



Contents lists available at ScienceDirect

Journal of Experimental Child Psychology

journal homepage: www.elsevier.com/locate/jecp



Animacy-induced conflict in sentence production and comprehension from late childhood to adolescence

Shi Hui Wu^{*}, Lisa-Marie Henderson, Silvia P. Gennari

Department of Psychology, University of York, York YO10 5DD, UK



ARTICLE INFO

Article history:

Received 22 March 2021

Revised 7 December 2021

Keywords:

Relative clauses
Sentence comprehension
Sentence production
Language development
Working memory
Vocabulary
Text exposure
Backward digit span

ABSTRACT

Some animacy configurations elicit parallel semantic interference in adult production and comprehension; for example, phrases with similar animate nouns like *the man that the girl is hugging* are more difficult than phrases like *the doll that the girl is hugging*. Yet little is known about how this interference manifests in development, particularly, beyond early childhood. Because frontal brain maturation and cognitive control improvements are known to occur across late childhood and adolescence, we investigated (a) how animacy-induced difficulty in production and comprehension vary with age throughout this period and (b) whether control processes reflected in the backward digit span (BDS) test uniquely explained these differences besides other language measures. In separate tasks, participants (8- to 15-year-old children; $N = 91$) heard auditory descriptions of depicted characters, produced characters' descriptions, and completed BDS, vocabulary, and reading experience tests. Results indicated that, as in adults, animacy modulated performance in production and comprehension across all ages. The animacy modulation interacted with age in production but not in comprehension, suggesting age-related animacy differences in production but relatively stable differences in comprehension despite processing speed improvements. Importantly, these age-related production differences were also modulated by the BDS scores; only participants with higher BDS scores displayed age-related animacy differences. Together, these results indicate that comprehension and production develop at different rates and that the development of BDS performance interacts with age-dependent changes in sentence planning from late childhood to adolescence.

^{*} Corresponding author.

E-mail address: shw502@york.ac.uk (S.H. Wu).

More generally, the study highlights tasks' disparities to be explained by cognitive and developmental models of language.

© 2021 Elsevier Inc. All rights reserved.

Introduction

Sentence comprehension and production involve a complex interplay of multiple cognitive and linguistic processes. In comprehension, words must be temporarily held in mind as a sentence unfolds to integrate them with upcoming ones. In production, speakers must maintain some representation of the intended message, find the right words, and plan and execute an output with a grammatical linear order. Because both these processes involve temporary maintenance of information, theories of working memory (WM) have played an important role in the study of language processing. In both developmental and adult research, WM approaches to language processing have generally focused on storage capacity (Just & Carpenter, 1992; Montgomery, 1995) or on WM mechanisms associated with executive or control processes (Lewis, Vasishth, & Van Dyke, 2006; Montgomery, Evans, & Gillam, 2009). For example, interference between similar items or retrieval cues in WM are thought to be responsible for comprehension difficulty.

Other approaches, in contrast, have examined how language experience and use and long-term knowledge may explain sentence processing (Kidd, Brandt, Lieven, & Tomasello, 2007; MacDonald, Pearlmuter, & Seidenberg, 1994; Matthews, Lieven, Theakston, & Tomasello, 2005). For example, words and phrase structures that occur more frequently in someone's experience are typically easier to acquire and process than rarer ones. During processing, probabilistic expectations of frequent patterns (e.g., that initial animate nouns are typically agents) may conflict with upcoming information leading to competition between alternative structures or interpretations. These contrasting theoretical approaches have developed in largely independent research strands and mainly differ in theoretical assumptions. At least in behavioral sentence processing research, these accounts can accommodate similar behavioral evidence. Nevertheless, some recent proposals have argued that the verbal WM construct emerges from other verbal processes ultimately grounded in verbal knowledge, prior experience, and learning (Jones et al., 2020; Jones & Macken, 2015; Kidd, 2013; MacDonald, 2016; MacDonald & Christiansen, 2002; Postle, 2006).

Both these approaches have highlighted the potential role of executive or control processes in adult production and comprehension. Control processes are typically associated with frontal lobe functions and involve the regulation of thought and behavior according to task demands or agents' goals, for example, dealing with conflicting cues for successful performance as in the Stroop task (Botvinick, Braver, Barch, Carter, & Cohen, 2001; Miyake & Friedman, 2012). Adult behavioral and brain research in sentence comprehension suggest that ambiguous sentences recruit control processes to manage and/or monitor conflicting cues for successful task performance (Hsu & Novick, 2016; Novick, Trueswell, & Thompson-Schill, 2005, 2010). For example, temporary ambiguous phrases such as *put the frog on the napkin on the box* are argued to share control processes with the Stroop task. Similarly, sentence production may involve monitoring or motor conflict resolution as it recruits a larger brain network along dorsal pathways, including conflict-related motor brain regions (Alario, Chainay, Lehericy, & Cohen, 2006; Hickok & Poeppel, 2007; Humphreys & Gennari, 2014; Nozari & Novick, 2017).

Developmentally, control processes reflected in various WM tasks and complex sentence comprehension are thought to emerge during early childhood (~4–6 years of age) but continue to undergo significant changes into adolescence (Gathercole, Pickering, Ambridge, & Wearing, 2004; Sesma, Mahone, Levine, Eason, & Cutting, 2009; Skeide & Friederici, 2016). Word production tasks during this period also show significant age-related changes in functional brain activity (Brown et al., 2005). More generally, independent of verbal tasks, a large body of neurodevelopmental research has reported considerable changes into adolescence in control tasks involving inhibition, WM load, and task switching

(Davidson, Amso, Anderson, & Diamond, 2006; Huizinga, Dolan, & van der Molen, 2006). Across the same developmental period, the prefrontal cortex undergoes significant functional and structural changes that are associated with behavioral performance (Casey, Tottenham, Liston, & Durston, 2005; Gogtay et al., 2004; Shaw et al., 2008; Skeide & Friederici, 2016). Prefrontal immaturity, for example, correlates with poorer performance on WM tasks (Bunge & Wright, 2007; Crone & Steinbeis, 2017) and tasks involving inhibition such as Go/No-Go and Stroop (Blasey et al., 2002; Constantinidis & Luna, 2019; Rubia et al., 2006; Tamm, Menon, & Reiss, 2002). On the whole, this research indicates that cognitive control processes mature and develop late into adolescence.

To date, however, most behavioral developmental studies, particularly those concurrently examining comprehension and production, have focused on early language acquisition stages with preschool and young children (Contemori & Garraffa, 2010; Håkansson & Hansson, 2000; Jensen De Lopez et al., 2014), or have targeted comprehension or repetition, rather than message formulation. Even in adult sentence processing, it is currently unclear which control processes might operate in each sentence production and comprehension and whether these might be shared and/or recruit domain-general control processes not specific to language processing (Gratton, Cooper, Fabiani, Carter, & Karayanidis, 2018; Miyake & Friedman, 2012). Nevertheless, these issues are important to delineate the architecture of the language system and how it might develop over time. Indeed, whereas vocabulary and grammatical knowledge are typically considered shared across language tasks (Chang, Dell, & Bock, 2006; Pickering & Ferreira, 2008), much less is known about how control processes vary across tasks.

With these issues in mind, we aimed to investigate how sentence comprehension and production performance varied with age from late childhood to adolescence and how these changes relate to measures of verbal knowledge and experience along with a task recruiting control processes. We used complex referential phrases containing relative clauses (RCs) that have been previously shown to elicit processing difficulty as a function of noun animacy. Below, we review previous work on these structures.

RC comprehension in children and adults

Active object relative clauses (ORCs), like that in item (a) below, have been extensively examined in multiple adults and children's comprehension studies in English, particularly in comparison with active subject relative clauses (SRCs), like that in item (b) below:

- (a) *Active ORC, animate target*: The man that the girl is hugging.
- (b) *Active SRC*: The man that is hugging the girl.
- (c) *Passive SRC*: The man that is being hugged by the girl.
- (d) *Active ORC, inanimate target*: The teddy bear that the girl is hugging.

Compared with SRCs, ORCs are more difficult to process because comprehenders are unable to interpret who did what to whom in a linear fashion as in typical subject–verb–object structures. Comprehenders need to overcome the tendency to interpret the first noun as the agent of the upcoming verb and understand it as the patient (e.g., *hug*). Thus, although SRCs elicit little difficulty in processing, children under 6 years of age make comprehension errors for ORCs and tend to master them later in development (Arnon, 2010; Diessel & Tomasello, 2000; Kidd & Bavin, 2002; Macdonald, Brandt, Theakston, Lieven, & Serratrice, 2020). Depending on the approach, comprehension difficulty in RCs has been argued to stem from either WM mechanisms or the relative infrequency of ORCs compared with SRCs (Gibson, 1998; Kidd et al., 2007; Macdonald et al., 2020; Real & Christiansen, 2007; Traxler, Morris, & Seely, 2002).

However, the much-investigated contrast between SRCs and ORCs is unable to distinguish the role of syntax and semantics in processing because these structures differ not only in meaning (who is acting on whom) but also in syntactic structure; thus, the difficulty associated with ORCs may potentially stem from the more infrequent word order in ORCs, the unexpected patient interpretation of the first head noun (which should be revised as the structure unfolds), or both. More recent adult and child studies have instead focused on minimally different contrasts such as item (a) versus (c) and item

(a) versus (d) in the list above, where the meaning or syntactic structure stays constant and only one factor changes across conditions (Gennari & MacDonald, 2008, 2009; Gennari, Mirković, & MacDonald, 2012; Macdonald et al., 2020; Montag & MacDonald, 2015). Note that although passives are generally infrequent in main sentences, they are very frequent in RCs, as demonstrated in both adult and child studies, corpus analyses, and production studies (Fox & Thompson, 1990; Montag & MacDonald, 2015; Roland, Dick, & Elman, 2007).

Importantly, these studies have emphasized that comprehension difficulty in ORCs depends on the animacy of the nouns involved. For example, ORCs with inanimate-target referents like item (d) above are typically interpreted more accurately and are understood faster than those with animate targets like item (a) above by both children and adults (Arosio, Guasti, & Stucchi, 2011; Corrêa, 1995; Mak, Vonk, & Schriefers, 2002; Traxler et al., 2002; Traxler, Williams, Blozis, & Morris, 2005). Interestingly, inanimate-target ORCs such as that in item (d) above are as difficult to understand as SRCs, suggesting that the unusual word order in ORCs is not the only source of difficulty in these structures (Mak et al., 2002; Traxler et al., 2002). Instead, in both probabilistic and WM approaches, meaning and interpretation play a role in these structures.

