LJMU Research Online

MacDonald, R, Brandt, S, Theakston, A, Lieven, E and Serratrice, L

 The role of animacy in children's interpretation of relative clauses in English: Evidence from sentence-picture matching and eye movements

http://researchonline.ljmu.ac.uk/id/eprint/13117/

Article

The Role of Animacy in Children's Interpretation of Relative Clauses in English: Evidence From Sentence–Picture Matching and Eye Movements

Ross Macdonald,[a] Silke Brandt,[b] Anna Theakston,[c] Elena Lieven,[c] Ludovica Serratrice[d]

[a]*School of Psychology, Liverpool John Moores University*
[b]*Department of Linguistics and English Language, Lancaster University*
[c]*Division of Human Communication, Development & Hearing, University of Manchester*
[d]*Centre for Literacy and Multilingualism, University of Reading*

Abstract

Subject relative clauses (SRCs) are typically processed more easily than object relative clauses (ORCs), but this difference is diminished by an inanimate head-noun in semantically non-reversible ORCs ("The book that the boy is reading"). In two eye-tracking experiments, we investigated the influence of animacy on online processing of semantically reversible SRCs and ORCs using lexically inanimate items that were perceptually animate due to motion (e.g., "Where is the tractor that the cow is chasing"). In Experiment 1, 48 children (aged 4;5–6;4) and 32 adults listened to sentences that varied in the lexical animacy of the NP1 head-noun (Animate/Inanimate) and relative clause (RC) type (SRC/ORC) with an animate NP2 while viewing two images depicting opposite actions. As expected, inanimate head-nouns facilitated the correct interpretation of ORCs in children; however, online data revealed children were more likely to anticipate an SRC as the RC unfolded when an inanimate head-noun was used, suggesting processing was sensitive to perceptual animacy. In Experiment 2, we repeated our design with inanimate (rather than animate) NP2s (e.g., "where is the tractor that the car is following") to investigate whether our online findings were due to increased visual surprisal at an inanimate as agent, or to similarity-based interference. We again found greater anticipation for an SRC in the inanimate condition, supporting our surprisal hypothesis. Across the experiments, offline measures show that lexical animacy influenced children's interpretation of ORCs, whereas online measures reveal that as RCs unfolded, children were sensitive to the perceptual animacy of lexically inanimate NPs, which was

Correspondence should be sent to Ross Macdonald, School of Psychology, Liverpool John Moores University, Tom Reilly Building, Byrom Street, Liverpool L3 3AF, UK. E-mail: r.g.macdonald1@ljmu.ac.uk

not reflected in the offline data. Overall measures of syntactic comprehension, inhibitory control, and verbal short-term memory and working memory were not predictive of children's accuracy in RC interpretation, with the exception of a positive correlation with a standardized measure of syntactic comprehension in Experiment 1.

1. Introduction

A key issue in the study of the development of sentence comprehension concerns the linguistic factors that contribute to children's interpretation of syntactic constructions, whether young learners pay attention to the same cues as adults, and whether the temporal resolution of their parsing preferences is the same as adults'. An extensive program of research has focused on children's comprehension of relative clauses (RCs), particularly on the well-attested asymmetry between subject and object relative clauses (SRCs and ORCs), and on the structural, semantic, and pragmatic factors that mitigate processing difficulties with ORCs.

In the present two studies, we used eye-tracking to investigate English-speaking 4- to 6-year-olds' online processing of SRCs and ORCs. We focused on the animacy of the head noun, and we included offline accuracy measures alongside eye-movement measures, and response times (RTs). To investigate the role of individual differences in RC interpretation, we added measures of verbal short-term memory (VSM) and verbal working memory (VWM), of inhibitory control, and of receptive syntactic skills. We also compared children's online and offline performance to a group of English-speaking adults to assess the extent to which interpretation, parsing strategies, and their time course show developmental continuity.

1.1. The nature of the SRC-ORC asymmetry

It is well attested that, all else being equal, ORCs like (2) are typically understood less accurately than SRCs like (1):

1. Where is the boy [$_{SRC}$ that pushed the girl]?
2. Where is the boy [$_{ORC}$ that the girl pushed]?

Cross-linguistically, children and adults typically find SRCs easier to process than ORCs (e.g., Adani, van der Lely, Forgiarini, & Guasti, 2010; Booth, MacWhinney, & Harasaki, 2000; Brandt, Kidd, Lieven, & Tomasello, 2009; Friedmann, Belletti, & Rizzi, 2009; Mak, Vonk, & Schriefers, 2002, 2006; Traxler, Morris, & Sealy, 2002), but the asymmetry is obliterated in languages like Chinese where the RC precedes the head noun (e.g., Hsiao & Gibson, 2003). Multiple accounts have been proposed to explain the extra demands imposed by ORCs, including their non-canonical word order, the complexity of

their syntactic derivation, and the additional memory demands involved in thematic role assignment (Gordon & Lowder, 2012, for a review).

In a language like English, the canonical and most frequent word order is SVO where a subject noun phrase (NP) is followed by a verb phrase (VP) and by an object NP. In an SRC like (1), this linear word order is preserved with NP1 (*the boy*) preceding the VP (*pushed*) and NP2 (*the girl*). An ORC like (2) displays a non-canonical word order where two NPs (*the boy*, *the girl*) precede the VP (*pushed*), and the object NP1 precedes the subject NP2. The word order in an ORC deviates from the highly frequent, and thus more likely expected SVO word order, and this violation may make the parsing and the interpretation of ORCs more demanding (cf. Wells, Christiansen, Race, Acheson, & MacDonald, 2009).

The asymmetrical parsing demands between SRCs and ORCs are also explained by syntax only/syntax-first accounts of language processing where comprehenders adopt an active filler strategy when processing RCs (Frazier, 1987). If there is a preference to initially assign the syntactic role of subject to the RC head-noun (NP1) upon encountering the complementizer *that*, the parser proceeds to analyze the incoming embedded sentence as an SRC. When this parse fails in the case of an ORC, the comprehender will need to resort to a time-consuming process of syntactic reanalysis.

Another suggested reason for the SRC-ORC asymmetry is connected to the structural complexity of ORCs compared to SRCs. According to locality-based theories of sentence comprehension (Frazier & Flores d'Arcais, 1989; Gibson, 1998, 2000; Grodner & Gibson, 2005; Warren & Gibson, 2002), the farther away the two dependent elements, the more demanding the dependency. In the ORC, the head NP must be maintained in memory for longer before it can be integrated with the post-verbal gap in the RC. Consequently, the ORC should incur a greater storage cost in working memory than the SRC to be processed correctly, as the object needs to be retrieved and integrated. In (1) and (2) repeated below as (3) and (4), both SRCs and ORCs contain a phonologically empty gap (coded as $\otimes$ in the examples) that is co-indexed with the extracted element, that is, the head NP of the RC (*the boy* in the examples below):

3. Where is the boy [$_{SRC}$ $\otimes$ that pushed the girl]?
4. Where is the boy [$_{ORC}$ that the girl pushed $\otimes$]?

Meaningful parsing of the sentence requires the anaphoric resolution of the dependency between the gap and the filler antecedent (i.e., the head NP1). In the SRC, the gap appears as the subject of the verb in the RC, while in the ORC it appears as the object of the verb in the RC.

Aside from the sheer distance between the head noun and the gap, in ORCs with two (animate) NPs, some have argued that it is the intervening NP2 in the RC that makes it harder to establish the dependency between the two (Friedmann et al., 2009; Grillo, 2009). Specifically, it is the structural and semantic similarity in an ORC like (4) between the embedded NP2 and the head NP1 (both NPs, both animate) that is the cause of the difficulty. When the structural/semantic similarity between NP1 and NP2 is lessened, comprehension accuracy increases. For instance, for both English-speaking and Italian-

speaking children, Adani et al. (2010, Adani, Forgiarini, Guasti, & Van Der Lely, 2014) found that ORCs where there was a number mismatch between NP1 and NP2 were comprehended more accurately than those where number was the same for both NPs, thus showing that grammatical features like Number, in addition to the number of full NPs, play a significant role in similarity. In contrast, a gender mismatch in Italian ORCs between NP1 and NP2 led to a significantly smaller facilitatory effect (Adani et al., 2010), suggesting that not all grammatical features are weighted equally. In terms of referential expression similarity, Haendler, Kliegl, and Adani (2015) reported that ORCs with a NP1 realized by a proper noun were processed more accurately when the NP2 was realized by a first-person pronoun rather than a third-person pronoun.

1.2. The role of animacy

Corpus analyses have shown that ORCs with two animate NPs of the type that are commonly used in experimental studies are not actually that frequent outside of the lab as ORCs tend to have an inanimate NP1 and an NP2 realized by a subject pronoun (e.g., *The book that I read*) (Arnon, 2010; Kidd, Brandt, Lieven, & Tomasello, 2007). More specifically, there is cross-linguistic evidence that in RC sentences where NP2 is realized by a personal pronoun, ORCs are significantly more frequent than SRCs (see Reali, 2014, for corpus and experimental evidence in Spanish, and Reali & Christiansen, 2007, for corpus and experimental evidence in English). The low frequency of ORCs with two animate full NPs, and hence the more limited experience with this type of construction, has been proposed as an additional reason why children—and even adults—find this type of ORC harder to process. The animacy of the NP1 has been repeatedly shown to be another significant predictor of children's accuracy in processing ORCs in offline studies (Arnon, 2010; Bentea, Durrleman, & Rizzi, 2016; Brandt et al., 2009). ORCs with an inanimate NP1 like (6) are typically interpreted more accurately by children than ORCs with an animate NP1 like (5).

5. The girl that the boy kicked
6. The ball that the boy kicked

Data showing an interaction between clause type (SRC vs. ORC) and animacy in adult studies (Baudiffier, Caplan, Gaonach, & Chesnet, 2011; Betancort, Carreira, & Sturt, 2009; Gennari & MacDonald, 2008; Mak, Vonk, & Schriefers, 2006; Traxler et al., 2002) also confirm that semantic information influences the syntactic choices listeners/readers make when confronted with an RC, and that the advantage for SRCs disappears when the subject NP in the ORC is animate and the object NP is inanimate (as in 6 above).

The extent to which animacy affects the interpretation of RCs, however, warrants further investigation in connection with two unresolved issues: the relationship between the animacy of NP1 and the semantic reversibility of the verb in the RC, and the relative contribution of lexical and perceptual animacy to syntactic role attribution. In the following, we will unpack the notion of semantic reversibility, and we will introduce the distinction between lexical and perceptual animacy that is relevant for our studies.

With the exception of Bentea et al.'s (2016) picture selection task, previous studies investigating children's performance on ORCs with animate and inanimate NP1 present a confound between the animacy of NP1 and the semantic reversibility of the RC. ORCs like (5) with an animate NP1 (*the girl*) are semantically reversible as either of the two animate referents could be the agent or the patient of the verb *kick*. In contrast, in an ORC with an inanimate NP1 (*the ball*) like in (6), the sentence is semantically non-reversible and the only plausible interpretation is one in which *the boy* is the agent (see O'Grady, 2011, for a similar point about the confound between animacy and semantic reversibility). It is therefore not clear whether the facilitation arises because of the semantic inanimacy of the NP1 or because of the semantic non-reversibility of the verb *read*. In their seminal experiment on children's ORC interpretation, Kidd et al. (2007) did not fully cross the reversibility of their two-NP ORCs with animate and inanimate NP1. While all of the four ORC items with an inanimate NP1 were semantically non-reversible, only two of the four ORCs with an animate head were semantically implausible when reversing the two NPs (i.e., *the man that the dog bit at the park yesterday; There is the girl that the cat licked in the garden today*). This design therefore creates some ambiguity in the interpretation of the animate ORCs, but not of the inanimate ORCs. Without fully crossing the semantic reversibility of the verb with the animacy of the NP1, it is impossible to tease apart the role of the animacy of the head noun from the semantic reversibility of the RC in the correct interpretation of ORCs. The presence of the relativizer *that* following an NP is a syntactic cue to the beginning of an RC, in a sentence like (6) and therefore correct interpretation of the sentence as an ORC could be rescued by the semantics of the verb. Even if the RC were initially incorrectly interpreted as an SRC—the incongruency between the inanimacy of the NP *the ball* and the verb *kick* requiring an animate agent—would trigger re-analysis of the RC.

