credible interval. Credible intervals that do not contain zero are highlighted in bold. Note that because sum contrasts only allow n-1 contrasts to be determined at one time (i.e., three comparisons for the four-level variable Type), for readability we combined the output of two models in order to show all contrasts in one table.

Number of observations: 576; Subjects: 12; Items: 24.

| Type.sumbecause | Intercept 2107.214 | 1078.201 | 3196.227 |
|-----------------|---------------------|-----------|-----------|
| Type.sumbefore  | −107.2104           | −874.341  | 620.5036  |
| Type.sumif      | 53.3714             | 832.9076  | 753.3426  |
| Type.sumbefore:ClauseOrderMain-sub:ClauseGivenGivenMain | 447.6711 | 230.4668 | 1236.0254 |
| Type.sumbecause:ClauseOrderMain-sub:ClauseGivenGivenMain | −420.1058 | −1138.686 | 319.5042 |
| Type.sumbecause:ClauseOrderMain-sub:ClauseGivenGivenMain | 209.2175 | −493.3061 | 1041.7582 |
| Type.sumbefore:ClauseOrderMain-sub:ClauseGivenGivenMain | 273.9238 | −493.3061 | 1041.7582 |

### References

Amidon, A. (1976). Children’s understanding of sentences with contingent relations: Why five-year-olds cannot understand before and after. *Advances in Psycholinguistics*, *2*, 417–423. https://doi.org/10.1015/2005.43.3.449.

Amidon, A. (1972). Why five-year-olds cannot understand before and after. *Journal of Experimental Child Psychology*, *22*(3), 423–437. https://doi.org/10.1016/0022-0965(76)90106-5.

Amidon, A., & Carey, P. (1972). Why five-year-olds cannot understand before and after. *Journal of Experimental Child Psychology*, *22*(3), 423–437. https://doi.org/10.1016/0022-0965(76)90106-5.

Amidon, A., & Carey, P. (1972). Why five-year-olds cannot understand before and after. *Journal of Experimental Child Psychology*, *22*(3), 423–437. https://doi.org/10.1016/0022-0965(76)90106-5.

Amidon, A., & Carey, P. (1972). Why five-year-olds cannot understand before and after. *Journal of Experimental Child Psychology*, *22*(3), 423–437. https://doi.org/10.1016/0022-0965(76)90106-5.

Amidon, A., & Carey, P. (1972). Why five-year-olds cannot understand before and after. *Journal of Experimental Child Psychology*, *22*(3), 423–437. https://doi.org/10.1016/0022-0965(76)90106-5.

Amidon, A., & Carey, P. (1972). Why five-year-olds cannot understand before and after. *Journal of Experimental Child Psychology*, *22*(3), 423–437. https://doi.org/10.1016/0022-0965(76)90106-5.

Amidon, A., & Carey, P. (1972). Why five-year-olds cannot understand before and after. *Journal of Experimental Child Psychology*, *22*(3), 423–437. https://doi.org/10.1016/0022-0965(76)90106-5.

Amidon, A., & Carey, P. (1972). Why five-year-olds cannot understand before and after. *Journal of Experimental Child Psychology*, *22*(3), 423–437. https://doi.org/10.1016/0022-0965(76)90106-5.

Amidon, A., & Carey, P. (1972). Why five-year-olds cannot understand before and after. *Journal of Experimental Child Psychology*, *22*(3), 423–437. https://doi.org/10.1016/0022-0965(76)90106-5.

Amidon, A., & Carey, P. (1972). Why five-year-olds cannot understand before and after. *Journal of Experimental Child Psychology*, *22*(3), 423–437. https://doi.org/10.1016/0022-0965(76)90106-5.