Consequently, it has been proposed that inanimate-target ORCs are easier to process than animate-target ones because inanimate targets are more frequent in ORCs, more distinguishable from animate nouns in WM, and strongly associated with patient-like roles. Thus, inanimate nouns are less likely to be the agent of the action verb (e.g., inanimate objects do not typically act on people) and thus are immediately interpreted as patients (Trueswell, Tanenhaus, & Garnsey, 1994). In contrast, the two animate nouns in item (a) above elicit competition or interference when establishing who is acting on who before and after the verb is encountered (Gennari & MacDonald, 2008; Humphreys, Mirković, & Gennari, 2016; Van Dyke & Lewis, 2003). Evidence for semantic competition comes from studies showing that the degree of conceptual similarity between the nouns modulates comprehension (and production) difficulty (Gennari, Mirković, & MacDonald, 2012; Humphreys, Mirković, & Gennari, 2016; Konopka & Kuchinsky, 2015). For example, phrases like *the boy that the girl is hugging* are more difficult than phrases like *the dog that the girl is hugging*, and these in turn are more difficult than inanimate-target ORCs. These findings suggest that interference or competition in English ORC comprehension increases when nouns are more difficult to distinguish from one another to determine who is acting on whom. Thus, even when these processing differences may ultimately be correlated with distributional characteristics in the language, animate-target ORCs are more problematic to process than inanimate-target ones, particularly if the two animate nouns are conceptually similar and the event described is reversible.

RC production in children and adults

Compared with comprehension studies, less research has been devoted to RC production. But it is well documented in both corpus and experimental studies that children and adults show a preference for certain types of RCs in a way that matches their comprehension patterns; SRCs are preferred over ORCs, and inanimate-target ORCs are more frequent and mastered earlier than animate-target ORCs (Kidd, Brandt, Lieven, & Tomasello, 2007; McDaniel, McKee, & Bernstein, 1998; Real & Christiansen, 2007). For example, in a picture description paradigm in English, both children (8- and 12-year-olds) and adults tended to avoid active ORCs to describe animate-target events (e.g., *the man that the girl is hugging*), and instead overwhelmingly produce passives (e.g., *the man that is being hugged by the girl*) around 90% of the time. In contrast, active ORCs are produced more frequently for inanimate targets (e.g., *the teddy bear that the girl is hugging*), around 75% of time by children and 50% of the time by adults (Gennari et al., 2012; Hsiao & MacDonald, 2016; Montag & MacDonald, 2015). The preference for passive RCs in describing animate-target events reflects the distributional statistics in adult and child English corpora (Fox & Thompson, 1990; Gennari & MacDonald, 2009; Montag & MacDonald, 2015; Roland, Dick, & Elman, 2007).

These corpus preferences stem from speakers' tendency to make the production process easier and more fluent. For example, speakers opt for passive structures in animate-target descriptions because they experience interference or competition between highly similar animate entities, which makes uttering the two nouns in close proximity difficult, as would be required in an active structure

(Allum & Wheeldon, 2007, 2009; Smith & Wheeldon, 2004). The selection of a noun to describe an animate-target ORC in the presence of alternative picture characters to be named is known to elicit interference (Arnold, 2010; Fukumura, van Gompel, Harley, & Pickering, 2011). Thus, in planning to talk about a man who is being hugged by a girl in a picture, speakers alleviate the competition between first naming the man or the girl by adopting a passive structure; they select the task-required target noun *the man* and inhibit the alternative noun *the girl*. Thus, this alternative noun is demoted to the end of the structure (via a by-phrase) or is omitted altogether, for example, *the man being hugged* (Gennari et al., 2012). Interestingly, as compared with adults, children's choice of passive structure is less prevalent. As suggested by Montag and MacDonald (2015), young children are less familiar with passives in RC structures due to less reading experience and thus are less likely to entertain a passive structure as a viable and efficient alternative in RC production, using an active ORC instead.

Despite this body of research, previous studies investigating RC comprehension or production have examined only preschool or school-age children. Thus, less is known about developmental patterns beyond childhood. Moreover, most comprehension studies have mainly focused on comparing ORCs with SRCs as in item (a) versus (c) above. In contrast, the current study focused on comparing animate- and inanimate-target descriptions to specifically examine the animacy-induced conflict while keeping the syntactic structure constant. Importantly, this study concurrently investigated production and comprehension and thus had the potential to reveal contrasting skills and developmental patterns across language tasks, both of which are critical to master communication.

Control processes in RC comprehension and production

In WM accounts, various tasks have been used to predict RC children's comprehension. These include backward and forward digit spans, reading and listening spans, and nonword repetition, among others. Simple span tasks are considered measures of short-term memory capacity, but many WM models suggest that control processes such as competitive cuing or interference and response suppression play a role in serial recall (Cowan, Sauls, Elliott, & Moreno, 2002; Hurlstone, Hitch, & Baddeley, 2014). In the backward digit span (BDS) task used here, these processes are accompanied by item manipulation (repeating a sequence backward). Like other complex span tasks, BDS is considered to involve the central executive—a domain-general attentional mechanism (Gathercole et al., 2004). Nevertheless, no task is a pure reflection of a given theoretical construct or executive function given that it necessarily includes variance related to uncontrolled factors, for example, stimulus processing and speed of articulation (Miyake & Friedman, 2012; Miyake et al., 2000).

A large number of developmental studies have reported positive associations between WM measures and RC comprehension, although not RC planning. Children with poorer WM scores perform less well in accuracy and/or processing times as compared with children with better WM scores (see Table SM1 in the online supplementary material). Notably, however, findings are mixed regarding which particular WM measure most strongly relate to RC comprehension, with different studies using different WM measures. Although some studies have suggested that limited capacity storage plays a role in ORC comprehension (Arosio et al., 2011; Arosio, Yatsushiro, Forgiarini, & Guasti, 2012; Bentea, Durrleman, & Rizzi, 2016; Booth, MacWhinney, & Harasaki, 2000), a majority of studies have used complex span tasks tapping on the central executive (Felser, Clahsen, & Munte, 2003; Finney, Montgomery, Gillam, & Evans, 2014; Montgomery, Evans, Fargo, Schwartz, & Gillam, 2018; Roberts, Marinis, Felser, & Clahsen, 2007; Rusli & Montgomery, 2017). Whereas this evidence suggests that control processes operate on ORC comprehension, it is still unclear which ones these are and how they interact with linguistic knowledge (Rusli & Montgomery, 2017).

The current study

Here, we aimed to illuminate our understanding of language development from late childhood to adolescence in both production and comprehension. Specifically, we examined two main questions: (a) how animacy-induced competition varies with chronological age and (b) how this variation relates to individual measures reflecting language-specific characteristics (vocabulary knowledge and reading

experience) and control processes involving maintenance, manipulation, and motor output in serial recall (BDS). We used picture-based tasks following previous studies (Humphreys et al., 2016; Montag & MacDonald, 2015). Participants either heard descriptions of picture characters with an ORC structure (comprehension task) or produced an RC description of a character (production task). Keeping the animacy of the agent constant, we specifically manipulated the animacy of the RC head noun, namely, the animacy of the character (man or teddy bear) that was the target of the description and the affected participant. From the comprehension task, accuracy and processing time measures were obtained. From the production task, we computed description accuracy and the proportion of passive utterances produced. This is a common measure in production research (Bock, Loebell, & Morey, 1992), which has been previously used in RC production (Gennari et al., 2012; Montag & MacDonald, 2015). Participants also performed standard tests probing vocabulary knowledge, reading experience, and BDS.

Because prior research suggests that animate-target RCs elicit more difficulty than inanimate-target RCs, we specifically examined whether the animacy effect would remain relatively stable or change from late childhood to adolescence. A change in performance would result in an interaction between animacy and age. Montag and MacDonald (2015) indeed reported such an interaction suggesting that 12-year-olds, like adults, produced more passives compared with 8-year-olds, but only for animate targets. Importantly, an interaction between animacy and individual measures would indicate a contrasting modulation in explaining the animate- versus inanimate-target RCs, suggesting a different role of the measure in explaining processing difficulty in each animacy condition. For example, as previously found for adults (Wu, Henderson, & Gennari, *in press*), vocabulary scores may better predict the animate- versus inanimate-target condition, suggesting a stronger role for vocabulary specifically in the more difficult condition. Given that developmental changes from childhood to adolescence are linked to improvements in control processes, as reviewed above, we expect that control processes such as those recruited in the BDS will interact with animacy beyond the influence of language measures in ways that vary across ages.

Method

Participants

A total of 92 participants aged 8;0 to 14;11 (years;months) (57 girls and 35 boys; $M_{\text{age}} = 11.66$ years, $SD = 2.00$, $Mdn = 12.083$) were recruited from the wider community in the city of York, United Kingdom. Age was not normally distributed (Kolmogorov–Smirnov = .097, $p = .031$; see also supplementary material for age distribution histogram). All participants were native English speakers without any known history of language problems or developmental delays. Of this sample, the comprehension performance of 1 participant (age 11;2) was not recorded due to program malfunctioning, and the production performance of 1 participant (age 8;2) was removed due to low accuracy (45%, suggesting that this participant did not fully understand the task instructions). Altogether, 91 participants' comprehension performance and 91 participants' production performance were included for analysis.