A second issue that has not been directly addressed by the animacy manipulation in previous developmental research is the extent to which lexical versus perceptual animacy is responsible for the observed animate–inanimate distinction in ORC interpretation. In this literature, animacy has typically been treated as a binary semantic feature (+/− animate) of a lexical item, for example, boy = animate, car = inanimate. While it is of course possible to think of animacy in these terms (lexical animacy), there is considerable evidence that the cognitive representation of an entity can vary along a continuum (Silverstein, 1976), and that in different contexts people can conceptualize lexically inanimate entities as more or less animate (perceptual animacy, e.g., Boudewyn, Blalock, Long, & Swaab, 2019; Nelson & Vihman, 2018; Nieuwland & Van Berkum, 2006). One contextual cue that can increase the perceptual animacy of lexically inanimate entities is motion (Scholl & Tremoulet, 2000; Vogels, Krahmer, & Maes, 2013). Motion and causing change of state are also two of the five proto-agent properties identified by Dowty (1991), and agency is intimately connected with animacy whereby agent subjects tend to be more frequently animate than inanimate—at least in English (Clark, 1965). Given this premise, it remains to be seen whether the animacy effects that have been reported in the RC interpretation literature for lexically inanimate entities (e.g., *pen*, *food*, *fence*, *ball* in Kidd et al., 2007) can be replicated for lexically inanimate entities when they can be

contextually perceived as higher on the perceptual animacy scale because they are implicated in a motion event, for example, a car chasing a cow.

1.3. Online and offline measures of RC interpretation

In the present two studies, we chose the visual world method (Tanenhaus, Spivey-Knowlton, Eberhard, & Sedivy, 1995) to study the online interpretation of RCs, and we took advantage of this paradigm to disentangle the effects of lexical animacy from the semantic reversibility of the event. Previous studies investigating the time course of RC interpretation as a function of animacy of the NP1 in adult populations have included written materials, an option that is not available when studying younger children with limited literacy skills. The rationale for choosing the visual world paradigm is based on the assumption that eye movements can be used to study the dynamic process of RC interpretation as it happens, while offline accuracy data can give us an insight into the final outcome of the process. The suitability of the visual world paradigm to study language processing and development in children rests on three crucial linking assumptions (Trueswell, 2008): (a) Eye position is a metric of spatial attention driven by properties of the visual stimulus and by the child's goals (i.e., understanding who does what to whom in a visual scene); (b) in tasks that require the mapping of linguistic expressions to visual referents, eye movements are a proxy for referential decisions (i.e., children map referential expressions in the spoken signal to referent in the visual world); and (c) eye movements to referents in the visual world can be used to infer syntactic parsing decisions (i.e., eye movements to a referent who is the patient of a transitive action—e.g., *the cat*—indicate correct interpretation of a passive sentence—e.g., *Look at the cat chased by the dog*).

In our study, we added forced-choice comprehension questions for each of the items in the two eye-tracking experiments to investigate the alignment between online and the offline comprehension measures. Our position is that grammar and the language processor are essentially the same system; therefore, we do not make a distinction between syntactic representations (i.e., grammar), and real-time comprehension and production processes (i.e., the parser). Both are part of the same cognitive system that builds representations that are used for speaking and for understanding (Lewis & Phillips, 2015). We are, however, open to the possibility that there may be a misalignment between offline and online interpretations as has been documented before, for example in the case of garden path sentences (Wonnacott, Joseph, Adelman, & Nation, 2016). The parser always starts by pursuing the most likely and grammatical syntactic analysis (e.g., an SRC upon encountering a NP followed by the relativizer *that*), so there is no sense in which parser and grammar are at odds. However, a misalignment arises when the parser's initial choice proves to be inconsistent with the rest of the sentence and syntactic re-analysis becomes necessary (e.g., the constituent following the relativizer *that* is not a verb but another NP, a cue that the sentence is an ORC and not an SRC). Moreover, even in cases of successful online syntactic re-analysis, the incorrect interpretation associated with the initial incorrect parse may persist (Christianson, Hollingworth, Halliwell, & Ferreira, 2001; van

Gompel, Pickering, Pearson, & Jacob, 2006). Most research on syntactic re-analysis has focused on referential ambiguity, that is, cases in which an NP can initially be misanalyzed as the direct object of the verb in the first clause when in fact it is the subject of the second clause (e.g., *While Anna dressed the baby spat up on the bed; When the man hunted the deer ran into the woods*). Of particular relevance for our studies of RC interpretation, in a series of experiments on the interpretation of head final Japanese RCs, Nakamura and Arai (2016) showed persistence of the initial syntactic parse even in the absence of referential ambiguity and when comprehension was assessed in the form of a forced-choice question as we did in our studies.

1.4. Predictions for the RC and animacy manipulation for children and adults

To investigate whether children differ from adults in the time course and accuracy of their RC processing by the time they start formal education (4- to 6-year olds), we also tested a group of adults in both studies.

The predictions are that both children and adults should be more accurate with SRCs than ORCs, and that they should fixate earlier (faster) on the target picture with SRCs. As outlined earlier, SRCs are more frequent, they follow canonical word order, they are less syntactically complex than ORCs, and they impose fewer memory demands involved in thematic role assignment. With respect to the animacy manipulation, if it is the lexical animacy of the noun that matters, then we expect that ORCs with an inanimate NP1 will be interpreted faster and more accurately than ORCs with an animate NP1. Conversely, if participants are sensitive to the perceptual animacy of the NP1, we expect less of a facilitation effect in ORCs with inanimate NP1, as these inanimate nouns are perceptually closer to animates than to inanimates. For the same reason, we would expect SRCs with an animate NP1 to be processed faster and more accurately than SRCs with an inanimate NP1 if lexical animacy matters, but for this difference to be diminished if participants are sensitive to perceptual animacy. We have no principled reasons to expect adult—child differences in terms of lexical versus perceptual animacy.

1.5. The role of memory, language skills, and inhibitory control in RC interpretation

The interpretation of RCs taps into a series of cognitive and linguistic competences that affect children's performance and which may—at least partially—account for individual differences in task success. VSM and VWM are implicated in sentence-level comprehension as information in the unfolding sentence needs to be stored and later retrieved and integrated. Syntactically complex structures like RCs, particularly ORCs, require children to keep track of the relationship between the head of the RC and a phonologically empty trace of the extracted element which necessarily taps into memory resources. Furthermore, children's concurrent sentence-level comprehension skills are an indirect measure of their parsing skills and should therefore positively correlate with their ability to comprehend RCs. Children who have better sentence-level comprehension skills are expected to be more accurate and faster than children with lower sentence-level comprehension skills.

Some previous studies on the comprehension of RCs have included measures of VSM, VWM (Arosio, Adani, & Guasti, 2009; Arosio, Guasti, & Stucchi, 2011; Arosio, Yatsushiro, Forgiarini, & Guasti, 2012; Bentea et al., 2016; Booth et al., 2000; Boyle, Lindell, & Kidd, 2013; Haendler et al., 2015), and language abilities (Haendler et al., 2015). Studies including verbal memory measures have used the forward digit span task as a proxy to the short-term storage of phonological information, and the backward digit span task which is thought to tap into the operation of the central executive, the locus of coordination, and manipulation of the information stored in the phonological loop. Findings are mixed, with some studies reporting an effect of VSM on children's accuracy in the interpretation of RCs (Arosio et al., 2012; Booth et al., 2000), others reporting an effect of VWM (Arosio et al., 2009, 2011), and one study reporting a significant positive effect for a composite of VSM and VWM (Haendler et al., 2015). Although the evidence so far is somewhat inconsistent as to the relative contribution of VSM versus VWM in RC interpretation, higher memory capacity predicts higher accuracy. In the present study, we measured both VSM and VWM and expected a positive correlation with accuracy and reaction times in the offline RC interpretation task and earlier looks to the target picture in the online task.

In addition, the selection between an SRC or an ORC interpretation in the visual world paradigm requires a degree of inhibitory control in what Trueswell (2008, p. 74) defines as information re-characterization. Because children have more experience with SRCs than with ORCs, and because—all other things being equal—SRCs are less demanding, their first parsing decision upon hearing the relativizer *that* following an NP will be to pursue a SRC parsing decision and to look at a scene in which NP1 is the subject/agent of the transitive action. In the case in which an active transitive verb immediately follows SRC, for example, *The cow that is chasing…* this initial parsing decision is in alignment with the syntactic unfolding of the sentence. However, when the relativizer *that* is followed by an NP rather than a verb, the parser gets a syntactic cue to rescind and revise the initial SRC interpretation, that is, to look away from a visual scene in which NP1 is the subject/agent and to shift their eye movements to a referent that is the object/patient of the upcoming verb. To fixate on, and choose the correct target picture, participants need both to suppress the response to select the competitor picture and to ignore the interference from the competitor picture. Hence, we expect that children with better inhibitory control skills should be faster and more accurate in a task where they hear one sentence and have to choose between two competing visual stimuli.

While no previous studies have investigated the role of inhibition in connection with the interpretation of RCs, particularly ORCs, there are theoretically principled reasons why children with better inhibitory skills should have an advantage in the processing of these taxing syntactic constructions. Inhibition essentially requires the suppression of a dominant response—response suppression—and the filtering out of irrelevant information—interference control. Although response suppression and interference control are separate constructs (Friedman & Miyake, 2004; Harnishfeger, 1995), they are related and they are both involved in goal-directed behavior (Barkley, 1999). The idea that cognitive control is implicated in syntactic re-analysis is not new (Novick, Trueswell, & Thomson-

Schill, 2005; Woodward, Pozzan, & Trueswell, 2016). In more recent work, training studies with adults report a positive correlation between improved performance on a training task targeting conflict-resolution processes and gains in garden path recovery in syntactically ambiguous sentences (Novick, Hussey, Teubner-Rhodes, Harbison, & Bunting, 2014). Thothathiri, Asaro, Hsu, and Novick (2018) showed that engaging inhibitory control via a Stroop task facilitated the resolution of a syntax–semantics conflict in thematic role assignment and concluded that this conflict adaptation effect supports a causal link between inhibitory control and thematic role assignment in online sentence parsing.

One of the tasks that has been widely used to measure interference control is the flanker task (Fan, Flombaum, McCandliss, Thomas, & Posner, 2003). In the visual world paradigm adopted in the present study, children listened to SRC and ORC sentences and had to select from two pictures the one that was consistent with the sentence they had heard. The prepotent response that should be suppressed is the processing preference for a SRC interpretation when the sentence, in fact, matches an ORC interpretation. The prediction is therefore that children who have better inhibitory skills—as measured by the flanker task—should be more accurate and faster in selecting the correct picture in ORC sentences as they are better able to suppress the preferred SRC interpretation.

2. Experiment 1: Method

2.1. Participants

An a priori power analysis indicated that a sample of 48 participants would yield a power of .81 to detect an interaction β value of .15 in a generalized linear mixed-effects model (GLMM) using RC type and animacy as fixed factors. Accordingly, 48 children took part in this study (22 girls, mean age = 5;5, range = 4;5–6;4). Participants were recruited from Reception and Year 1 classes from a primary school in the North of England after obtaining ethical approval from the University Research Ethics Committee of the University of Manchester. All of the children were monolingual speakers of English and were developing typically according to class teachers' reports. The school received a book token as thanks for their participation. In all, 32 adults (24 women, mean age = 25, range = 18–36) also took part in the eye-tracking task (visual—world paradigm) only, and they were not compensated for their participation. The adult participants were undergraduate and postgraduate university students and university administrators at the University of Manchester.

2.2. Experimental task: Materials and design

The eye-tracking experiment used a 2 × 2 within-subjects design. The independent variables were the lexical animacy of the NP1 (animate or inanimate) and the type of RC used in the sentence (SRC or ORC). All lexically inanimate nouns were high on the perceptual animacy continuum as they were paired with just four verbs: "*following*," "*chasing*," "*bumping*," and "*hitting*." These verbs were chosen as they allowed for semantically plausible reversible sentences with a lexically inanimate head.

With six items in each condition, 24 experimental items were used in this experiment. Each item was made up of an audio sentence and a visual display. The sentences had four versions (see Table 1; Supplementary Material 1), one for each of the four 2 (RC type: SRC, ORC) × 2 (animacy of NP1: animate, inanimate) conditions: The sentences either had a SRC structure ("*Where is the [NP1] that is following the [NP2]?*") or an ORC structure ("*Where is the [NP1] that the [NP2] is following?*"). The NP2 was always an animate noun (one of twelve animal characters), but the NP1 was either one of these animates or one of twelve inanimate objects so that animacy was of the type NP1 *animate*–NP2 *animate* (animate conditions) or NP1 *inanimate*–NP2 *animate* (inanimate conditions). Each of the four verbs ("*following,*" "*chasing,*" "*bumping,*" and "*hitting*") was used in six experimental items.