Comprehension and production tasks

Materials

A total of 40 gray-scale pictures were selected from a large set used in previous adult studies (Humphreys et al., 2016; Wu et al., *in press*). These comprised 20 experimental pictures and 20 filler pictures (see Tables SM3 and SM4 in supplementary material). Each experimental picture contained two events to be described with the same transitive verb, for example, *hug*, *carry*, or *spray*. The events contained either an animate- or inanimate-target character (animacy manipulation) that was acted on by another character. In Fig. 1, for example, the targets for the right-side picture were either the teddy bear or the man being hugged. In the comprehension task, each picture was also accompanied by a recorded phrase referring to a picture character. This phrase was in an active ORC form and included nouns and verbs that participants in a previous production study had more often used to refer to the

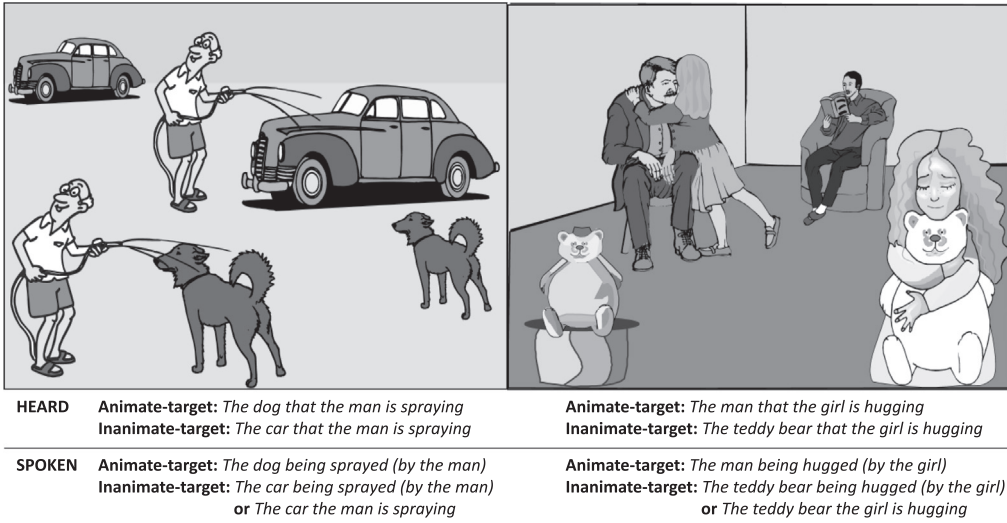


Fig. 1. Examples of picture stimuli used in the production and comprehension tasks. Below the images, the text illustrates the heard stimuli in the comprehension task and the most typical spoken answers provided by adults in the production task.

characters and action depicted, for example, *the teddy bear/man that the girl is hugging* (Humphreys et al., 2016). These RC phrases were recorded by a female native British English speaker in a sound-proof booth using Cool Edit Pro 2 (Synttrillium Corporation, Phoenix, USA). The sound files were normalized to 68 dB SPL (sound pressure level) to minimize intensity differences throughout the recording session.

Design

Following a similar design from adult studies (Humphreys et al., 2016; Wu et al., in press), Animate or inanimate targets were allocated to two different lists (Latin square design) so that a picture was seen in a given task only once, but all participants were exposed to both animate and inanimate target descriptions. Participants completed the comprehension task before the production task but were exposed to different lists in each task. If they had heard the animate-target reference in comprehension for a given picture (e.g., *the man being hugged*), they would then be prompted to describe the inanimate target in the production task and vice versa. This arrangement allowed participants to become familiar with the targeted phrase structures for the production task but did not contain repetitions of the same targets across both production and comprehension. As suggested by previous studies, speakers often use simpler structures such as *the man on the left* to identify the relevant character, and thus the comprehension task prompted participants to focus on the character of the depicted events. These tasks were conducted on E-Prime 2 (Psychology Software Tools, Pittsburgh, PA, USA). In both tasks, participants received 10 trials targeting animate entities and 10 trials targeting inanimate entities plus 20 filler items. Among the animate-target phrases, 5 were reversible (e.g., *the man that the girl is hugging*) and 5 were nonreversible (e.g., *the dog that the girl is hugging*). This factor was examined in previous adult studies (Humphreys et al., 2016; Wu et al., in press), but there was no effect of this variable in these data.

Comprehension task procedure

Participants were instructed to indicate whether the description they heard over the headphones was an accurate description of the character highlighted with a red square in the picture (see Fig. 2). They pressed one of two buttons on a keyboard to indicate their response (i.e., yes or no). The experimental trials in a list required a yes response, but the filler trials elicited no responses, thereby balancing the number of yes/no responses throughout the task. The order of trial presentation was

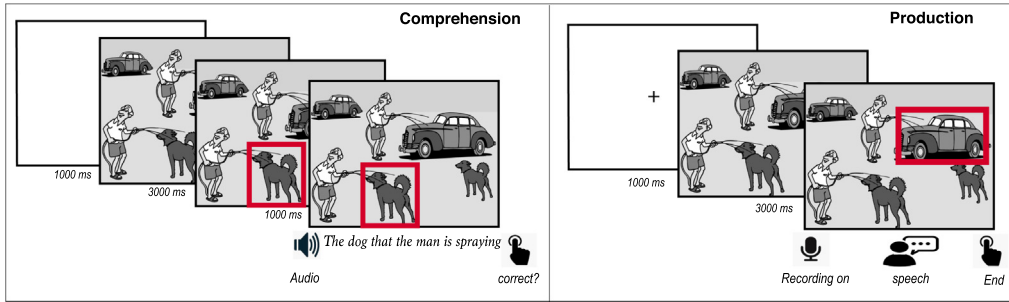


Fig. 2. Schematic trial structure in the production and comprehension tasks. A participant saw the same picture under different conditions in each task.

random. The trial structure is shown in Fig. 2. After a response was recorded, an intervening screen requested a key press to see the next trial. Filler trials in this task contained a variety of phrase structures (e.g., *the boy playing with a ball*, *the woman that is washing the car*, *the nurse being pushed by the violinist*). To prevent strategic performance, a third of the fillers did not match the picture due to the action being wrongly described or the characters being reversed, so that participants could not simply guess the response from the nouns heard and would need to hear the verb to be sure of their responses. Participants' reaction time (RT) for each correct trial was computed by subtracting the audio length of the recorded phrase from the E-Prime RT, which was computed from the audio onset. In a few cases, this subtraction resulted in a negative value if participants responded while the verb phrase was playing and they could recognize the verb from the first syllable. Because the RC stimuli ended with the critical verb and comprehension difficulty in RCs is typically observed at verb positions, where the nouns are integrated with the meaning of the verb, our RT measure captures the intended role-to-argument mappings characteristic of RC comprehension difficulty (determining who is acting on who) (Gennari & MacDonald, 2008; Gibson, 1998; Humphreys et al., 2016; Traxler et al., 2002).

Production task procedure

Participants saw the same 20 pictures as in the comprehension task but in a different animacy condition (i.e., a different character was highlighted). The additional filler items elicited a variety of structures (e.g., *the tree on the playground*, *the dog burying the bone in the sand*). The order of trial presentation ensured that a filler always occurred between experimental items, but experimental items followed a random order. Participants in this task were instructed to verbally describe the highlighted character, and their responses were recorded using a microphone. Practice trials and instructions indicated to participants that they should give descriptions uniquely to identify the character and use the actions being performed rather than location or shape characteristics (e.g., *the man on the left*, *the short girl*). The trial structure was similar to that of comprehension (see Fig. 2). After participants produced a verbal response, the experimenter pressed a key to move onto the next trial. If a participant failed to produce a description with the targeted structure because he or she failed to notice the competitor (e.g., giving a response such as *the dog standing* for Fig. 2), the experimenter pointed out the competitor and asked the participant to try again and make sure the target entity was distinguished from the competitor. The corrected trials were included in the analysis because young children often failed to notice the competitor. After the main experimental tasks, all participants completed standard tests of vocabulary, reading experience, and BDS (described below). We also included two other individual measures that did not yield any significant results (Stop Signal task and a lexical ambiguity task, as reported in Wu et al., in press) (see supplementary material). All tasks were administered individually within a 1-h session.

Individual difference measures

Vocabulary measures and text exposure

We used two measures of vocabulary knowledge: the Vocabulary subtest from the Wechsler Abbreviated Scale of Intelligence–Second Edition (WASI; Wechsler & Hsiao-Pin, 2011), which is a standardized measure of expressive vocabulary (or “depth” of vocabulary knowledge), and the Peabody Picture Vocabulary Test–Fourth Edition (PPVT; Dunn & Dunn, 2007), which is a measure of receptive vocabulary (or vocabulary “breadth”). To account for the influence of reading experience, we used a U. K. version of a Children’s Author Recognition Test (text exposure) developed by Harlaar, Dale, and Plomin (2007). In this test, participants were presented with a list of 21 names of real children’s authors and 21 foils. They were asked to identify which ones were names of real authors. These three measures were strongly correlated with each other (for WASI and PPVT, $r = .65$; for text exposure and WASI or PPTV, $r = .52$ or $r = .51$, respectively; all $ps < .05$, two-tailed), as previously reported (Cunningham & Stanovich, 1991; Stanovich, 1989). To reduce the number of collinear predictors, we used the raw scores from each test to compute z-scores that were then averaged in a composite score (referred to as *vocabulary + text exposure*) by averaging corresponding z-scores of the three measures. Means and standard deviations are shown in Table 1.

Backward digit span

Participants completed the BDS subtest of the Working Memory Test Battery for Children (Pickering & Gathercole, 2001). On each trial, participants were required to recall a sequence of spoken digits in its reverse order (e.g., 6–1–3–9–5–2). Test trials began with three digits and increased by one digit at each level (the levels ranged from three to seven digits, with 6 trials at each level), and the task ended when the participant was unable to recall 4 correct trials at any given level. Task score was calculated based on correct number of trials. Means and standard deviations are shown in Table 1. Correlations between individual measures in our sample are reported in Table 2.

Data coding and analysis

In the comprehension task, only responses to the experimental trials were coded for accuracy and analyzed. The accuracy was generally very high for the comprehension task, so most of the trials were included for further analyses. For comprehension RTs, we first removed extreme outliers suggesting involuntary or mistaken early key presses (e.g., before the verb was heard). Based on the audio files, the lower bound was set at -300 ms (i.e., 300 ms before the recorded RT) to guarantee that initial verb information was heard before a response was made. Second, we removed long responses 5 standard deviations from the mean (i.e., comprehension RTs > 5000 ms for an overall mean of 742 ms and a standard deviation of 723). This exclusion criterion was liberal because we expected large variability across ages and we aimed to retain most data points at this stage. Third, we computed z-scores for

Table 1
Descriptive statistics for cognitive measures.

Measure	N	Range	M	SD
Vocabulary + text exposure z-score	91	–2.19 to 1.76	0	0.84
Backward digit span	91	12 to 36	21.49	5.13

Table 2
Pearson’s correlations between age and individual measures.

	Age	Vocabulary + text
Vocabulary + text exposure	.600**	
Backward digit span	.325**	.443**

** Significant at .01 level (two-tailed).

each individual participant and animacy condition. Values falling above 2.5 standard deviations from the individual's mean by condition (i.e., for animate and inanimate conditions separately) were removed. This resulted in the removal of 22 correct comprehension responses out of a total of 1774 correct trials (see data file in Wu & Gennari, 2021, showing raw scores, z-scores, and used RTs). The production data were first coded for accuracy. If a description was skipped, was inaccurate, or did not include the targeted structure (e.g., *the girl sitting down, the man looking scared, the apple on the pole*), it was excluded from the analyses (216 responses were excluded from a total of 1820 responses). Of the 216 excluded responses, 84 were animate-target descriptions and 132 were inanimate-target ones, and these were roughly equally distributed across ages. The remaining RC responses were then coded as active or passive phrases.

Analyses were conducted using linear mixed-effects models in R Version 3.5.3 with BOBYQA optimizer and maximum iterations set at 100,000 for dichotomous dependent variables (DVs) (Kuznetsova, Brockhoff, & Christensen, 2017). All the initial mixed-effects models included the maximal random-effects structures allowed by the design (by-participant and by-item intercepts, by-participant and by-item random slopes for the animacy condition). In cases of nonconvergence, we simplified the model by removing the random item slopes and report the fullest model that converged (Barr, Levy, Scheepers, & Tily, 2013). Note that our experimental design does not afford statistical mixed models simultaneously evaluating production and comprehension performance because participants comprehended and produced RCs about different characters in different events. So, separate models were used for each task.

Our statistical models proceeded in two steps. For each DV, a first model was built including age, animacy, and the interaction between these two variables to investigate whether there were any age-related changes in the animacy effect in each task. For the animacy conditions, the animate-target trials were coded as 1 and the inanimate-target trials were coded as -1. Thus, a positive coefficient of the condition main effect would represent a higher value in the more difficult condition (i.e., animate condition). Nonsignificant main effects or interactions were removed to identify the simplest model (see Barr et al., 2013).

To examine the unique contribution of different individual difference predictors, we entered scaled predictors (Iacobucci, Schneider, Popovich, & Bakamitsos, 2016) and their interactions with age and animacy in a previously selected order. The combined vocabulary + text exposure score was entered first to account for the role of lexical knowledge and reading experience. BDS scores were entered second to investigate whether control processes explained additional variance after controlling for the effects of vocabulary + text exposure, two well-known predictors of comprehension (Brazz, Tabor, Shankweiler, & Mencl, 2007; Cunningham & Stanovich, 1991; Perfetti, 2007). Again, nonsignificant main effects or interactions were pruned at each step to allow models to be clearly interpretable. For simplicity, our result tables report only this full final model after nonsignificant main effects or interactions were removed. The data and the R code used in the analyses can be found in Wu and Gennari (2021).

Results

Comprehension performance

Comprehension accuracy

In the first model, there was no significant main effect of age, suggesting that younger participants performed as accurately as older participants on this task. There was also no main effect of animacy or vocabulary + text exposure. Only BDS yielded a significant main effect ($p < .01$) (see Table 3); participants with higher BDS scores achieved higher comprehension accuracy (see Fig. 3).

Comprehension RT

As predicted, there was a significant main effect of animacy condition (see Table 4); animate-target phrases took longer to comprehend than inanimate-target ones. There was also a significant main effect of age; younger participants were slower than older participants ($p < .01$). But the Animacy *

Table 3

Coefficient estimates of linear mixed-effects models predicting comprehension accuracy from age, animacy, and cognitive predictors.

	Coefficient	SE	z-score	p value
Intercept	4.02	0.29	14.06	<.01*
Backward digit span	0.82	0.20	4.06	<.01*

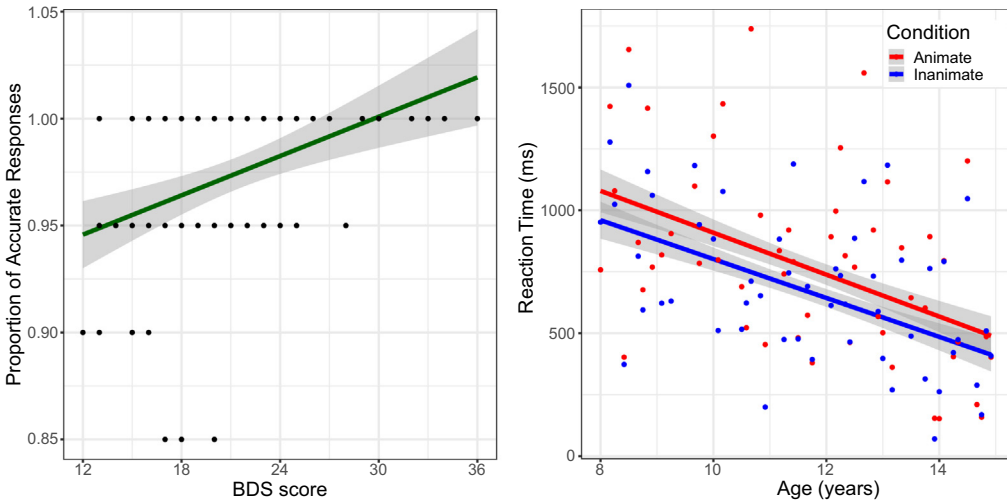


Fig. 3. Comprehension results: Overall proportions of accurate responses as a function of backward digit span (BDS) score (left panel) and mean comprehension reaction times by animacy condition and age (right panel).

Table 4

Coefficient estimates of linear mixed-effects models predicting comprehension reaction times (ms) from age, animacy, and cognitive predictors.

	Coefficient	SE	df	t score	p value
Intercept	1572.58	197.06	95.55	7.98	<.01*
Animacy	97.21	46.22	20.26	2.10	.048*
Age	−77.05	16.41	91.11	−4.69	<.01*

Age interaction was nonsignificant, suggesting that the size of the animacy effect does not vary with age (Table 4).

Because none of the individual predictors significantly explained comprehension RTs beyond age but were initially correlated with age (see Table 2), we explored whether these variables predicted RTs when age was not in the model. We found a significant main effect of vocabulary + text exposure ($p < .01$), which did not interact with the animacy condition; participants with better vocabulary and more text exposure made faster responses overall (see Table SM5 in supplementary material). This suggests that age and vocabulary + text exposure explain some shared portion of the model's variance, but age variability explains RTs best.

In sum, whereas comprehension accuracy was high across ages, we found significant age and animacy effects in comprehension RTs but no interaction between them, suggesting that accuracy in active ORCs does not vary with age. Therefore, this result indicates a relatively stable pattern in dealing with animacy-related difficulty throughout this developmental period. Moreover, despite high accuracy performance, which was similar across animacy conditions, this measure was predicted by

the BDS scores; adolescents with higher scores had more accurate comprehension than those with lower scores but did not interacted with animacy. Finally, none of the individual predictors significantly explained comprehension RTs beyond the animacy and age effects. Nevertheless, age and vocabulary + text exposure appear to play a similar role in predicting performance given that they were strongly correlated with each other; both age and vocabulary were similarly associated with RTs, suggesting an improvement in processing speed as age and vocabulary + text exposure increased.

Production performance

Production accuracy

Unlike comprehension accuracy, there was a marginally significant effect of animacy ($p = .057$) and a significant main effect of vocabulary + text exposure ($p < .01$) beyond age; speakers with better vocabulary and more text exposure provided more accurate responses using the targeted RC structures (see Fig. 4). There was no influence of BDS in this model, which was removed from Table 5.

The effect of animacy on production accuracy was due to more errors occurring in the inanimate-target condition. In these cases, participants sometimes described the wrong target; for example, the animate entity was described rather than the inanimate one (e.g., *the girl hugging the teddy bear* instead of *the teddy bear being hugged by the girl*). There was a trend indicating that this effect was larger in younger participants, as shown in Fig. 4 (left panel), but the interaction with age was not statistically significant. This type of error is also found in adults (Wu et al., in press), and these errors likely stem from a tendency to use the more frequent subject RC and to focus on the human character

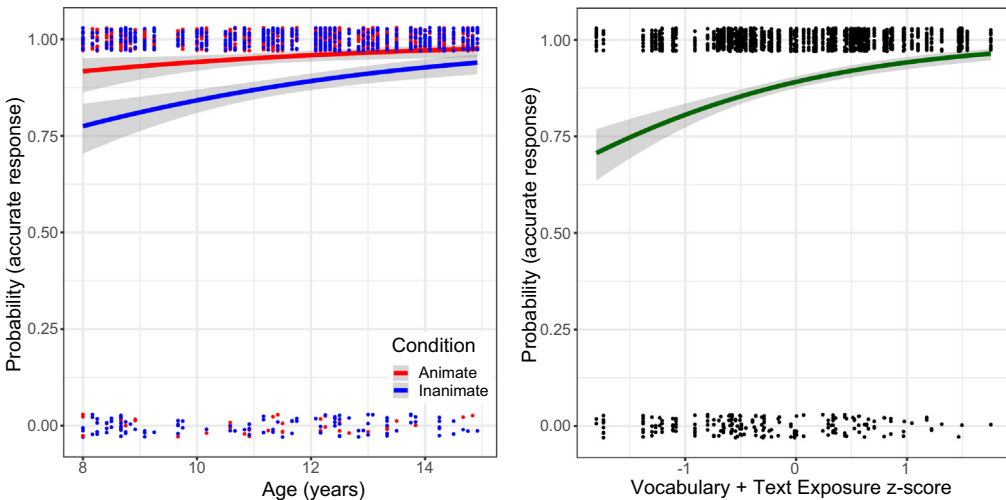


Fig. 4. Production accuracy: Predicted probabilities of accurate description for animacy conditions as a function of age (left panel) and predicted probabilities of accurate description as a function of vocabulary + text exposure (right panel).

Table 5

Coefficient estimates of linear mixed-effects models predicting production accuracy from age group, head noun animacy, and cognitive predictors.