The visual displays featured two transitive scenes in which the agent and the patient roles were reversed, for example a deer chasing a cow and a cow chasing a deer (Fig. 1a). Each item had an associated display with either an animate or inanimate head. The images of the depicted actions were 450 × 280 pixels in size and were displayed on the left and right of a 1,280 × 720 screen, centered 25% and 75% along the *x*-axis, respectively. Each display was counterbalanced so that each action appeared equally often across participants on the left and the right of the screen and so that each action was equally directed leftwards and rightwards. In total, there were eight unique versions of each display per item, which, combined with the SRC and ORC sentences, provided 16 unique sentence–display pairs.

In all, 12 filler items were produced with audio sentences that matched the experimental trials for word length (e.g., *Where is the gorilla jumping with the silly frog?*). These sentences were paired with a visual display that showed two pictures, one on the left and another on the right. One of the pictures matched the audio sentence (target), whereas the other did not (competitor). Crucially, as with the experimental sentences, it was only clear that the competitor picture did not match the sentence after the final word was uttered (for the example sentence above, the competitor was of a picture of a gorilla jumping with a monkey). Three of the fillers involved two animate characters and another three depicted just inanimate objects. The remaining six involved an animate character and an inanimate object. Six practice trials were also used. Four of these followed the format of the fillers and two featured just one animate character (e.g., *Where is the little donkey?*). See Supplementary Material 1 for a full list of filler items.

Table 1
Example subject relative clause (SRC) and object relative clause (ORC) sentences in the animate and inanimate conditions

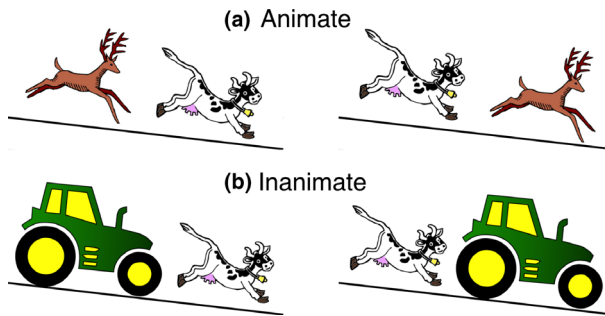| Relative Clause Condition	Animacy Condition	
	Animate	Inanimate
SRC	"*Where is the deer that is chasing the cow?*"	"*Where is the tractor that is chasing the cow?*"
ORC	"*Where is the deer that the cow is chasing?*"	"*Where is the tractor that the cow is chasing?*"

Fig. 1. Example images used in Experiment 1 for the (a) animate (with "deer" as NP1) and the (b) inanimate (with "tractor" as NP1) conditions.

2.3. Hardware, software, and eye-movement recording

The eye-tracking procedure was carried out on a Dell Precision M 4700 laptop computer and a Dell Latitude E 7450 Ultrabook, the latter of which has a 14-inch display that was used for stimulus presentation. The experiment was scripted and run using the SR Research Experiment Builder software. Eye-movement behavior was captured using a desk-mounted SR Research EyeLink 1000-Plus eye-tracker. This system uses corneal reflection and pupil position to calculate where a participant is fixating. Participants were positioned approximately 50 cm from the monitor and wore target stickers on their heads so that the tracker could track head position. Calibration involved the participant fixating on nine markers on the screen. Once calibrated, a verification procedure took place. If the verification procedure found mean spatial accuracy error to be more than 1.5 degrees or if any one of the spatial accuracy errors was >2 degrees, calibration and verification procedures were repeated. Before each trial, participants fixated a marker in the middle of the screen. This "Drift Checking" procedure allowed the experimenter to see the estimated fixation point on their display and required the experimenter to accept the fixation to begin the trial. If the error for this procedure exceeded 1.5 degrees of visual angle on three consecutive trials, the calibration procedure was repeated. A Microsoft Sidewinder gamepad was used for participant responses.

2.3.1. Language assessment

The test for reception of grammar (TROG-2; Bishop, 2003) was used to measure children's receptive syntactic skills. The test is a sentence–picture matching task with 20 blocks of four sentences each. The assessment was conducted and scored following the guidelines set out in the TROG-2 Manual (see Section 2.4. for details).

2.3.2. Executive function assessment

We used two tests from the computer-based Examiner battery (Kramer et al., 2014): the flanker task to measure inhibitory control, and the n-back task to measure visual working memory. These tasks were edited to suit the age range of this study; text was removed from the presentation, stimuli were enlarged, and presentation time was slowed.

The flanker and n-back tasks were conducted on a 14″ Lenovo laptop using the Examiner battery software and PSYCHOPY (Version 1.73.2; Peirce, 2007). For the flanker task, after the rules of the game were explained to the participant, a practice session of eight trials was initiated. If more than 75% of these trials were correct, then the test block began. If fewer than 75% of trials were correct, the practice session was repeated. There were 60 trials in the test block, split evenly between congruent and incongruent trials and randomly ordered.

In the n-back task, the participant was shown a 2.4 cm white square for 2,000 ms in one of 15 locations on a blank screen, followed by a centrally located number (1–9), which the participant had to immediately say out loud. After the number, another square would appear, either in the same location as the previous square or a different location. The participant was tasked with deciding if the square was in the same or different location as the previous square. They responded with the same keyboard keys as the flanker task ("M" for different, "Z" for same). There were 30 trials in this task, each involving one square. As with the flanker task, there was a practice session in which the participant had to score at least 70% before the test phase could begin.

In addition, we used the forward and backward Digit Span task from the Wechsler Intelligence Scale for Children (WISC-V; Wechsler, 2014) as a measure of VWM. The experimenter read digits from a record sheet and the children responded orally.

2.4. Procedure

Children took part in two sessions approximately 1 week apart. In the first session, they were administered the language assessment, and the executive function assessment over approximately 45 min; in the second session, they took part in the 20-min eye-tracking task. Adults only took part in the eye-tracking task. The order of the assessment tasks was kept constant across children: TROG-2, forward and backward digit span, flanker task, and n-back task.

2.4.1. Eye-tracking task

Testing for the children took place on school premises in a quiet space near their classroom. Adults were tested in a university lab and completed the same task as the children. Each participant was told that they would be playing a word and picture game. They were informed they would see two pictures on either side of the screen and that they would hear the recording of a lady speaking, after which they would choose the picture she was referring to by using the buttons on the gamepad. The participant then practiced pressing the "left" and "right" buttons on the gamepad. Once the experimenter was satisfied that the participant was comfortable with the gamepad, the eye-tracker was set up (see *hardware, software, and eye-movement recording*), and the practice session started—see Supplementary Material 1 for a full list of practice trial sentences. In each practice trial, as well as the experimental and filler trials, the picture was displayed for 2,000 ms before the sentence onset. At the point of onset of the final word in a sentence, the participant was able to press one of the two response buttons on the gamepad. Once a button was pressed on the

gamepad, the visual display would disappear. In the first three practice trials, the participant was shown a tick or a cross after the display disappeared, indicating whether their response was correct or incorrect. If correct, the participant was congratulated and encouraged to carry on. If incorrect, the experimenter explained why the response was incorrect and encouraged the participant to make sure they listened carefully and that they only pressed the button once they knew which picture the lady was speaking about. The final three practice trials did not involve the feedback stage. After completion of the practice stage, the experimental/filler session began. Participants each carried out 36 randomized trials, using each experimental and filler item once. As there were 16 versions of each experimental item, we used 16 item-lists that were balanced for conditions, target location, and action-direction. Each list was used for four participants, meaning that each version of each item was used four times across all participants.

2.5. Analysis

2.5.1. Test for reception of grammar

For each participant, the raw score of number of blocks passed was converted into a standardized score based on the children's age (in 6-month brackets) from the TROG-2 manual.

2.5.2. Digit span tasks

The raw scores from the forward and the backward digit span tasks were combined into a composite VWM score. This included the number of number-strings that were correctly repeated in the two tasks.

2.5.3. Flanker task

A flanker score was calculated using the EXAMINER software. This 0–10 range score is the sum of a score out of five for accuracy and RT for incongruent trials. The accuracy score is simply the proportion of correct responses multiplied by five. The RT scored is inversely proportional to the log (base 10) transformation of the median RT.

2.5.4. N-back task

A 1-back score was derived from the difference between the hit-rate and false-positive rate using the EXAMINER software.

2.5.5. Eye-movement task

The sample-level eye-movement data (500 samples per second) were outputted using the SR Research DATA VIEWER software and analyzed in R (R Core Team, 2016) using the *eyetrackingR* (Dink & Ferguson, 2016) package.

We used accuracy in the selection of the target picture, and RT as our behavioral-dependent variables in this task and also analyzed the eye movements of participants after the onset of the RC at the relativizer "that." During the RC, increased looks to the target

in the SRC condition relative to the ORC condition would be considered evidence of anticipation of an SRC.

2.6. Statistical analyses

2.6.1. Language and executive function measures

We ran simple correlations between each of these measures and our dependent variables. In cases where there was a significant correlation, we then included these in our GLMMs/LMMs.

2.6.2. Eye-movement experiment

We ran GLMMs on accuracy and linear mixed-effects models (LMMs) on RT. All of our models used RC type and animacy as fixed factors and participant and item as random factors. We used models with maximal random structures where possible, but in cases of non-convergence we simplified our models and followed the steps described by Barr, Levy, Scheepers, and Tily (2013). If a maximal model did not converge, we first removed the correlations between random effects. Then we removed random interaction slopes, followed by random slopes, until we found the model with the maximal random structure that successfully converged.

For our eye-movement data, we carried out cluster-based permutation analysis (Maris & Oostenveld, 2007) to identify areas of significant divergence of fixations between conditions. This involved two steps: First, for the animate and inanimate conditions, we carried out GLMMs on the proportion of target fixations between the RC conditions on each 50 ms time-bin for the 2,500 ms following the onset of the RC. We identified any groups of adjacent time-windows with $z > 1.96$ as time-clusters and summed each cluster's z-statistics (i.e., a cluster of three adjacent time-bins each with a z-statistic value of 2.5 would have a sum-statistic of 7.5). Next, we repeated this process on 2,000 shuffled (randomized) datasets and found the proportion of these datasets that have clusters with sum-statistics as large as or larger than the sum-statistics of the clusters in our data. This proportion tells us the chances of finding any of our clusters in our data, assuming there was no real effect of RC condition. Therefore, this value was used as the p value for the divergence between the two conditions. These analyses were carried out in R (R Core Team, 2016) using the *eyetrackingR* package (Dink & Ferguson, 2016). The raw data and analysis scripts for Experiments 1 and 2 can be found on our project page on the Open Science Framework website (https://osf.io/v59yk/)

3. Experiment 1: Results and discussion

3.1. Children

We found no significant correlation between age and accuracy, $r(46) = .166$, $p = .260$ for the children's data, and so did not use this factor in any of our models.

The accuracy results are reported in Fig. 2. As hypothesized, children were more accurate in selecting the target picture with SRC than ORC sentences, and this was found to be significant in a GLMM of accuracy, $\beta = .983$, $SE = .164$, $t = 5.995$, $p < .001$ (Model 1. A full list of the models is shown Supplementary Material 2). Animacy did not have an overall influence on accuracy, $\beta = .221$, $SE = .125$, $t = 1,263$, $p = .210$ (Model 2^1); however, there was a significant interaction between RC type and animacy in both models, $\beta = .712$, $SE = .323$, $t = -2.203$, $p = .029$ (Model 1), $\beta = .699$, $SE = .324$, $t = -2.154$, $p = .033$ (Model 2). Children performed better with ORCs in the inanimate condition compared to the animate condition, suggesting that overall they were sensitive to the lexical animacy of the NP1 when encountering ORCs.

The RT results (from the onset of the RC) are shown in Fig. 3. A LMM of ($\log_{10}$) RT (Model 3) showed a significant effect of RC type, $\beta = .130$, $SE = .016$, $t = 8.382$, $p < .001$, with shorter RTs for SRCs, a smaller, but still significant effect of animacy, $\beta = .052$, $SE = .014$, $t = 3.617$, $p < .001$, with longer RTs for RCs with an animate head, and no interaction between RC type and animacy, $\beta = -.031$, $SE = .029$, $t = -1.091$, $p = .274$.

To investigate children's online processing of these sentences, we analyzed their eye movements after the onset of the RCs (*that...*). We separated the 2,500 ms (the approximate duration of the RC) after the onset of the RC into 50 ms time-bins. In Fig. 4, we have plotted the proportion of looks to the target picture (out of all looks to either the target or competitor) for SRC and ORC sentences in the animate (Fig. 4a) and inanimate conditions (Fig. 4b).