	Coefficient	SE	z-score	p value
Intercept	1.03	0.81	1.27	.20
Animacy	0.66	0.35	1.91	.057
Age	0.11	0.07	1.65	.10
Vocabulary + text exposure	0.61	0.17	3.65	<.01*

rather than the objects in visually prompted descriptions. As suggested by many adult production studies, animate entities are more conceptually and visually salient, so participants in this condition need to inhibit the tendency to attend to and talk about animate characters (Arnold & Griffin, 2007; Bock & Warren, 1985; Fukumura et al., 2011; McDonald, Bock, & Kelly, 1993; Montag & MacDonald, 2014; Vogels, Krahmer, & Maes, 2013). Because younger participants have relatively less experience in producing inanimate-target RC descriptions (whether passive or active) compared with SRCs, this configuration is somewhat difficult for them if they attempt to follow the more frequent pattern (Roland et al., 2007). This interpretation is consistent with the fact that the main effect of age became nonsignificant ($p = .10$) (see Table 5) when vocabulary + text exposure scores were added to the model (age was significant in a model without predictors), indicating that despite these measures being correlated, vocabulary and reading experience fit the data better and may ultimately underpin performance.

Active/passive utterance choices

As indicated in the Introduction, animate targets are generally described in passives because preference for passives reflects a choice that speakers make online to maintain fluency in the face of interference (Gennari et al., 2012; MacDonald, 2013). The most explanatory model in Table 6 revealed a significant interaction between BDS and animacy condition ($p = .01$) and a three-way interaction among age, animacy, and BDS ($p < .01$). We explored this three-way interaction by median-splitting the data by age and BDS scores and examining the effects of age, BDS, and animacy in each subgroup. These are reported in Table 6, showing only the significant effects for brevity. Fig. 5 (top panel) shows the animacy modulation on structural choices across ages (left) and across BDS scores (right). The bottom panel of Fig. 5 shows the animacy modulation on subgroups defined by the median split across ages and BDS scores.

Recall that BDS scores and age were correlated with each other so that, overall, lower BDS scores correspond to younger participants and vice versa. Nevertheless, across the whole group, the animacy effect showed opposite trends as a function of age and BDS scores; the probability of producing passives increased with age, particularly for animate targets. In contrast, the probability of producing passives for animate targets decreased as a function of BDS, but that for inanimate targets increased. This

Table 6

Coefficient estimates of linear mixed-effects models predicting utterance choices (active/passive) from age group, animacy, and individual difference measures.

	Variable	Coefficient	SE	z-score	p value
<i>Full model</i>	Intercept	1.42	1.84	0.77	0.44
	Animacy	−0.33	1.18	−0.28	0.78
	Age	0.00	0.15	0.03	0.97
	BDS	0.21	1.93	0.11	0.92
	Animacy * Age	0.22	0.10	2.22	0.03*
	Animacy * BDS	−4.43	1.48	−3.00	<0.01*
	Age * BDS	0.00	0.16	−0.02	0.98
	Animacy * Age * BDS	0.32	0.12	2.68	<0.01*
<i>Split models</i>					
8–12 years old	Intercept	1.22	0.45	2.71	0.01*
	Animacy	2.05	0.27	7.47	<.001*
	Age * BDS	1.06	0.46	2.28	0.02*
	Animacy * BDS	−0.91	0.29	−3.12	0.002*
12–15 years old	Intercept	1.36	0.44	3.07	<.001*
	Animacy	2.53	0.29	8.67	<.001*
	Age * Animacy	0.85	0.30	2.78	0.01*
	Intercept	1.20	0.51	2.35	0.02*
Low BDS	Animacy	2.86	0.33	8.64	<.001*
	Intercept	1.70	0.41	4.16	<.001*
High BDS	Animacy	1.70	0.25	6.94	<.001*
	Age * Animacy	0.71	0.24	2.91	<.001*
	Intercept	1.70	0.25	6.94	<.001*

Note. BDS, backward digit span.

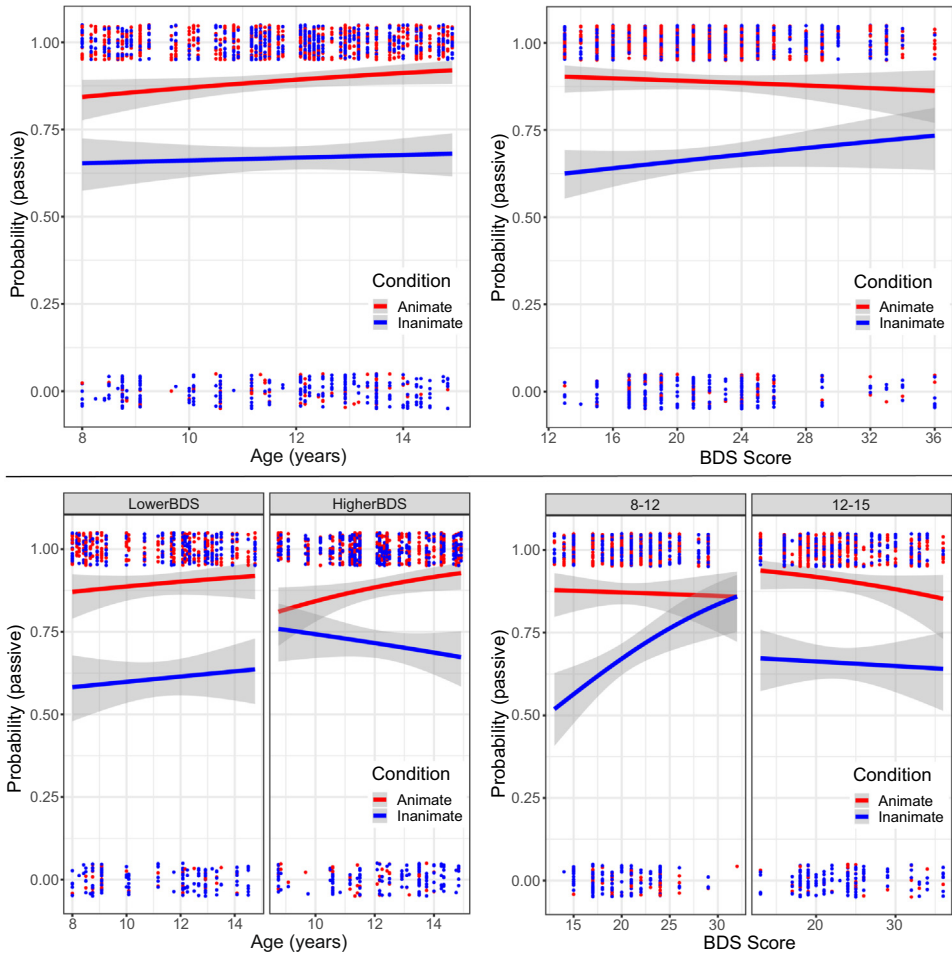


Fig. 5. Production of passive relative clauses: Top panel: Predicted probabilities of passive utterances as a function of age (left) and backward digit span (BDS) scores (right) across all participants. Bottom panel: Predicted probabilities of passive utterances median-split by participants' age (left) and BDS scores (right).

suggests that participants with better BDS scores were more able to utter active structures instead of passives for animate targets, but it also showed greater flexibility in adopting passives for inanimate targets.

The Age * Animacy interaction was also observed in the separate models fitting the data for each age subgroup. Both the younger and older groups showed an animacy effect and an Age * Animacy interaction (Table 6), suggesting that production choices continued to change with age. However, only 8- to 12-year-olds showed a BDS * Animacy interaction such that the animacy effect became smaller as BDS increased (Fig. 5, bottom right panel). Because 8- to 12-year-olds generally displayed more variation in BDS scores (including the lowest and highest values), their BDS scores strongly modulated production performance in this subgroup, where participants with scores as high as those of adolescents showed smaller differences between animacy conditions, whereas participants with lower scores showed large differences as in the overall BDS pattern (Fig. 5, top right panel).

Importantly, across the entire age range, the BDS subgroups indicate that participants with lower BDS scores only showed an animacy effect and behaved similarly across ages (Fig. 5, bottom left panel). In contrast, participants with higher BDS scores displayed an Age * Animacy interaction similar

to that obtained across the whole group, that is, an increase in passive use for animate-targets. These contrasting patterns suggest that age-related improvements in behavior are observed only in those participants with better BDS scores. Thus, despite age being correlated with BDS, BDS modulated production performance in the opposite direction to age and the rate of change on performance across ages was more pronounced in those individuals with higher BDS scores. Therefore, because chronological age and its correlates (including vocabulary + text exposure) are generally expected to modulate performance during development, the current results suggest that developmental improvements in production are strongly linked to performance in the BDS task.

Taken together, the current production results indicate an association of vocabulary + text exposure with description accuracy. Importantly, unlike comprehension, animacy interacted with age and BDS scores, suggesting that BDS and age played a more interactive role in planning production than in comprehension across animacy conditions. This result confirms our expectation in the Introduction that control-related processes are associated with performance improvements in the transition from childhood to adolescence beyond the influence of linguistic predictors, albeit only in production.

Comparing production and comprehension

Because different RCs and alternating animacy conditions were used in the production and comprehension tasks for a given participant and picture (cf. Fig. 2), mixed-effects models including animacy and DVs from both tasks were not possible. Nevertheless, we compared overall accuracy performance between tasks, averaging across animacy conditions. Comprehension accuracy was on average higher ($M = 97\%$) than production accuracy ($M = 88\%$). A linear regression model predicting the proportion of accurate performance as a function of task (production vs. comprehension) and age revealed a main effect of task and an interaction between age and task, as shown in Table 7 and Fig. 6. These results indicate large changes in production performance as a function of age but relatively stable performance in comprehension, suggesting more prominent developmental changes in production compared with comprehension.

Discussion

This study examined how animacy-induced processing difficulty changes with chronological age in both sentence production and comprehension as well as how these changes were associated with individual measures assessing vocabulary knowledge and experience and control-related BDS. Specifically, we asked whether the animacy effect would remain stable across ages and, moreover, whether our control measure (BDS) would interact with animacy and age beyond linguistic measures, thereby suggesting a role for control processes in explaining animacy-induced difficulty from childhood to adolescence. A summary of our main results across production and comprehension is provided in Table 8.

We found that although animacy and chronological age independently accounted for comprehension performance, animacy did not interact with age, indicating little developmental change in understanding animate-target ORCs compared with inanimate-target ones. Although comprehension times became generally faster with age, processing speed improved to a similar degree in both animacy conditions (cf. Fig. 4). In contrast, animacy-induced performance in production interacted with age even

Table 7
Regression results predicting accuracy from age and task (production vs. comprehension).