Fig. 4a shows that in the animate condition, after the onset of the RC, looks to the target increase in SRC sentences, but decrease in ORC sentences, indicating a preference for the SRC picture and thus suggesting anticipation of an SRC. Our GLMMs on each 50 ms time-bin revealed a cluster of five significant time-bins (650–1,000 ms). However, cluster-based permutation analysis (Maris & Oostenveld, 2007) showed this cluster to be nonsignificant (summed $z = 17.844$, $p = .108$). We also found a smaller cluster of two time-bins (1,750–1,850 ms), which was also found to be nonsignificant (summed $z = -4.385$, $p = .379$). In contrast to the accuracy and RT data, there is no evidence for a significant advantage of clause type in the looking data.
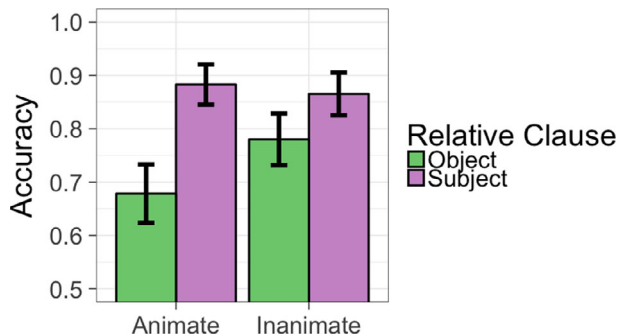

Fig. 2. Accuracy results for children in Experiment 1. Data are split by animacy and relative clause condition. Error bars show 95% confidence interval.
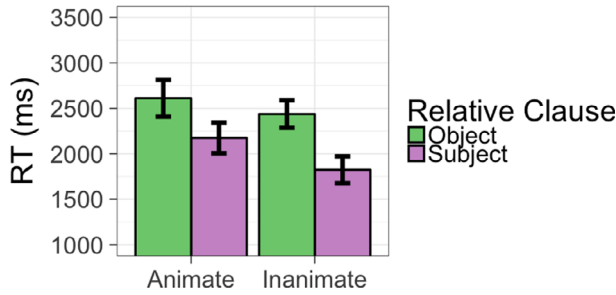
R. Macdonald et al. / Cognitive Science 44 (2020)

Fig. 3. The children's response time (RT; after the onset of the relative clause) for correct trials in the eye-tracking experiment. Error bars show 95% confidence intervals.
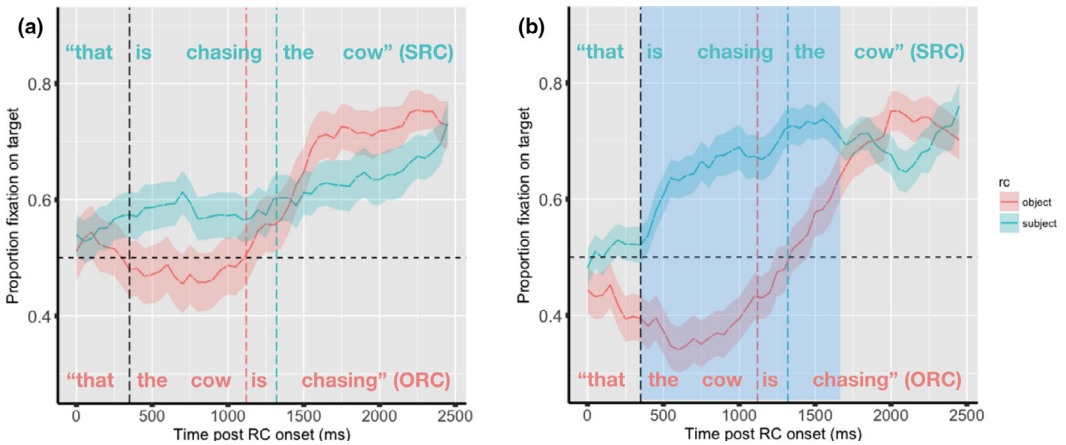

Fig. 4. The proportion of fixations on the target (out of all fixations to the target and competitor) for children for the 2,500 ms following the onset of the relative clause (RC) for the (a) animate and (b) inanimate conditions. Shaded area shows the area of significant divergence between the RC conditions.

Fig. 4b shows the much more extreme divergence between the SRC and ORC conditions. With an inanimate NP1, children were more likely to fixate on the picture depicting a SRC, regardless of RC type. The GLMMs of each time-bin showed a cluster of 26 significant time-bins (350–1,650 ms). Cluster-based permutation analysis showed the divergence at this cluster to be significant (summed $z = -117.779$, $p < .001$).

The eye-tracking data are in contrast to the accuracy data. In the animate condition, our eye-tracking results do not show evidence for a higher proportion of target fixations for SRCs. In the inanimate condition, the fixation patterns do not indicate that children are more likely to treat the lexically inanimate NP1 as the head of an ORC. On the contrary, children fixated on the picture in which the inanimate referent was the subject/agent, regardless of RC type. We discuss this pattern further after the presentation of the adult data.

3.2. Adults

The accuracy results for adults are shown in Fig. 5a. As expected, adult participants performed at ceiling in each condition. The RT results for adults are shown in Fig. 5b. A LMM of ($\log_{10}$) RT (Model 4) showed a significant effect of RC type, $\beta = .130$, $SE = .012$, $t = 10.628$, $p < .001$, but no significant effect of animacy, $\beta = .010$, $SE = .018$, $t = .582$, $p < .551$, nor any interaction between these factors, $\beta = -.024$, $SE = .024$, $t = -1.036$, $p = .304$.

We analyzed adult eye movements in the same way as the children's. Fig. 6 shows the proportion of looks to the target picture (out of all looks to either the target or competitor) for SRC and ORC sentences in the animate (Fig. 6a) and inanimate conditions (Fig. 6b).

Fig. 6a shows that the pattern of looks to the target after RC onset in the animate condition did not vary between SRC and ORC sentences. The GLMMs of each time-bin revealed no significant clusters of divergence between the RC conditions. Fig. 6b, however, does show an early difference between SRC and ORC conditions, showing that adults were more likely to look at the SRC image after RC onset in the inanimate condition. This 16-time-bin (200–1,000 ms) cluster of divergence between the conditions was found to be significant (summed $z = 50.210$, $p = .004$). Two other clusters were identified, but neither of these was found to be significant (0–100 ms, summed $t = 4.694$, $p = .368$; 1,800–2,100 ms, summed $z = 13.024$, $p = .177$). Like the children's eye-tracking data, the adult data provide evidence for the anticipation of SRCs in the inanimate condition but not the animate condition.

3.3. Language measures

The standardized TROG scores were normally distributed (mean score = 102, range = 62–137) and they correlated significantly with accuracy, $r(46) = .464$, $p < .001$ (Fig. 7). Since we are interested in whether TROG score significantly interacted with animacy and RC condition, we included it in a GLMM for accuracy. TROG score was found
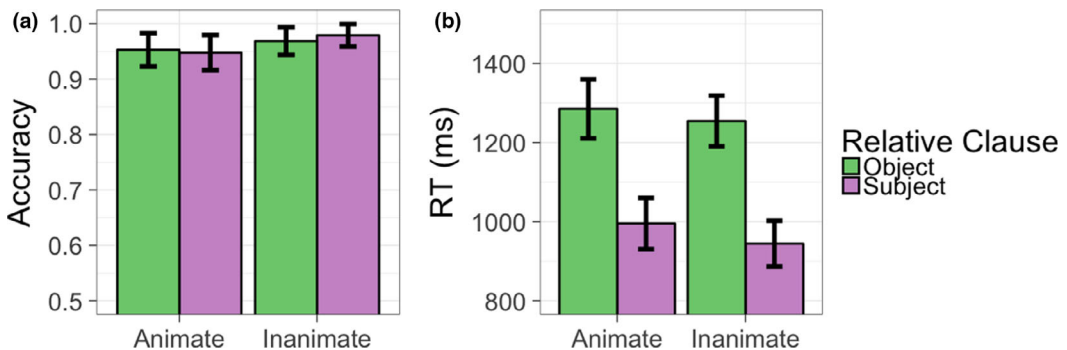

Fig. 5. (a) Accuracy results for adults and (b) response time (RT; after the onset of the relative clause [RC]) for correct trials. Data are split by animacy and RC condition. Error bars show 95% confidence interval.
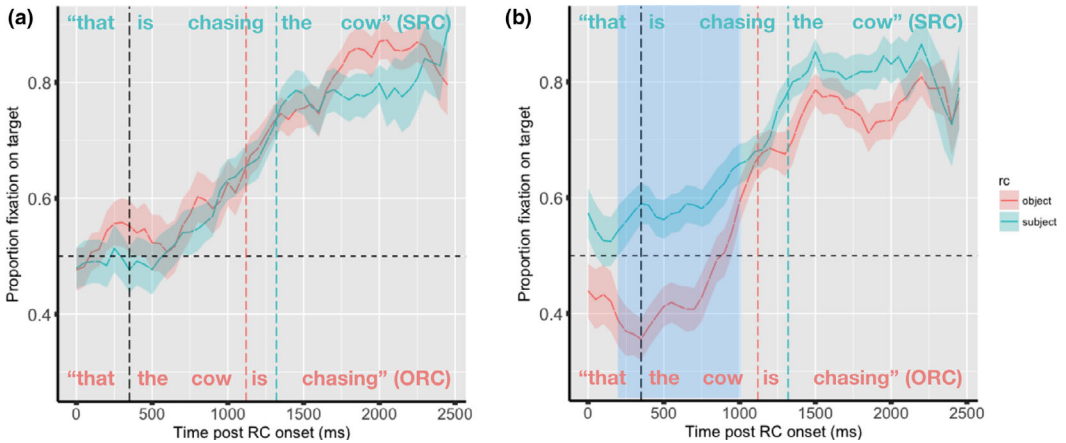
Fig. 6. The proportion of fixations on the target (out of all fixations to the target and competitor) for adults for the 2,500 ms following the onset of the relative clause (RC) for the (a) animate and (b) inanimate conditions. Shaded area shows the area of significant divergence between the RC conditions.

to be significant, $\beta = .370$, $SE = .099$, $z = 3.789$, $p < .001$ (Model 5). In this model, RC type was again found to be significant, $\beta = -.950$, $SE = .166$, $z = -5.718$, $p < .001$, and the interaction between RC type and animacy marginally significant, $\beta = -.647$, $SE = .331$, $z = -1.953$, $p = .051$. No significant three-way interaction or any two-way interactions were found between TROG score and animacy or RC type (all ps $> .125$). We found no significant correlations between TROG score and RT.

3.4. Executive function measures

Six children did not complete the flanker task and 18 failed to complete the n-back task; therefore, scores were not collected for these children and they feature in the analyses as missing data. The score on the flanker task was a composite of accuracy and speed and it had a maximum of 10, the distribution was negatively skewed with the majority of children getting scores between 6 and 8 (mean score = 6.47, range = 2.02–8.48). The digit span task score was a composite of the forward and backward recall and had a maximum score of 16; the distribution had a wide range, but it was slightly positively skewed with the majority of children getting scores between 6 and 10 (mean score = 8.7, range = 2–15). The n-back task was particularly challenging for children and although the distribution of scores was broadly standard the scores were overall low (mean score = 1.05, range = −1.39 to 2.33). We found no correlation between accuracy and digit span score, $r(46) = .162$, $p = .272$, flanker task score, $r(40) = .155$, $p = .328$, or n-back score, $r(28) = .273$, $p = .145$. Because of the absence of a correlation, we did not include these factors in any of our models. We found no correlations between RT and any of our executive function measures.
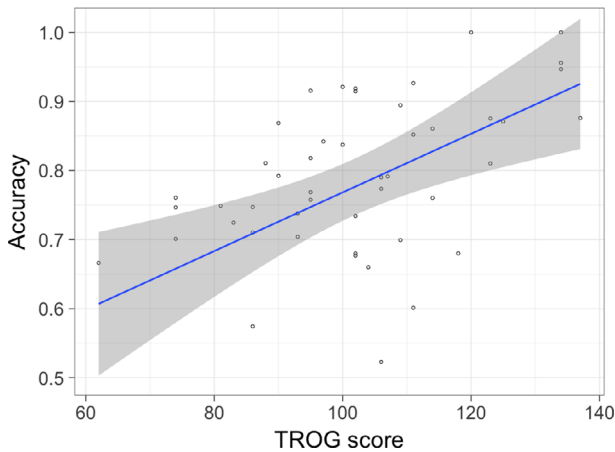
Fig. 7. The accuracy and test for the reception of grammar (TROG) scores for each participant aggregated across condition.

4. Experiment 1 summary

Experiment 1 investigated the effects of RC type and NP1 animacy on children's and adults' online and offline processing of RCs. For the children, we also factored in measures of language proficiency and executive control, none of which were either significantly correlated with accuracy (executive function measures) or interacted with RC type or animacy (language measure).

The children's offline accuracy data are consistent with the hypothesis that SRCs are easier to comprehend than ORCs, and with the hypothesis that ORCs with a lexically inanimate NP1 are easier to comprehend than ORCs with an animate NP1. The RT data also show that SRCs with either an animate or an inanimate NP1 are processed more quickly than ORCs. These results are in line with previous findings in the developmental literature (Adani et al., 2010; Booth et al., 2000; Brandt et al., 2009). The adults were at ceiling in terms of accuracy for all RC types, and they were faster with SRCs, regardless of head animacy, again a result that is consistent with the literature (Mak, Vonk, & Schriefers, 2002, 2006; Traxler et al., 2002).

In the children's online eye-tracking data, however, we did not find a facilitatory effect of RC type in the animate condition, and we found no evidence that encountering an inanimate NP1 led them to expect an ORC. This pattern of eye movements was again found for the adult participants, with no differences between SRCs and ORCs in the animate condition, and an early preference for an incorrect interpretation of ORCs as SRC in the inanimate condition. For the adults, the temporal window in which they fixated longer on the SRC picture for ORC sentences was restricted to the 200–1,000 ms after the onset of the RC, while for the children the window lasted for an additional 500 ms in the 350–1,650 ms interval.

These two disparate sets of findings can be reconciled if we consider the lexical versus perceptual animacy of the inanimate NP1 and how the offline and online measures are

differently affected by these two facets of animacy. The eye-tracking data tap into the unfolding interpretation of the spoken sentences while participants were inspecting the visual scenes. Because of the semantic reversibility of the transitive scenes, we had to select verbs where both an animate and an inanimate referent could plausibly serve as the subject/agent. This restriction led us to settle on verbs that were associated with motion of some description ("*chasing*," "*following*," "*hitting*," and "*bumping*"). The inanimate referents included in these scenes were therefore perceptually animate to some extent, and their perceptual animacy made them more agent-like and more animate than would be expected by their lexical animacy status alone. Given this visual setup, the fact that an inanimate NP1 like *tractor*, when paired with an animate NP2 like *cow* in a chasing/following scene, could be initially construed as the head of a SRC is not particularly implausible. As participants process an incoming RC in the eye-tracking experiment, they are faced with a choice between two visual stimuli where the lexically inanimate referent is perceptually more like an animate than an inanimate and therefore are temporarily led to consider the inanimate as a potential agent and subject of the RC. By 1,000 ms after the onset of the RC adults revise their incorrect SRC analysis for an ORC, while the children persevere in looking at the incorrect picture for 650 more ms.

What is less obvious is why a lexically inanimate NP1, albeit one that has some degree of animacy due to its motion property, would drive more SRC interpretations than a bona fide lexically animate NP1. We have considered two possible explanations for this seemingly counter-intuitive finding. First, it may be that there is increased interest in the sentences with an inanimate NP1 driven by a *surprisal* effect triggered by the mismatch between the lexical inanimacy of a referent (e.g., tractor) and its agency in the visual scene (e.g., chasing a cow). Inanimate entities as a rule are not involved in agent-like actions; therefore, seeing a lexically inanimate referent that is behaving like an animate entity is unexpected and surprising. Assuming a general online anticipation of an SRC, surprisal may result in a boost in fixations in the inanimate condition immediately after RC onset. Alternatively, it may simply be that the SRC image was easier to immediately identify in the inanimate condition than the animate condition because in the latter there is similarity-based interference between two similar animates (Gordon, Hendrick, Johnson, & Lee, 2006; Humphreys, Mirkovic, & Gennari, 2016; Jäger, Engelmann, & Vasishth, 2017). A deer and a cow are perceptually more similar—and hence more likely to compete with each other for the subject/agent role (Fig. 1a)—than a cow and a tractor (Fig. 1b). We carried out Experiment 2 to discriminate between these two alternative explanations. Experiment 2 was largely similar to Experiment 1, but crucially differed in that the NP2 in each sentence was inanimate. Therefore, instead of using NP1*animate*–NP2*animate* and NP1*inanimate*–NP2*animate* conditions as in Experiment 1, we used NP1*animate*–NP2*inanimate* and NP1*inanimate*–NP2*inanimate* conditions. By altering the design in this way, our *surprisal* and *similarity-based interference* explanations lead to opposing predictions. If the increase in fixations to the SRC image in the inanimate condition was caused by a *surprisal* effect, then we should replicate these findings in Experiment 2. NP1*inanimate*–NP2*inanimate* configurations (inanimate condition) should lead to higher anticipation of SRCs than NP1*animate*–NP2*inanimate* configurations (animate

condition). However, if the *similarity-based interference* hypothesis is correct, then in Experiment 2 the animate condition (NP1*animate*–NP2*inanimate*) should be easier than the inanimate (NP1*inanimate*–NP2*inanimate*) condition, resulting in an increase in fixations to the SRC image in the animate condition.

5. Experiment 2: Method

5.1. Participants

In all, 45 monolingual English-speaking children who had not taken part in Experiment 1 participated in this study (21 girls, mean age = 5;11, range = 4;5–6;9). In all, 40 participants were recruited from Reception and Year 1 classes from another primary school in the North of England after obtaining ethical approval from the University Research Ethics Committee of the University of Manchester. All of the children were developing typically according to class teachers' reports. A further five were recruited from a participant database at the University of Manchester. The school received a book token as thanks for their participation. The caregivers of the five children tested outside of school were reimbursed for their travel costs. In total, 32 adults (23 women, mean age = 26, range = 18–41) also took part in the eye-tracking task only; the adults were not compensated for their time. The adult participants were undergraduate and postgraduate university students and university administrators at the University of Manchester.

5.2. Materials

Similarly to Experiment 1, 24 experimental items were used, with sentences following the same structure as those in Experiment 1a (SRC = "*Where is the [NP1] that is following the [NP2]*", ORC = "*Where is the [NP1] that the [NP2] is following*"). However, in this experiment, NP2 was always inanimate, and NP1 was either animate or inanimate. Table 2 shows the example sentences in each of our four conditions (full list in Supplementary Material 3). Instead of using the four verbs used in Experiment 1a, we used only "*following*" and "*chasing*" because of the difficulty in depicting an inanimate "*bumping*" or "*hitting*" another inanimate. New visual stimuli were created to match these sentences following the same specification as Experiment 1. We also reused all of the filler items,

Table 2
Example subject relative clause (SRC) and object relative clause (ORC) sentences in the animate and inanimate conditions

Relative Clause Condition	Animacy Condition	
	Animate	Inanimate
SRC	*"Where is the elephant that is chasing the ball?"*	*"Where is the bike that is chasing the ball?"*
ORC	*"Where is the elephant that the ball is chasing?"*	*"Where is the bike that the ball is chasing?"*

except for one because one of the objects was used in our Experimental items. This filler item was replaced with a new filler item (see Supplementary Material 3).

The same language measure (TROG-2) and executive function measures (flanker task and composite VWM from forward and backward digit recall) were included in this experiment for the children, with the exception of the visual working memory task from the procedure (n-back task), due to the large number of children that could not successfully complete it in Experiment 1 ($N = 18$).

5.3. Procedure and analysis

The procedures and analyses were identical to those of Experiment 1.

6. Experiment 2: Results and discussion

6.1. Children

The accuracy data from the children are shown in Fig. 8a. A GLMM found that accuracy was significantly higher for SRC sentences compared to ORC sentences, $\beta = -.170$, $SE = .287$, $t = -3.727$, $p < .001$ (Model 6). No significant main effect was found for animacy, $\beta = .512$, $SE = .275$, $t = 1.860$, $p = .069$ (Model 7), nor was any interaction found between RC type and animacy, $\beta = .046$, $SE = .432$, $t = .106$, $p = .916$ (Model 6), $\beta = .038$, $SE = .430$, $t = .089$, $p = .930$ (Model 7). Similarly, a LMM for the reaction-time data revealed a significant difference between the SRC and ORC conditions, with SRC sentences responded to more quickly than ORC sentences, $\beta = .152$, $SE = .018$, $t = 8.397$, $p < .001$ (Model 8). No significant effect of animacy was found, $\beta = -.027$, $SE = .018$, $t = -1.523$, $p = .130$ (Model 8), nor any interaction, $\beta = .043$, $SE = .035$, $t = 1.223$, $p < .219$ (Model 8).

The eye-tracking data from the children (Fig. 9) show that in the animate condition, there was no significant difference in looks to the target between the RC conditions
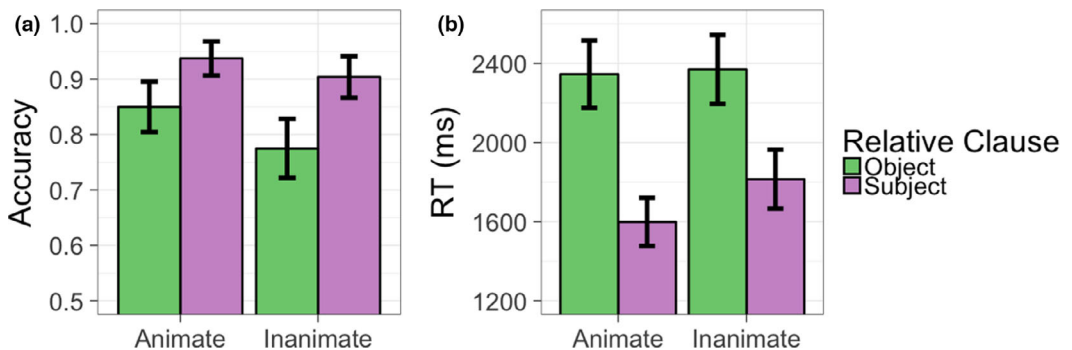

Fig. 8. (a) Accuracy results for children and (b) response time (RT; after the onset of the relative clause) for correct trials. Data are split by animacy and relative clause (RC) condition. Error bars show 95% confidence intervals.

immediately after the onset of the RC. However, throughout most of the RC, there was a higher proportion of looks to the target in the SRC, but this was only significant during one period (1,000–1,850 ms, summed $z = -48.396$, $p = .001$). In the inanimate condition, there was a large cluster (350–1,300 ms) in which there was a significantly higher proportion of looks to the target in SRC sentences compared to ORC sentences, summed $z = -78.567$, $p < .001$. These data follow the same pattern as those found in Experiment 1, showing increased looks to the SRC image after RC onset in the inanimate condition, but not the animate condition. This provides support for our *surprisal* explanation of the results from Experiment 1, rather than a similarity-based interference account, as the strong preference for the SRC image after RC onset was again present in the condition with an inanimate NP1 (i.e., an inanimate–inanimate pairing) in the presence of two inanimate nouns.

6.2. Adults

As with Experiment 1, adult accuracy reached ceiling (Fig. 10a) and RT was significantly affected by RC type (Fig. 10b), $\beta = .152$, $SE = .018$, $t = 8.400$, $p < .001$ (Model 9), with SRC sentences responded to more quickly than ORC sentences. There was no effect of animacy, $\beta = -.027$, $SE = .018$, $t = -1.550$, $p = .130$ (Model 9), nor was there an interaction between these two factors, $\beta = .043$, $SE = .035$, $t = .122$, $p = .219$ (Model 9).