	Coefficient	SE	t score	p value
Intercept	0.97	0.04	24.29	<.0001*
Task	−0.25	0.06	−4.31	<.0001*
Age	0.00	0.00	0.03	0.98
Task * Age	0.02	0.00	3.32	.001*

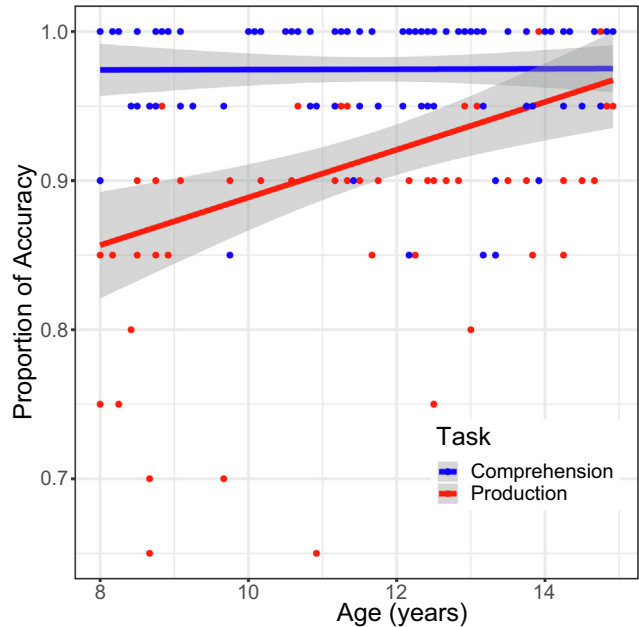


Fig. 6. Proportions of accurate performance in production and comprehension tasks as a function of age.

Table 8
Summary of results.

Model	Dependent variable	Main effects	Predictors	Interaction
Comprehension	Accuracy		BDS	
Production	Reaction time Accuracy	Animacy, age Age, animacy ¹	Vocabulary + text exposure	
Production vs. comprehension	Passive rate Accuracy	Animacy Task		Animacy * BDS * Age Task * Age

Note. BDS, backward digit span.

¹ Marginally significant effect ($p < .06$).

when older participants were separately examined. These results indicate that animacy-induced differences remain relatively stable in auditory comprehension but continue to change for spoken production into adolescence. Moreover, production accuracy was generally lower than comprehension accuracy and interacted with age (Fig. 6), suggesting pronounced developmental changes in production but not in comprehension. Importantly, BDS interacted with animacy and age in production, whereas it correlated only with overall accuracy in comprehension. This interaction in the production data is consistent with our expectation that control processes might be relevant in explaining differences in processing difficulty at this developmental stage. Taken together, the contrasts found between production and comprehension suggest that the ability to master RC production lags behind that for comprehension, in part because production is more difficult and engages more complex processes and brain structures than comprehension. As a result, control processes play a more prominent role in the development of production skills.

Developmental pattern in RC comprehension and production

Given that language comprehension and processing speed generally improve with age (Ferrer et al., 2007), it is not surprising that age-related improvements were observed in both RC comprehension and production; older participants were faster at comprehending active ORCs, provided more accurate RC descriptions in production, and used more passives for animate targets at a similar rate to that of adults. Importantly, target animacy modulated performance in both production and comprehension; animate-target RCs elicited longer comprehension responses and were more often described in passives compared with inanimate-target RCs, confirming previous findings in the developmental and adult literature (Booth et al., 2000; Humphreys et al., 2016; Macdonald et al., 2020; Montag & MacDonald, 2015). These results suggest greater difficulty associated with animate-head RCs due to competition or interference between similar animate nouns during interpretation or planning, consistent with previous adult results (Gennari et al., 2012; Humphreys et al., 2016; Wu et al., in press).

Importantly, although comprehension becomes faster with age, the size of the animacy effect in comprehension did not change given that there was no Animacy * Age interaction. In contrast, production choices displayed an Age * Animacy interaction across all ages (Table 6), suggesting that structural choices continue to change as a function of animacy. The likelihood of producing passives increased with age for animate targets, but not so much for inanimate targets (Fig. 5). A similar contrast between language tasks was also evident in performance accuracy, with marked age-related changes in production but not in comprehension (Fig. 6). Together, the current results suggest that participants experience more difficulty in production than in comprehension despite the fact that similar pictures and events were used in both tasks.

Different task demands are partly responsible for this contrast. Answering *yes* or *no* to an auditory description is surely easier than planning and outputting a complex syntactic structure. Nevertheless, these contrasting task demands reflect naturally occurring differences in ordinary communication; meaning understanding is often partially completed and/or facilitated by discourse contexts (including visual scenes), but spoken references cannot be left incomplete because interlocutors would not be able to identify the intended referent. This intrinsic contrast between production and comprehension is consistent with the observation that the ability to master production lags behind comprehension during early childhood (Clark & Hecht, 1983) and that production engages additional processes absent in comprehension (e.g., planning, ordered motor output). As evident in neurobiological studies, the production network is larger than the comprehension network and involves additional subcortical and control-related regions such as supplementary motor and cingulate cortices (de Zubicaray, Wilson, McMahon, & Muthiah, 2001; Humphreys & Gennari, 2014; Nozari & Novick, 2017; Whitney et al., 2008). Taken together, these observations and the current results suggest that RC production takes longer to master than comprehension and that the inherent complexity of production may engage control processes absent in comprehension, as discussed below.

Associations with individual measures

Consistent with the developmental differences between production and comprehension, BDS scores were also differentially associated with each task. Whereas BDS was positively associated with overall accuracy in comprehension (Fig. 3), it interacted with animacy and age in production. This three-way interaction indicated that age and BDS displayed opposite trends on the animacy effect (Fig. 5). Irrespective of age, higher-BDS participants more flexibly mixed actives and passives across animacy conditions, whereas lower-BDS participants showed a stronger tendency for opposite structures as a function of animacy. Importantly, higher-BDS participants displayed more pronounced animacy-induced changes as a function of age than lower-BDS ones, suggesting that higher BDS scores accelerate developmental changes in production. Thus, the ability to overcome animacy-induced competition is linked to BDS-related individual skills across ages.

The specific aspects of the BDS task that may matter for RC production are open to interpretation because this complex task can be construed in various ways, as suggested in our Introduction. Working memory approaches assume that BDS and other complex span tasks reflect memory capacity and executive or control processes like item manipulation and interference resolution (Boyle et al., 2013;

Montgomery et al., 2018; Roberts et al., 2007; Rusli & Montgomery, 2017). But the BDS is a verbal task and, as such, also recruits verbal processes (e.g., speech rate) and is modulated by prior knowledge and experience as other verbal tasks are (Jones et al., 2020). This does not exclude the involvement of attention or control mechanisms, but the problem of task impurity mentioned above (Miyake & Friedman, 2012) makes interpretation difficult.

In principle, maintaining and manipulating digits to be outputted in a specific order, as in the BDS task, seems relevant for RC production, where a choice between an active structure or a passive structure entails different word orders. Similarly, interference between items and positions in serial recall might relate to the interference between conceptually similar words in RC production. In this view, lower-BDS participants more often resort to passive RCs than to active RCs to avoid interference between similar candidate nouns for the same sentential position. In contrast, high-BDS participants were less constrained by animacy in their descriptions, more flexibly choosing passive or active structures and more easily managing interference. Whichever is the right interpretation, changes in BDS skills were strongly related to changes in structure planning, suggesting similar processes operating in both tasks and interdependent improvements over time.

Finally, the remaining associations with individual measures in our results are consistent with previous children's findings (e.g., Arosio et al., 2011, 2012; Montag & MacDonald, 2015). Specifically, BDS was associated with ORC comprehension accuracy irrespective of animacy, whereby lower-BDS participants made more errors. As suggested above, this association might stem from susceptibility to interference between nouns, their event roles, or pictured characters during understanding or decision making, although other alternative possibilities cannot be excluded. Moreover, vocabulary + text exposure explained production accuracy beyond age; speakers with higher scores produced more accurate descriptions than those with lower scores. This finding is consistent with the suggestion that reading experience provides exposure to syntactically complex structures that may be otherwise infrequent in speech such as passive RCs (Cunningham & Stanovich, 1991; Montag & MacDonald, 2015; Stanovich, 1989). Nevertheless, because all our individual measures were initially correlated with each other, the absence of a relationship with performance beyond age does not exclude the possibility that a relationship may exist when age-related variability is not so high. Indeed, recent adult studies examining individual differences in RC production or comprehension found significant interactions between vocabulary + text exposure and experimental conditions in both tasks (Van Dyke, Johns, & Kukona, 2014; Wu et al., *in press*), suggesting that there is still a role for this variable when fine-grained differences between same-age individuals are examined.

Overall, the current findings contribute to the understanding of developmental changes from late childhood to adolescence. First, they showed improvements across ages in comprehension speed, suggesting that typical increases in language experience and cognitive development associated with age contribute to improvements in processing speed. Second, our results showed that production is strongly linked to the control processes involved in serial recall and mental manipulation, consistent with developmental studies arguing for a significant development of control skills throughout this developmental period (Bunge & Wright, 2007; Casey et al., 2005; Constantinidis & Luna, 2019; Crone & Steinbeis, 2017; Gogtay et al., 2004; Tamm, Menon, & Reiss, 2002). Finally, the results showed different developmental patterns for production and comprehension when resolving animacy-induced interference, with each task relating to the same control measure in distinct ways. These findings suggest that each production and comprehension may rely on distinct control processes and develop at different rates across the life span, consistent with the partially distinct neural networks found in brain studies (Hickok & Poeppel, 2007; Humphreys & Gennari, 2014; Kreher et al., 2008). These contrasts between tasks thus point to a complex cognitive language architecture likely composed of a common knowledge base, as previously suggested, but linked to different task-specific or control processes.

Appendix A. Supplementary material

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.jecp.2021.105350>.