Fig. 11 shows the eye-movement data during the RC for adults in Experiment 2. The pattern of results is very similar across the two animacy conditions. There is an initial preference for the SRC image in both conditions after the onset of the RC, but this difference does not reach significance in either the animate (summed $z = 7.008$, $p = .328$) or the inanimate (summed $z = 17.029$, $p = .067$) conditions.
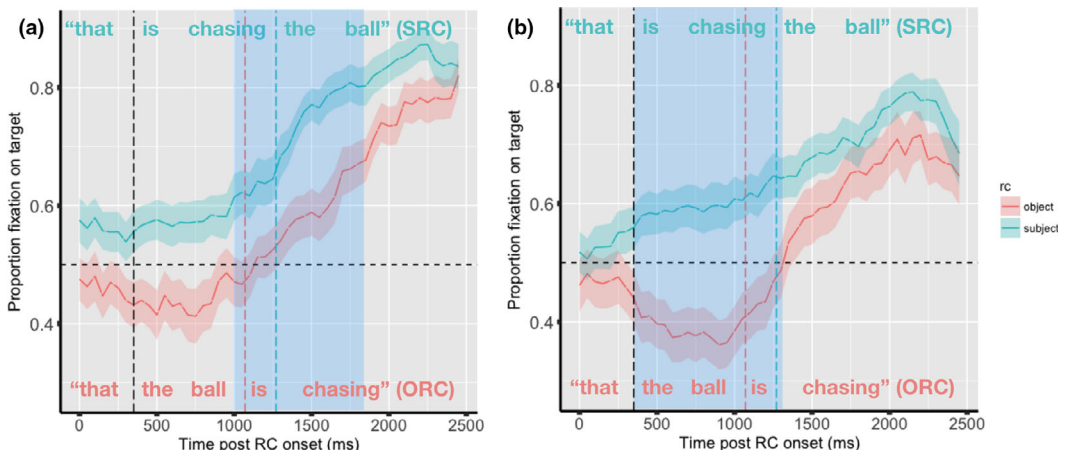

Fig. 9. The proportion of fixations on the target (out of all fixations to the target and competitor) for children in Experiment 2 for the 2,500 ms following the onset of the relative clause (RC) for the (a) animate and (b) inanimate conditions. Shaded area shows the area of significant divergence between the RC conditions.
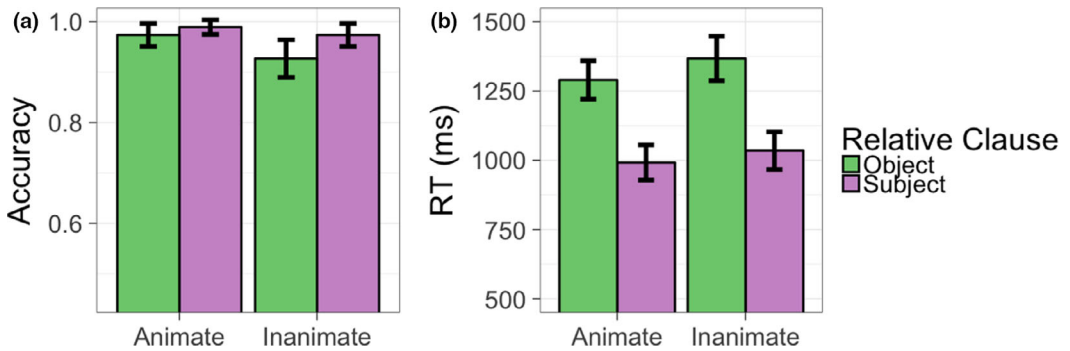
Fig. 10. (a) Accuracy results for adults and (b) response time (RT; after the onset of the relative clause) for correct trials. Data are split by animacy and relative clause (RC) condition. Error bars show 95% confidence interval bars.
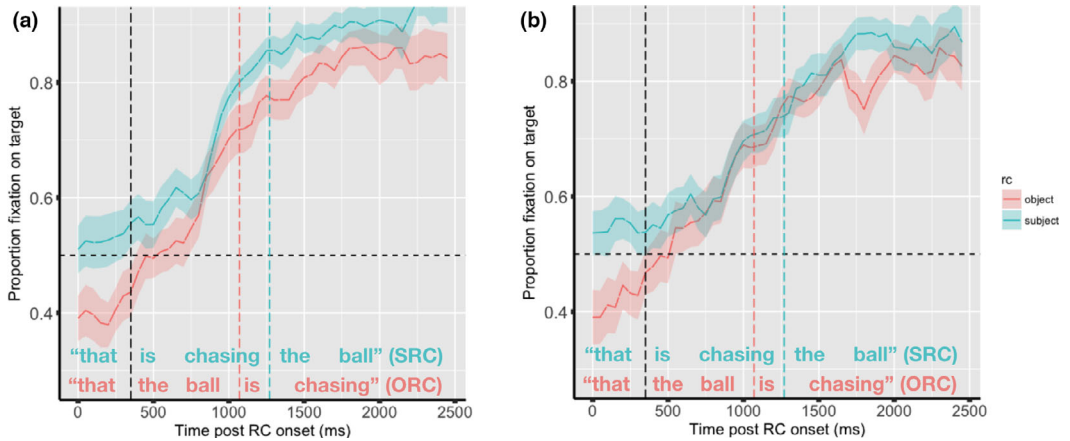

Fig. 11. The proportion of fixations on the target (out of all fixations to the target and competitor) for adults in Experiment 2 for the 2,500 ms following the onset of the relative clause for the (a) animate and (b) inanimate conditions.

6.3. Individual differences measures

We found no significant correlation between age (in months) and accuracy in Experiment 2, $r(43) = .269$, $p = .074$, and so did not use this factor in any of our models.

Unlike for the children in Experiment 1, where the TROG standard scores were normally distributed, in Experiment 2 we found a negatively skewed distribution (mean = 100, range = 62–130). Five children did not successfully complete the flanker task and, similarly to what we found in Experiment 1, the scores were negatively skewed with the majority of children getting scores between 6 and 8 (mean score = 6.89, range = 2.96–8.47). The composite working memory score was slightly positively skewed (mean = 9.5, range = 5–13). Due to the challenging nature of the n-back task and the

large number of children who could not complete the test in Experiment 1, we did not include it in Experiment 2.

There were no significant correlations between accuracy and TROG score, $r(43) = .260$, $p = .085$, flanker score, $r(41) = .011$, $p = .942$, or digit-span score, $r(43) = .271$, $p = .071$; therefore, we did not include these measures of individual differences in the models.

7. Experiment 2 summary

In Experiment 1, children were more accurate with NP1*inanimate*–NP2*animate* ORCs than NP1*animate*–NP2*animate* ORCs, but we found evidence of greater anticipation for an SRC as the RC unfolded in the NP1*inanimate*–NP2*animate* condition compared to the NP1*animate*–NP2*animate* condition. With Experiment 2, we investigated whether the findings of the eye-tracking task in Experiment 1 were due to similarity-based interference between the two animate referents in the NP1*animate*–NP2*animate* condition, or to surprisal at the depiction of an inanimate as an agent in the NP1*inanimate*–NP2*animate* condition. We found, again, that there was an increase in looks to the SRC image when NP1 was inanimate in the NP1*inanimate*–NP2*inanimate* condition, but not when the NP1 was animate in the NP1*animate*–NP2*inanimate* condition, suggesting that the SRC preference for this image was due to surprisal at the unexpectedness of seeing an inanimate-as-agent in the pictures. It is perhaps important to note, however, that this preference was not as strong for the children as it was in Experiment 1, and this preference was not found for adults at all in Experiment 2. This may be because the depiction of an inanimate agent and an animate patient was more surprising/interesting in Experiment 1 than the depiction of an inanimate agent and an inanimate patient in Experiment 2. An inanimate agent is unexpected and surprising and therefore salient, but our findings show that this salience is crucially also a function of the animacy of the patient. When the patient is also inanimate as in the NP1*inanimate*–NP2*inanimate* condition in Experiment 2, there is no animacy differential between the two NPs; this lack of an animacy mismatch between the agent and the patient seems to be flattening the interest in the fact that the agent is an inanimate.

Behavioral results for the adults followed the same pattern in Experiments 1 and 2, but the patterns differed for children. In Experiment 2, there was no interaction between RC type and animacy. Therefore, there was no indication that children found ORCs with animate NP1 heads (and inanimate NP2 subjects) any more difficult than ORCs with inanimate NP1 heads (and inanimate NP2 subjects).

8. General discussion

In two studies, we investigated the effect of head-noun lexical and perceptual animacy on English-speaking children's and adults' offline and online processing of SRCs and ORCs. For the children, we also explored whether individual differences in receptive syntactic skills, verbal and non-VWM, and inhibitory control affected the accuracy of their

interpretation. In the following, we discuss the results of the offline data, the online data, and the role of individual differences.

In Experiment 1 when NP2 was always animate, children were more accurate and faster at comprehending SRCs than ORCs, and they were more accurate with ORCs with inanimate NP1 heads than animate NP1 heads. For the adults, we only found an effect of RC type with overall faster RTs for SRCs. The eye-movement data, however, paint a somewhat different picture. In the animate condition, there was no facilitation for SRCs for either the children or the adults; in both groups, there were significantly more looks to the target picture after the relativizer for both RC types. In the inanimate condition, a lexically inanimate NP1 did not drive children or adults to expect an ORC. On the contrary, both groups were more likely to anticipate a SRC when the head of the RC was lexically inanimate (e.g., "*tractor*"). In Experiment 2, when the NP2 was always inanimate, children were more accurate and faster with SRCs, while RC type only affected adults' RTs as they were at ceiling in terms of accuracy. The eye-movement data did show increased anticipation for a SRC in the inanimate condition for the children—but not the adults—as the RC unfolded.

With the exception of receptive syntactic skills in Experiment 1, the individual difference measures we collected for the children did not show significant correlations with the accuracy data and therefore did not contribute any meaningful additional data to our modelling. Below we outline the possible implications of our findings from the offline data, the online data, and the individual differences results.

8.1. Offline data

Children's accuracy in Experiment 1 indicates that when it came to ORCs with an animate NP2, they were most successful when the head NP1 was lexically inanimate. This suggests that the presence of an inanimate NP1 facilitates children's comprehension of ORCs, supporting previous research on adults and children alike (Betancort et al., 2009; Kidd et al., 2007; Traxler et al., 2002). The relative ease of comprehending these inanimate ORCs could be due to the degree of exposure to these particular types of construction: Active ORCs (7) with inanimate head nouns are more commonly used than passive SRCs with the same meaning (8) (Gennari & MacDonald, 2009). Conversely, active ORCs with an animate head (9) are used more rarely than passive SRCs (10) (Humphreys et al., 2016). Children's increased familiarity with constructions such as (7) compared to (9) may have resulted in their greater accuracy in the inanimate ORC condition.

7. The truck the boy is pulling
8. The truck being pulled by the boy
9. The girl the boy is pulling
10. The girl being pulled by the boy

Aside from previous experience with ORCs headed by lexically inanimate nouns, the semantic appropriateness of the NP1 as a subject—which is intimately related to frequency—is likely to have played a role. In English, the NP1 is favored as the subject of

a sentence (Järvikivi, van Gompel, Hyönä, & Bertram, 2005). According to the *topichood hypothesis* (Mak, 2001), the subject of an RC is determined by the *topicworthiness* (or appropriateness as topic) of a noun. Head nouns are generally highly topic worthy, but this can be modulated by semantic factors, such as their animacy. Since, conceptually, inanimates are less likely to act on animates than other animates are, an inanimate NP1 is less topic-worthy than an animate NP1 and therefore less likely to be the head of a SRC. Following this hypothesis, children in Experiment 1 may have been more inclined to attribute an animate NP1 in the ORC to the role of subject, leading to more errors, and providing evidence that frequency and lexical semantics significantly affect syntactic interpretation in 4- to 6-year-old children. In contrast, unlike the children, the adults were at ceiling in the accuracy task, regardless of RC type and head animacy. For these mature and competent speakers, reliance on the word order syntactic information overrode the semantic and frequency factors that affected the younger learners in their interpretation of ORCs.

In Experiment 2, when NP2 was always inanimate, there was no interaction between the animacy of NP1 and RC type in terms of accuracy. Children in this experiment were no more accurate with ORCs with an inanimate NP1 than with an animate NP1. Considering the sentences in isolation, the *topichood hypothesis* cannot account for these data, as the animacy of the NP2 should have no effect on the *topicworthiness* of the NP1. However, in our experiments, the sentences were not heard in isolation, but were accompanied by visual depictions that were presented from 2 s before the sentence onset. It is possible that by inspecting the images before sentence onset in the inanimate–inanimate condition, children were aware they were about to hear a sentence with two inanimates; therefore, the effect of an inanimate head-noun on *topicworthiness* may have been diminished relative to Experiment 1 because of the inanimate status of the NP2.

The behavioral results discussed above provide insight into the effect of lexical animacy on the explicit comprehension of RCs and are consistent with the results of previous research on the offline interpretation of RCs with animate and inanimate heads (Adani, Stegenwallner-Schütz, & Niesel, 2017; Brandt et al., 2009; Kidd et al., 2007). Comparing children and adults with the same experimental materials has also highlighted the degree to which children—unlike adults—are affected by the frequency of SRCs, and the lexical animacy of the NP1 in ORCs.

8.2. Online data

By measuring their eye movements, alongside accuracy, and RT data, we have also been able to gain insight into the implicit parsing choices children and adults make during RC processing. In Experiment 1, there was a clear and significant preference for the SRC image in the NP1inanimate–NP2animate condition as the RC unfolded, but not in the NP1animate–NP2animate condition. A similar, but less marked effect was found for adult participants. These results suggest increased anticipation for a SRC when hearing an inanimate NP1 rather than an ORC, a result which is prima facie in contrast to the accuracy data results. Differently from previous developmental studies investigating the

role of animacy on RC interpretation (Brandt et al., 2009; Kidd et al., 2007—but see Adani et al., 2017; Bentea et al., 2016 for different animacy manipulations), we used semantically reversible sentences in all conditions, regardless of the lexical animacy of NP1—thus removing the semantics of the verb cue to disambiguating between SRC and ORC readings. Using semantically reversible scenes constrained us in the type of verbs we could use, and it automatically increased the perceptual animacy of the inanimate referents as they had to be plausibly interpreted as subjects/agents. The early preference for a SRC reading upon hearing an inanimate noun like *tractor* is unexpected on the assumption that the lexical animacy of the head is a strong enough probabilistic cue to bias an ORC interpretation. However, the head noun was associated with a motion event and all pictures were displayed on an incline suggesting downward movement (see Fig. 1). We argue that the motion context boosted the perceptual animacy of the inanimate referents and consequently it increased the likelihood that an inanimate noun could be taken to be the head of an SRC rather than the head of an ORC. This account, however, does not explain why participants should be more likely to think that an inanimate noun—even one with contextually given perceptual animacy—should drive more SRC interpretations than an actual animate head noun. Furthermore, these data also seem incongruent with our behavioral results. It is possible that the image depicting an inanimate agent acting on an animate patient may have been a more visually interesting image to look at than all the others. However, prior to the onset of the RC, there was no preference for these images. A general bias for SRC anticipation coupled with a surprisal-driven interest in the inanimate-as-subject image may have led to a boost in fixations to this image. Alternatively, we speculated that the increased preference for the SRC image in the inanimate condition might have been due to similarity-based interference between the two animate agents in the animate condition. The SRC image in the inanimate condition could have been easier for the children to quickly identify, resulting in what appears to be increased anticipation for a SRC. This finding also speaks to the need to consider the role of the animacy of the NP1 in the broader context of the RC, including the way in which the animacy of the NP2 can jointly affect the interpretation of the sentence. This was the rationale for changing the animacy of the NP2 in Experiment 2, while still manipulating the animacy of the NP1; this allowed us to distinguish between a *surprisal* and *a similarity-based interference* account. Even when both NP1 and NP2 were inanimate, we again found a large preference for fixations to the SRC image, suggesting that it is the increased interest in the image depicting an inanimate agent, combined with an online preference for anticipating SRCs, that led to increased looks to this image as the RC unfolded. The preference for fixations to the SRC image with an inanimate NP1 in the second experiment is, however, smaller than that found in Experiment 1 for the children and it is not there for the adults. This may be the result of a greater surprisal effect for the pairing of an inanimate agent and an animate patient, than for an inanimate agent and an inanimate patient.

While the potential semantic interest of some of the visual depictions may make it difficult to interpret the eye-movement data in our experiments (i.e., the depiction of an inanimate object acting on something or someone else may be of more interest than an

animate character acting on another character or object), we suggest that there are two broad conclusions that we can draw from our data. First, throughout both experiments, we found no evidence of any increase in anticipation for an ORC when a lexically inanimate head-noun was used. Prima facie this finding suggests a strong syntax-first bias to treat the NP preceding a relativizer as the head of a SRC where comprehension is primarily guided—both for 4- to 6-year-olds and adults—by a structural cue. At the same time, we argue that in the context of a visual world paradigm task, perceptual animacy, rather than lexical animacy, affected participants' initial looking behavior, if not their ultimate picture selection. In the context of the task, participants had to construct a mental model of the situation including both linguistic information—lexical animacy (the experimental sentences)—and visual information—perceptual animacy (the experimental pictures). The fact that lexically inanimate nouns elicited a significant proportion of looks to the picture matching an SRC indicates a role for perceptual animacy over lexical animacy, thus confirming the importance of non-linguistic factors in language comprehension, and it speaks to the importance of context broadly construed. Although language is more than a running commentary on referents and events we can see and describe, as soon as we incorporate a non-linguistic/visual dimension to language use we need to factor this dimension into the discourse model. Work on speakers' audience design has addressed this very issue in determining the relative weight of linguistic and non-linguistic variables in adults' choices of referential expressions (Arnold & Griffin, 2007; Fukumura, 2015; Fukumura, Van Gompel, Harley, & Pickering, 2011; Fukumura, Van Gompel, & Pickering, 2010). Recent work on the production of referential expressions in monolingual (Serratrice, 2013) and bilingual children (Serratrice & De Cat, 2019) has shown that 5-year-old children are affected by non-linguistic perceptual information, for example, number of referents, animacy of referents, visual access to the referent by their interlocutor, when choosing pronouns versus full lexical NPs to identify an animate entity in a referential communication task. In the current study, we have shown that in comprehension, too, non-linguistic information in the form of the perceptual animacy of a lexically inanimate referent affects children as well as adults when interpreting ORCs. Children, unlike adults, are more susceptible to perceptual surprisal effects as they process incoming ORCs, and they are more likely to entertain perceptually plausible interpretations (SRC) that are inconsistent with the syntactic information encoded in the word order of the RCs they are parsing. Adults, who have more entrenched syntactic knowledge than 5-year olds, can overcome this perceptual bias faster than children (Experiment 1) or disregard it completely (Experiment 2).

The second important finding relates to the discrepancy between offline and online behavior—particularly in the children's case. The implicit online parsing choices that children made while listening to RCs did not necessarily predict their explicit comprehension of these clauses as measured by comprehension questions. Specifically, in Experiment 1, we found that children looked less at the ORC target image during the RC in the inanimate condition than the animate condition, yet they performed better in the former condition. From our behavioral measures alone, we would not be able to identify the strong anticipation for SRCs as the RCs unfolded in the inanimate condition, and we may

even assume less of an online preference for SRCs in this condition. Our online and offline results in combination suggest that the greater accuracy found for inanimate ORCs compared to animate ORCs is not due to differences in the initial syntactic choices made by the children as they heard the RCs, but rather to later stages of the interpretation process. The eye-tracking data have also provided an insight into adults' time course of RC interpretation. Although the adults were at ceiling in all conditions in the accuracy task, they too had an initial bias to fixate more to the incorrect picture in the inanimate–animate ORC sentences, albeit for a shorter period than the children. These differences highlight the importance of using online measures, such as eye tracking, to investigate the processing of complex sentences as they unfold.

8.3. Individual differences

We initially hypothesized that VSM and VWM, inhibitory control, and syntactic comprehension skills would be predictive of children's task performance in terms of accuracy and RT. Our analyses revealed that the only significant predictor was the TROG score for the accuracy data in Experiment 1. An inspection of the distribution of the scores revealed that the standard TROG scores in Experiment 1 were the only ones that were clearly normally distributed and where therefore we had a good spread of syntactic abilities in our sample. The scores for the TROG in Experiment 2, and the scores for the flanker task and the digit recall tasks in both experiments showed a positive or a negative skew revealing a much lower range of systematic variation. We believe this lack of systematic variation is the likely reason for the lack of a correlation with the accuracy scores. This brings us to the much wider issue of the reliability paradox in the study of individual differences (Hedge, Powell, & Sumner, 2018). Low between-subjects variability, which is a desirable feature of an experimental task, causes low reliability for individual differences and therefore the argument is that these experimental executive function tasks are ill-suited for the purpose of studying individual differences. In experimental research, a reliable effect must be replicable, it must be observed in most participants, and it produces consistent effect sizes. In contrast, a task that is suitable to study individual differences must reliably discriminate between participants and in essence ensure that the measure consistently ranks individuals. This tension between the correlational approach that examines differences between individuals in a population (within-subjects variance), and the experimental approach that aims at finding out the typical, average response to a manipulation (within-subjects variance) is methodologically problematic.

Our results are in line with a recent study that also investigated the role of individual differences in the comprehension of complex subordinate clauses (De Ruiter, Theakston, Brandt, & Lieven, 2018). De Ruiter et al. (2018) did not find evidence for a significant role of individual differences in memory, executive function, and general language ability in complex sentence comprehension. This lack of an effect of individual variation on comprehension accuracy leaves open the question as to the role of VWM and inhibition in language processing. The relationship between individual differences and language processing is far from straightforward, and it is not yet entirely clear to what extent

standardized measures are an age-appropriate tool to tap into complex language processes (see Kidd, 2013 for a review on the role of WM in acquisition).

8.4. Conclusion

The offline findings of Experiment 1 show that although children were indeed facilitated by the lexical animacy of the head noun in correctly interpreting ORCs with an inanimate head noun in the sentence–picture matching task, this cue did not facilitate online processing. On the contrary, we argue that the perceptual animacy of the head NP1—coupled with the lack of semantic reversibility of the RC—conspired to make even the inanimate ORCs potential candidates for an SRC interpretation, at least in the initial stages of parsing as shown by the eye-movement data. The plausibility of this interpretation is corroborated by similar findings in the adults' online data, although the more competent speakers stop looking at the incorrect picture earlier than children do.

Our second experiment provided evidence that this seemingly increased anticipation for SRCs was due to increased interest in the visual depiction of an inanimate entity acting on another entity, thus a surprisal effect rather than an effect due to the absence of similarity-based interference, and one that is confined to the children's data, as the adults did not display this behavior in Experiment 2.

The use of online and offline measures together in these experiments, combined with the visual dimension of the tasks, has refined our understanding of the interplay between RC type and animacy in the more challenging ORC structure. Specifically, our offline measures show that the lexical animacy of a head noun can influence the interpretation of an ORC, but our online measures revealed the extent to which perceptual animacy can influence the real-time processing of RCs.

Note

1. A maximal random-effects structure failed to converge for this analysis. Following the advice of Barr et al. (2013), we carried out two versions of the model, with different random structures. For relative clause effects, we used a version of the model with random slopes of RC-type (Model 1) and for animacy effects we used a model with random slopes of animacy condition (Model 2).

References

Adani, F., Stegenwallner-Schütz, M., & Niesel, T. (2017). The peaceful co-existence of frequency and structural intervention effects in relative clause comprehension. *Frontiers in Psychology*, 8, 1590.

Adani, F., van der Lely, H. K. J., Forgiarini, M., & Guasti, M. T. (2010). Grammatical feature dissimilarities make relative clauses easier: A comprehension study with Italian children. *Lingua*, 120(9), 2418–2466.

Adani, F., Forgiarini, M., Guasti, M. T., & Van Der Lely, H. K. (2014). Number dissimilarities facilitate the comprehension of relative clauses in children with (Grammatical) Specific Language Impairment. *Journal of Child Language*, 41, 811–841.

Arnold, J. E., & Griffin, Z. (2007). The effect of additional characters on choice of referring expression: Everyone competes. *Journal of Memory and Language*, 56, 521–536.

Arnon, I. (2010). *Starting big: The role of multi-word phrases in language learning and use* (Unpublished PhD dissertation). Stanford University.

Arosio, F., Adani, F., & Guasti, M. T. (2009). Grammatical features in the comprehension of Italian relative clauses by children. In J. M. Brucart, A. Gavarrò, & J. Solà (Eds.), *Merging features: Computation, interpretation and acquisition* (pp. 138–155). Oxford, UK: Oxford University Press.

Arosio, F., Guasti, M. T., & Stucchi, N. A. (2011). Disrupting information and memory resources in children's processing of Italian relative clauses. *Journal of Psycholinguistic Research*, 40(2), 137–154.

Arosio, F., Yatsushiro, K., Forgiarini, M., & Guasti, M. T. (2012). Morphological information and memory resources in children's processing of relative clauses in German. *Language Learning and Development*, 8, 340–364.

Barkley, R. A. (1999). Response inhibition in attention deficit hyperactivity disorder. *Mental Retardation and Developmental Disabilities Research Reviews*, 5, 177–184.

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

Baudiffier, V., Caplan, D., Gaonach, D., & Chesnet, D. (2011). The effect of noun animacy on the processing of unambiguous sentences: Evidence from French relative clauses. *Quarterly Journal of Experimental Psychology*, 64, 1896–1905.

Bentea, A., Durrleman, S., & Rizzi, L. (2016). Refining intervention: The acquisition offeatural relations in object A-bar dependencies. *Lingua*, 169, 21–41.