References

- Alario, F., Chainay, H., Lehericy, S., & Cohen, L. (2006). The role of the supplementary motor area (SMA) in word production. *Brain Research*, 1076, 129–143.
- Allum, P. H., & Wheeldon, L. R. (2007). Planning scope in spoken sentence production: The role of grammatical units. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 33, 791–810.
- Allum, P. H., & Wheeldon, L. (2009). Scope of lexical access in spoken sentence production: Implications for the conceptual–syntactic interface. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 35, 1240–1255.
- Arnold, J. E. (2010). How speakers refer: The role of accessibility. *Linguistics and Language Compass*, 4, 187–203.
- Arnold, J. E., & Griffin, Z. M. (2007). The effect of additional characters on choice of referring expression: Everyone counts. *Journal of Memory and Language*, 56, 521–536.
- Arnon, I. (2010). Rethinking child difficulty: The effect of NP type on children's processing of relative clauses in Hebrew. *Journal of Child Language*, 37, 27–57.
- Arosio, F., Guasti, M. T., & Stucchi, N. (2011). Disambiguating information and memory resources in children's processing of Italian relative clauses. *Journal of Psycholinguistic Research*, 40, 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.
- 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.
- Bentea, A., Durrleman, S., & Rizzi, L. (2016). Refining intervention: The acquisition of featural relations in object A-bar dependencies. *Lingua*, 169, 21–41.
- Blasey, C. M., Gl, G. H., Adelman, N. E., Menon, V., Blasey, C. M., White, C. D., ... Reiss, A. L. (2002). A developmental fMRI study of the Stroop color–word task. *NeuroImage*, 16, 61–75.
- Bock, J. K., Loebell, H., & Morey, R. (1992). From conceptual roles to structural relations: Bridging the syntactic cleft. *Psychological Review*, 99, 150–171.
- Bock, J. K., & Warren, R. K. (1985). Conceptual accessibility and syntactic structure in sentence formulation. *Cognition*, 21, 47–67.
- Booth, J. R., MacWhinney, B., & Harasaki, Y. (2000). Developmental differences in visual and auditory processing of complex sentences. *Child Development*, 71, 981–1003.
- Botvinick, M. M., Braver, T. S., Barch, D. M., Carter, C. S., & Cohen, J. D. (2001). Conflict monitoring and cognitive control. *Psychological Review*, 108, 624–652.
- Braze, D., Tabor, W., Shankweiler, D. P., & Mencl, W. E. (2007). Speaking up for vocabulary. *Journal of Learning Disabilities*, 40, 226–243.
- Brown, T. T., Lugar, H. M., Coalson, R. S., Miezin, F. M., Petersen, S. E., & Schlaggar, B. L. (2005). Developmental changes in human cerebral functional organization for word generation. *Cerebral Cortex*, 15, 275–290.
- Bunge, S. A., & Wright, S. B. (2007). Neurodevelopmental changes in working memory and cognitive control. *Current Opinion in Neurobiology*, 17, 243–250.
- Boyle, W., Lindell, A. K., & Kidd, E. (2013). Investigating the role of verbal working memory in young children's sentence comprehension. *Language Learning*, 63(2), 211–242. <https://doi.org/10.1111/lang.12003>.
- Casey, B. J., Tottenham, N., Liston, C., & Durston, S. (2005). Imaging the developing brain: What have we learned about cognitive development? *Trends in Cognitive Sciences*, 9, 104–110.
- Chang, F., Dell, G. S., & Bock, J. K. (2006). Becoming syntactic. *Psychological Review*, 113, 234–272.
- Clark, E. V., & Hecht, B. F. (1983). Comprehension, production and language acquisition. *Annual Review of Psychology*, 34, 325–349.
- Constantinidis, C., & Luna, B. (2019). Neural substrates of inhibitory control maturation in adolescence. *Trends in Neurosciences*, 42, 604–616.
- Contemori, C., & Garraffa, M. (2010). Comparison of modalities in SLI syntax: A study on the comprehension and production of non-canonical sentences. *Lingua*, 120, 1940–1955.
- Corrêa, L. M. S. (1995). An alternative assessment of children's comprehension of relative clauses. *Journal of Psycholinguistic Research*, 24, 183–203.
- Cowan, N., Sauls, J. S., Elliott, E. M., & Moreno, M. V. (2002). Deconfounding serial recall. *Journal of Memory and Language*, 46, 153–177.
- Crone, E. A., & Steinbeis, N. (2017). Neural perspectives on cognitive control development during childhood and adolescence. *Trends in Cognitive Sciences*, 21, 205–215.
- Cunningham, A. E., & Stanovich, K. E. (1991). Tracking the unique effects of print exposure in children: Associations with vocabulary, general knowledge, and spelling. *Journal of Educational Psychology*, 83, 264–274.
- Davidson, M. C., Amso, D., Anderson, L. C., & Diamond, A. (2006). Development of cognitive control and executive functions from 4 to 13 years: Evidence from manipulations of memory, inhibition, and task switching. *Neuropsychologia*, 44, 2037–2078.
- de Zubicaray, G., Wilson, S., McMahon, K., & Muthiah, S. (2001). The semantic interference effect in the picture–word paradigm: An event-related fMRI study employing overt responses. *Human Brain Mapping*, 14, 218–227.
- Diessel, H., & Tomasello, M. (2000). The development of relative clauses in child speech. *Cognitive Linguistics*, 11, 131–151.
- Dunn, L. M., & Dunn, D. M. (2007). *PPVT-4: Peabody Picture Vocabulary Test*. San Antonio, TX: Pearson Assessments.
- Felscher, C., Clahsen, H., & Munte, T. F. (2003). Storage and integration in the processing of filler-gap dependencies: An ERP study of topicalization and wh-movement in German. *Brain and Language*, 87, 345–354.
- Ferrer, E., McArdle, J. J., Shaywitz, B. A., Holahan, J. M., Marchione, K., & Shaywitz, S. E. (2007). Longitudinal models of developmental dynamics between reading and cognition from childhood to adolescence. *Developmental Psychology*, 43, 1460–1473.
- Finney, M. C., Montgomery, J. W., Gillam, R. B., & Evans, J. L. (2014). Role of working memory storage and attention focus switching in children's comprehension of spoken object relative sentences. *Child Development Research*, 2014, 450734.
- Fox, B. A., & Thompson, S. A. (1990). A discourse explanation of the grammar of relative clauses in English conversation. *Language*, 66, 297–316.

- Fukumura, K., van Gompel, R. P. G. G., Harley, T., & Pickering, M. J. (2011). How does similarity-based interference affect the choice of referring expression? *Journal of Memory and Language*, 65, 331–344.
- Gathercole, S. E., Pickering, S. J., Ambridge, B., & Wearing, H. (2004). The structure of working memory from 4 to 15 years of age. *Developmental Psychology*, 40, 177–190.
- Gennari, S. P., & MacDonald, M. C. (2008). Semantic indeterminacy in object relative clauses. *Journal of Memory & Language*, 58, 161–187.
- Gennari, S. P., & MacDonald, M. C. (2009). Linking production and comprehension processes: The case of relative clauses. *Cognition*, 111, 1–23.
- Gennari, S. P., Mirković, J., & MacDonald, M. C. (2012). Animacy and competition in relative clause production: A cross-linguistic investigation. *Cognitive Psychology*, 65, 141–176.
- Gibson, E. (1998). Linguistic complexity: Locality of syntactic dependencies. *Cognition*, 68, 1–76.
- Gogtay, N., Giedd, J. N., Lusk, L., Hayashi, K. M., Greenstein, D., Vaituzis, A. C., ... Thompson, P. M. (2004). Dynamic mapping of human cortical development during childhood through early adulthood. *Proceedings of the National Academy of Sciences of the United States of America*, 101, 8174–8179.
- Gratton, G., Cooper, P., Fabiani, M., Carter, C. S., & Karayanidis, F. (2018). Dynamics of cognitive control: Theoretical bases, paradigms, and a view for the future. *Psychophysiology*, 55, 1–29.
- Håkansson, G., & Hansson, K. (2000). Comprehension and production of relative clauses: A comparison between Swedish impaired and unimpaired children. *Journal of Child Language*, 27, 313–333.
- Harlaar, N., Dale, P. S., & Plomin, R. (2007). Reading exposure: A (largely) environmental risk factor with environmentally-mediated effects on reading performance in the primary school years. *Journal of Child Psychology and Psychiatry and Allied Disciplines*, 48, 1192–1199.
- Hickok, G., & Poeppel, D. (2007). The cortical organization of speech processing. *Nature Reviews Neuroscience*, 8, 393–402.
- Hsiao, Y., & MacDonald, M. C. (2016). Production predicts comprehension: Animacy effects in Mandarin relative clause processing. *Journal of Memory and Language*, 89, 87–109.
- Hsu, N. S., & Novick, J. M. (2016). Dynamic engagement of cognitive control modulates recovery from misinterpretation during real-time language processing. *Psychological Science*, 27, 572–582.
- Huizinga, M., Dolan, C. V., & van der Molen, M. W. (2006). Age-related change in executive function: Developmental trends and a latent variable analysis. *Neuropsychologia*, 44, 2017–2036.
- Humphreys, G. F., & Gennari, S. P. (2014). Competitive mechanisms in sentence processing: Common and distinct production and reading comprehension networks linked to the prefrontal cortex. *NeuroImage*, 84, 354–366.
- Humphreys, G. F., Mirković, J., & Gennari, S. P. (2016). Similarity-based competition in relative clause production and comprehension. *Journal of Memory and Language*, 89, 200–221.
- Hurlstone, M. J., Hitch, G. J., & Baddeley, A. D. (2014). Memory for serial order across domains: An overview of the literature and directions for future research. *Psychological Bulletin*, 140, 339–373.
- Iacobucci, D., Schneider, M. J., Popovich, D. L., & Bakamitsos, G. A. (2016). Mean centering helps alleviate “micro” but not “macro” multicollinearity. *Behavior Research Methods*, 48, 1308–1317.
- Jensen De Lopez, K., Sundahl Olsen, L., Chondrogianni, V., De -López, K. J., Olsen, L. S., & Chondrogianni, V. (2014). Annoying Danish relatives: Comprehension and production of relative clauses by Danish children with and without SLI. *Journal of Child Language*, 41, 51–83.
- Jones, G., Justice, L. V., Cabiddu, F., Lee, B. J., Iao, L. S., Harrison, N., & Macken, B. (2020). Does short-term memory develop? *Cognition*, 198, 104200.
- Jones, G., & Macken, B. (2015). Questioning short-term memory and its measurement: Why digit span measures long-term associative learning. *Cognition*, 144, 1–13.
- Just, M. A., & Carpenter, P. A. (1992). A capacity theory of comprehension: Individual differences in working memory. *Psychological Review*, 99, 122–149.
- Kidd, E. (2013). The role of verbal working memory in children's sentence comprehension: A critical review. *Topics in Language Disorders*, 33, 208–223.
- Kidd, E., & Bavin, E. L. (2002). English-speaking children's comprehension of relative clauses: Evidence for general-cognitive and language-specific constraints on development. *Journal of Psycholinguistic Research*, 31, 599–617.
- Kidd, E., Brandt, S., Lieven, E., & Tomasello, M. (2007). Object relatives made easy: A cross-linguistic comparison of the constraints influencing young children's processing of relative clauses. *Language and Cognitive Processes*, 22, 860–897.
- Konopka, A. E., & Kuchinsky, S. E. (2015). How message similarity shapes the timecourse of sentence formulation. *Journal of Memory and Language*, 84, 1–23.
- Kreher, W., Schnell, S., Ku, D., Vry, M.-S., Umarova, R., Musso, M., ... Weiller, C. (2008). Ventral and dorsal pathways for language. *Proceedings of the National Academy of Sciences of the United States of America*, 105, 18035–18040.
- Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest package: Tests in linear mixed effects models. *Journal of Statistical Software*, 82(13). <https://doi.org/10.18637/jss.v082.i13>.
- Lewis, R. L., Vasishth, S., & Van Dyke, J. A. (2006). Computational principles of working memory in sentence comprehension. *Trends in Cognitive Sciences*, 10, 447–454.
- MacDonald, M. C. (2016). Speak, act, remember: The language-production basis of serial order and maintenance in verbal memory. *Current Directions in Psychological Science*, 25, 47–53.
- MacDonald, M. C., & Christiansen, M. H. (2002). Reassessing working memory: Comment on Just and Carpenter (1992) and Waters and Caplan (1996). *Psychological Review*, 109, 35–74.
- MacDonald, M. C., Pearlmuter, N., & Seidenberg, M. (1994). Lexical nature of syntactic ambiguity resolution. *Psychological Review*, 101, 676–703.
- Macdonald, R., Brandt, S., Theakston, A., Lieven, E., & Serratrice, L. (2020). The role of animacy in children's interpretation of relative clauses in English: Evidence from sentence-picture matching and eye movements. *Cognitive Science*, 44, 3–35.
- Mak, W. M., Vonk, W., & Schriefers, H. (2002). The influence of animacy on relative clause processing. *Journal of Memory and Language*, 47(1), 50–68. <https://doi.org/10.1006/jmla.2001.2837>.