Betancort, M., Carreira, M., & Sturt, P. (2009). The processing of subject and object relative clauses in Spanish: An eye-tracking study. *Quarterly Journal of Experimental Psychology*, 62(10), 195–1929.

Bishop, D. V. M. (2003). *TROG-2: Test for the reception of grammar 2*. Oxford, UK: Medical Research Council.

Booth, J. R., MacWhinney, B., & Harasaki, Y. (2000). Developmental differences in visual and auditory processing of complex sentences. *Child Development*, 71(4), 981–1003.

Boudewyn, M. A., Blalock, A. R., Long, D. L., & Swaab, T. Y. (2019). Adaptation to animacy violations during listening comprehension. *Cognitive, Affective, & Behavioral Neuroscience*, 19(5), 1247–1258.

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

Brandt, S., Kidd, E., Lieven, E., & Tomasello, M. (2009). The discourse bases of relativization: An investigation of young German and English-speaking children's comprehension of relative clauses. *Cognitive Linguistics*, 20(3), 539–570.

Christianson, K., Hollingworth, A., Halliwell, J. F., & Ferreira, F. (2001). Thematic roles assigned along the garden path linger. *Cognitive Psychology*, 42(4), 368–407.

Clark, H. H. (1965). Some structural properties of simple active and passive sentences. *Journal of Verbal Learning and Verbal Behavior*, 4, 365–370.

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

Dink, J., & Ferguson, B. (2015). *eyetrackingR: An R library library for eye-tracking data analysis*. Available at: http://www.eyetracking-r.com Accessed June 17, 2020.

Dowty, D. (1991). Thematic proto-roles and argument selection. *Language*, *67*(3), 547–https://doi.org/10.2307/415037

Fan, J., Flombaum, J. I., McCandliss, B. D., Thomas, K. M., & Posner, M. I. (2003). Cognitive and brain consequences of conflict. *NeuroImage*, *18*, 42–57.

Frazier, L. (1987). Sentence processing: A tutorial review. In M. Coltheart (Ed.), *Attention and performance* (pp. 559–586). Hillsdale, NJ: Lawrence Erlbaum Associates.

Frazier, L., & Flores D'Arcais, G. B. (1989). Filler driven parsing: A study of gap filling in Dutch. *Journal of Memory and Language*, *28*, 331–344.

Friedman, N. P., & Miyake, A. (2004). The relations among inhibition and interference control functions: A latent variable analysis. *Journal of Experimental Psychology: General*, *133*, 101–135.

Friedmann, N., Belletti, A., & Rizzi, L. (2009). Relativized relatives. Types of intervention in the acquisition of A-bar dependencies. *Lingua*, *119*, 67–88.

Fukumura, K. (2015). Interface of linguistic and visual information during audience design. *Cognitive Science*, *39*(6), 1419–1433.

Fukumura, K., van Gompel, R. P. G., Harley, T., & Pickering, M. J. (2011). How does similarity-based interference affect the choice of referring expression? *Journal of Memory and Language*, *65*(3), 331–344.

Fukumura, K., van Gompel, R. P. G., & Pickering, M. J. (2010). The use of visual context during the production of referring expressions. *Quarterly Journal of Experimental Psychology*, *63*(9), 1700–1715.

Gennari, S., & MacDonald, M. C. (2008). Semantic indeterminacy in object relative clauses. *Journal of Memory and Language*, *58*(2), 161–187.

Gibson, E. (1998). Linguistic complexity: Locality of syntactic dependencies. *Cognition*, *68*(1), 1–76.

Gibson, E. (2000). The dependency locality theory: A distance-based theory of linguistic complexity. In Y. Miyashita, A. Marantz, & W. O'Neil (Eds.), *Image, language, brain* (pp. 95–126). Cambridge, MA: MIT Press.

Gordon, P. C., Hendrick, R., Johnson, M., & Lee, Y. (2006). Similarity-based interference during language comprehension: Evidence from eye tracking during reading. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *32*, 1304–1321.

Gordon, P. C., & Lowder, M. W. (2012). Complex sentence processing: A review of theoretical perspectives on the comprehension of relative clauses. *Linguistics and Language Compass*, *6*(7), 403–415.

Grillo, N. (2009). Generalized minimality: Feature impoverishment and comprehension deficits in agrammatism. *Lingua*, *119*, 1426–1443.

Grodner, D., & Gibson, E. (2005). Some consequences of the serial nature of linguistic input. *Cognitive Science*, *29*(2), 261–290.

Haendler, Y., Kliegl, R., & Adani, F. (2015). Discourse accessibility constraints in children's processing of object relative clauses. *Frontiers in Psychology*, *6*, 860.

Harnishfeger, K. K. (1995). The development of cognitive inhibition: Theories, definitions, and research evidence. In F. N. Dempster & C. J. Brainerd (Eds.), *Interference and inhibition in cognition* (pp. 175–204). San Diego, CA: Academic Press.

Hedge, C., Powell, G., & Sumner, P. (2018). The reliability paradox: Why robust cognitive tasks do not produce reliable individual differences. *Behavior Research Methods*, *50*(3), 1166–1186.

Hsiao, F., & Gibson, E. (2003). Processing relative clauses in Chinese. *Cognition*, *90*(1), 3–27.

Humphreys, G. F., Mirkovic, J., & Gennari, S. P. (2016). Similarity-based competition in relative clause production and comprehension. *Journal of Memory and Language*, *89*, 200–221.

Jäger, L. A., Engelmann, F., & Vasishth, S. (2017). Similarity-based interference in sentence comprehension: Literature review and Bayesian meta-analysis. *Journal of Memory and Language*, *94*, 316–339.

Järvikivi, J., van Gompel, P. G., Hyönä, J., & Bertram, R. (2005). Ambiguous pronoun resolution: Contrasting the first-mention and subject-preference accounts. *Psychological Science*, 16, 260–264.

Kidd, E. (2013). The role of verbal working memory in children's sentence comprehension: A critical review. *Topics in Language Disorders*, 33(3), 208–223.

Kidd, E., Brandt, S., Lieven, E., & Tomasello, M. (2007). Object relatives made easy: A cross-linguistic comparison of the constraints influencing young children's processing of relative clauses. *Language and Cognitive Processes*, 22(6), 860–897.

Kramer, J. H., Mungas, D., Possin, K. L., Rankin, K. P., Boxer, A. L., Rosen, H. J., ... Widmeyer, M. (2014). NIH EXAMINER: Conceptualization and development of an executive function battery. *Journal of the International Neuropsychological Society*, 20(1), 11–19.

Lewis, S., & Phillips, C. (2015). Aligning grammatical theories and language processing models. *Journal of Psycholinguistic Research*, 44(1), 27–46.

Mak, W. M. (2001). *Processing relative clauses: Effects of pragmatic, semantic, and syntactic variables* (Unpublished doctoral dissertation). University of Nijmegen.

Mak, W. M., Vonk, W., & Schriefers, H. (2002). The influence of animacy on relative clause processing. *Journal of Memory & Language*, 47, 50–68.

Mak, W. M., Vonk, W., & Schriefers, H. (2006). Processing relative clauses in Dutch: When rocks crush hikers. *Journal of Memory & Language*, 54, 466–490.

Maris, E., & Oostenveld, R. (2007). Nonparametric statistical testing of EEG- and MEG-data. *Journal of Neuroscience Methods*, 164(1), 177–190.

Nakamura, C., & Arai, M. (2016). Persistence of initial misanalysis with no referential ambiguity. *Cognitive Science*, 40(4), 909–940.

Nelson, D., & Vihman, V. A. (2018). Bringing the toys to life: Animacy, reference, and anthropomorphism in Toy Story. *International Journal of Language and Culture*, 5(2), 203–223.

Nieuwland, M. S., & Van Berkum, J. J. A. (2006). When peanuts fall in love: N400 evidence for the power of discourse. *Journal of Cognitive Neuroscience*, 18(7), 1098–1111.

Novick, J. M., Hussey, E., Teubner-Rhodes, S., Harbison, J. I., & Bunting, M. F. (2014). Clearing the garden-path: Improving sentence processing through cognitive control training. *Language, Cognition and Neuroscience*, 29(2), 186–217.

Novick, J. M., Trueswell, J. C., & Thompson-Schill, S. L. (2005). Cognitive control and parsing: Reexamining the role of Broca's area in sentence comprehension. *Cognitive, Affective & Behavioral Neuroscience*, 5, 263–281.

O'Grady, W. (2011). Relative clauses: Processing and acquisition. In E. Kidd (Ed.), *The acquisition of relative clauses: Processing, typology and function* (pp. 13–38). Amsterdam, The Netherlands: John Benjamin.

Peirce, J. W. (2007). PsychoPy – Psychophysics software in Python. *Journal of Neuroscience Methods*, 162 (1–2), 8–13.

R Core Team. (2016). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing.

Reali, F. (2014). Frequency affects object relative clause processing: Some evidence in favor of usage-based accounts. *Language Learning*, 64(3), 685–714.

Reali, F., & Christiansen, M. H. (2007). Processing of relative clauses is made easier by frequency of occurrence. *Journal of Memory and Language*, 57(1), 1–23.

Scholl, B. J., & Tremoulet, P. D. (2000). Perceptual causality and animacy. *Trends in Cognitive Sciences*, 4, 299–308.

Serratrice, L. (2013). The role of number of referents and animacy in children's use of pronouns. *Journal of Pragmatics*, 56, 31–42.

Serratrice, L., & De Cat, C. (2019). Individual differences in the production of referential expressions: The effect of language proficiency, language exposure and executive function in bilingual and monolingual children. *Bilingualism: Language and Cognition*. https://doi.org/10.1017/S1366728918000962

Silverstein, M. (1976). Hierarchy of features and ergativity. In R. M. W. Dixon (Ed.), *Grammatical categories in Australian languages* (pp. 112–171). Canberra, Australia: Australian Institute of Aboriginal Studies.

Tanenhaus, M. K., Spivey-Knowlton, M. J., Eberhard, K. M., & Sedivy, J. C. (1995). Integration of visual and linguistic information in spoken language comprehension. *Science*, *268*(5217), 1632–1634.

Thothathiri, M., Asaro, C. T., Hsu, N. S., & Novick, J. M. (2018). Who did what? A causal role for cognitive control in thematic role assignment during sentence comprehension. *Cognition*, *178*, 162–177.

Traxler, M. J., Morris, R. K., & Sealy, R. E. (2002). Processing subject and object relative clauses: Evidence from eye-movements. *Journal of Memory & Language*, *47*, 69–90.

Trueswell, J. (2008). Using eye movements as a developmental measure within psycholinguistics. In I. A. Sekerina, E. M. Fernández, & H. Clahsen (Eds.), *Developmental psycholinguistics: On-line methods in children's language processing* (pp. 73–96). New York, NY: John Benjamins.

Van Gompel, R. P., Pickering, M. J., Pearson, J., & Jacob, G. (2006). The activation of inappropriate analyses in garden-path sentences: Evidence from structural priming. *Journal of Memory and Language*, *55*(3), 335–362.

Vogels, J., Krahmer, E., & Maes, A. (2013). When a stone tries to climb up a slope: The interplay between lexical and perceptual animacy in referential choices. *Frontiers in Psychology*, *4*, 154.

Warren, T., & Gibson, E. (2002). The influence of referential processing on sentence complexity. *Cognition*, *85*, 79–112.

Wechsler, D. (2014). *Wechsler Intelligence Scale for Children* (5th ed). Bloomington, MN: Pearson.

Wells, J. B., Christiansen, M. H., Race, D. S., Acheson, D. J., & MacDonald, M. C. (2009). Experience and sentence processing: Statistical learning and relative clause comprehension. *Cognitive Psychology*, *58*, 250–327.

Wonnacott, E., Joseph, H. S., Adelman, J. S., & Nation, K. (2016). Is children's reading "good enough"? Links between online processing and comprehension as children read syntactically ambiguous sentences. *The Quarterly Journal of Experimental Psychology*, *69*(5), 855–879.

Woodard, K., Pozzan, L., & Trueswell, J. C. (2016). Taking your own path: Individual differences in executive function and language processing skills in child learners. *Journal of Experimental Child Psychology*, *141*, 187–209.

Supporting Information

Additional supporting information may be found online in the Supporting Information section at the end of the article:

Supplementary Material 1. List of the 24 experiment items in each condition and the 12 filler items and six practise items used in Experiment 1.

Supplementary Material 2. List of the linear mixed-effects models (LMMs) and generalized linear mixed-effects models (GLMMs) used in Experiments 1 and 2.

Supplementary Material 3. List of the 24 experimental items in each condition and the 12 filler items and six practise items used in Experiment 2.