- Matthews, D., Lieven, E., Theakston, A., & Tomasello, M. (2005). The role of frequency in the acquisition of English word order. *Cognitive Development*, 20, 121–136.
- McDaniel, D., McKee, C., & Bernstein, J. B. (1998). How children's relatives solve a problem for minimalism. *Language*, 74, 308–334.
- McDonald, J. L., Bock, J. K., & Kelly, M. H. (1993). Word and world order: Semantics, phonological, and metrical determinants of serial position. *Cognitive Psychology*, 25, 188–230.
- MacDonald, M.C. (2013). How language production shapes language form and comprehension. *Frontiers in psychology*, 4, 226. <http://doi.org/10.3389/fpsyg.2013.00226>.
- Miyake, A., & Friedman, N. P. (2012). The nature and organization of individual differences in executive functions: Four general conclusions. *Current Directions in Psychological Science*, 21, 8–14.
- Miyake, A., Friedman, N. P., Emerson, M. J., Witzki, A. H., Howerter, A., & Wager, T. D. (2000). The unity and diversity of executive functions and their contributions to complex “frontal lobe” tasks: A latent variable analysis. *Cognitive Psychology*, 41, 49–100.
- Montag, J. L., & MacDonald, M. C. (2014). Visual salience modulates structure choice in relative clause production. *Language and Speech*, 57, 163–180.
- Montag, J. L., & MacDonald, M. C. (2015). Text exposure predicts spoken production of complex sentences in 8- and 12-year-old children and adults. *Journal of Experimental Psychology: General*, 144, 447–468.
- Montgomery, J. W. (1995). Sentence comprehension in children with specific language impairment: The role of phonological working memory. *Journal of Speech, Language, and Hearing Research*, 38, 187–199.
- Montgomery, J. W., Evans, J. L., Fargo, J. D., Schwartz, S., & Gillam, R. B. (2018). Structural relationship between cognitive processing and syntactic sentence comprehension in children with and without developmental language disorder. *Journal of Speech, Language, and Hearing Research*, 61, 2950–2976.
- Montgomery, J. W., Evans, J. L., & Gillam, R. B. (2009). Relation of auditory attention and complex sentence comprehension in children with specific language impairment: A preliminary study. *Applied Psycholinguistics*, 30, 123–151.
- 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.
- Novick, J. M., Trueswell, J. C., & Thompson-Schill, S. L. (2010). Broca's area and language processing: Evidence for the cognitive control connection. *Language and Linguistics Compass*, 4, 906–924.
- Nozari, N., & Novick, J. (2017). Monitoring and control in language production. *Current Directions in Psychological Science*, 26, 403–410.
- Perfetti, C. (2007). Reading ability: Lexical quality to comprehension. *Scientific Studies of Reading*, 11, 357–383.
- Pickering, M. J., & Ferreira, V. S. S. (2008). Structural priming: A critical review. *Psychological Bulletin*, 134, 427–459.
- Pickering, S., & Gathercole, S. E. (2001). *Working Memory Test Battery for Children (WMTB-C)*. London: Psychological Corporation.
- Postle, B. R. (2006). Working memory as an emergent property of the mind and brain. *Neuroscience*, 139, 23–38.
- Real, F., & Christiansen, M. H. (2007). Processing of relative clauses is made easier by frequency of occurrence. *Journal of Memory and Language*, 57, 1–23.
- Roberts, L., Marinis, T., Felser, C., & Clahsen, H. (2007). Antecedent priming at trace positions in children's sentence processing. *Journal of Psycholinguistic Research*, 36, 175–188.
- Roland, D., Dick, F., & Elman, J. L. (2007). Frequency of basic English grammatical structures: A corpus analysis. *Journal of Memory and Language*, 57, 348–379.
- Rubia, K., Smith, A. B., Woolley, J., Nosarti, C., Heyman, I., Taylor, E., & Brammer, M. (2006). Progressive increase of frontostriatal brain activation from childhood to adulthood during event-related tasks of cognitive control. *Human Brain Mapping*, 27, 973–993.
- Rusli, Y. A., & Montgomery, J. W. (2017). Children's comprehension of object relative sentences: It's extant language knowledge that matters, not domain-general working memory. *Journal of Speech, Language, and Hearing Research*, 60, 2865–2878.
- Sesma, H. W., Mahone, E. M., Levine, T., Eason, S. H., & Cutting, L. E. (2009). The contribution of executive skills to reading comprehension. *Child Neuropsychology*, 15, 232–246.
- Shaw, P., Kabani, N. J., Lerch, J. P., Eckstrand, K., Lenroot, R., Gogtay, N., ... Wise, S. P. (2008). Neurodevelopmental trajectories of the human cerebral cortex. *Journal of Neuroscience*, 28, 3586–3594.
- Skeide, M. A., & Friederici, A. D. (2016). The ontogeny of the cortical language network. *Nature Reviews Neuroscience*, 17, 323–332.
- Smith, M., & Wheeldon, L. (2004). Horizontal information flow in spoken sentence production. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 30, 675–686.
- Stanovich, K.E., & West, R.F. (1989). Exposure to print and orthographic processing. *Reading research quarterly*, 402–433. <http://dx.doi.org/10.2307/747605>.
- Tamm, L., Menon, V., & Reiss, A. L. (2002). Maturation of brain function associated with response inhibition. *Journal of the American Academy of Child & Adolescent Psychiatry*, 41, 1231–1238.
- Traxler, M. J., Morris, R. K., & Seely, R. E. (2002). Processing subject and object relative clauses: Evidence from eye movements. *Journal of Memory and Language*, 47, 69–90.
- Traxler, M. J., Williams, R. S., Blozis, S. A., & Morris, R. K. (2005). Working memory, animacy, and verb class in the processing of relative clauses. *Journal of Memory and Language*, 53, 204–224.
- Trueswell, J. C., Tanenhaus, M. K., & Garnsey, S. M. (1994). Semantic influences on parsing: Use of thematic role information in syntactic ambiguity resolution. *Journal of Memory and Language*, 33, 285–318.
- Van Dyke, J. A., Johns, C. L., & Kukona, A. (2014). Low working memory capacity is only spuriously related to poor reading comprehension. *Cognition*, 131, 373–403.
- Van Dyke, J. A., & Lewis, R. L. (2003). Distinguishing effects of structure and decay on attachment and repair: A cue-based parsing account of recovery from misanalyzed ambiguities. *Journal of Memory and Language*, 49(3), 285–316. [https://doi.org/10.1016/S0749-596X\(03\)00081-0](https://doi.org/10.1016/S0749-596X(03)00081-0).
- Vogels, J., Krahmer, E., & Maes, A. (2013). Who is where referred to how, and why? The influence of visual saliency on referent accessibility in spoken language production. *Language and Cognitive Processes*, 28, 1323–1349.

- Wechsler, D., & Hsiao-Pin, C. (2011). *WASI II: Wechsler Abbreviated Scale of Intelligence* (2nd ed.). San Antonio, TX: Psychological Corporation.
- Whitney, C., Weis, S., Krings, T., Huber, W., Grossman, M., & Kircher, T. (2008). Task-dependent modulations of prefrontal and hippocampal activity during intrinsic word production. *Journal of Cognitive Neuroscience*, 21, 697–712.
- Wu, S. H., & Gennari, S. (2021). *Sentence comprehension and production from late childhood to adolescence*. Retrieved from osf.io/vursx.
- Wu, S. H., Henderson, L., & Gennari, S. P. (in press). Animacy interactions with individual variability in sentence production and comprehension reveal similar lexically-driven competitive processes. *Journal of Experimental Psychology: Learning, Memory, and Cognition*.