LANGUAGE LEARNING
A Journal of Research in Language Studies

Language Learning ISSN 0023-8333

Investigating the Role of Verbal Working Memory in Young Children's Sentence Comprehension

Whitney Boyle,[a] Annukka K. Lindell,[a] and Evan Kidd[b]

[a]La Trobe University and [b]The Australian National University

This study considers the role of verbal working memory in sentence comprehension in typically developing English-speaking children. Fifty-six ($N = 56$) children aged 4;0–6;6 completed a test of language comprehension that contained sentences which varied in complexity, standardized tests of vocabulary and nonverbal intelligence, and three tests of memory that measured the three verbal components of Baddeley's model of Working Memory (WM): the phonological loop, the episodic buffer, and the central executive. The results showed that children experienced most difficulty comprehending sentences that contained noncanonical word order (passives and object relative clauses). A series of linear mixed effects models were run to analyze the contribution of each component of WM to sentence comprehension. In contrast to most previous studies, the measure of the central executive did not predict comprehension accuracy. A canonicity by episodic buffer interaction showed that the episodic buffer measure was positively associated with better performance on the noncanonical sentences. The results are discussed with reference to capacity-limit and experience-dependent approaches to language comprehension.

Keywords first language acquisition; Working Memory; sentence comprehension; word order; sentence repetition; passive voice; relative clauses

---

We thank all of the children, kindergartens, and schools that participated and the Department of Early Education and Childhood Development for their cooperation. Thanks also go to Grzegorz Krajewski for extremely helpful statistical advice and to Professors Lourdes Ortega and Marianne Gullberg in addition to three anonymous reviewers for detailed and very helpful comments. This experiment was conducted as part of the first author's Ph.D. candidature. Preparation was supported by a Charles La Trobe Research Fellowship to Kidd.

Correspondence concerning this article should be addressed to Whitney Boyle, School of Psychological Sciences, La Trobe University, Melbourne 3086, Australia; or to Evan Kidd, Research School of Psychology, The Australian National University, Canberra ACT 0200, Australia. E-mail: w.boyle@latrobe.edu.au or evan.kidd@anu.edu.au

Language Learning 63:2, June 2013, pp. 211–242

DOI: 10.1111/lang.12003

Introduction

The role of verbal working memory (WM) in language acquisition is poorly understood. Procedurally, both language comprehension and production necessarily involve access to linguistic knowledge stored in long-term memory (LTM), which must be assembled into a grammatically permissible form, a process which is likely to at least partially involve WM. Debates about the role of WM in adult sentence processing have been ongoing for some time (e.g., Just & Carpenter, 1992; MacDonald & Christiansen, 2002; Waters & Caplan, 1996; Wells, Christiansen, Race, Acheson, & MacDonald, 2009). However, an understanding of the role of WM in acquisition has been hampered by inconsistencies in the selection of items and WM tests across studies. In the current study we systematically investigated the role of WM in children's comprehension of sentences that vary in complexity.

Background

WM is often conceptualized as a construct describing the concurrent maintenance and manipulation (i.e., processing) of information (Oberauer & Lewandowsky, 2010). A number of different theoretical models of WM exist (e.g, Baddeley, 2007; Cowan, 2005; Just & Carpenter, 1992; Waters & Caplan, 1996, see also papers in Barrouillet & Gaillard, 2011). For instance, a broad distinction can be made between approaches that conceptualize WM as consisting of multiple domain-specific components (e.g., Baddeley, 2007), and those that in contrast conceptualize WM as a domain-general system that controls the focus of attention (e.g., Cowan, 2005). While there are definite structural and conceptual differences between the theories (see Shah & Miykae, 1999), recent theoretical accounts have argued that such differences may be a matter of emphasis rather than being substantive (Logie, 2011). With respect to language processing and acquisition, there is broad empirical agreement that both processes implicate the short-term store of phonological information and the binding and manipulation of linguistic units.

In the current study we used the conceptual framework of one prominent model of WM—Baddeley's (2007) multiple component model of WM—to investigate the role of WM in 4- to 6-year-old children's sentence comprehension. We chose to use the multiple component framework because it is the most thoroughly studied and best-attested model of WM in children within the age range in which we were interested. Several large studies (Alloway, Gathercole, & Pickering, 2006; Alloway, Gathercole, Willis, & Adams, 2004; Gathercole, Pickering, Ambridge, & Wearing, 2004) have supported the presence of a WM system in children akin to that described by Baddeley and Hitch (1974) and Baddeley (2007). Most relevant to the current study, Alloway et al. (2004) tested 633 children aged 4 to 6 years on multiple measures of memory and found that the best fitting structural model of children's memory included the same components that have been well established in adults, namely, the central executive, the episodic buffer, and the phonological loop and visuospatial sketchpad. We next briefly describe each of these subcomponents of WM.

Baddeley's (2007) Model of WM

The phonological loop (or phonological short-term memory [STM]) and the visuospatial sketchpad are stimulus-specific subsystems of the WM model; the phonological loop is specialized for auditory information, and the visuspatial sketchpad for visual information. The phonological loop is argued to have two roles: first, to hold incoming memory traces of auditory information for a short period (seconds) and, second, to rehearse information in order to prevent it from fading (Baddeley, 2007). The phonological loop is capacity limited; its functioning or capacity is often tested by requiring participants to recall various types of lists. The visuospatial sketchpad performs a similar role to the phonological loop but takes visual and spatial information as its input. Like the phonological loop, the visuospatial sketchpad also has a limited capacity (approximately 3 to 4 objects). In the current study we refer to the phonological loop as phonological STM.

The episodic buffer was added to the WM model by Baddeley (2000). In its current formulation (Baddeley, Allen, & Hitch, 2010), the episodic buffer is conceptualized as a multidimensional yet passive store that receives input from LTM and components of WM. Its inclusion in the model was driven by the need to incorporate the influence of long-term knowledge and skills in WM performance (e.g., Ericsson & Kintsch, 1995; see also Shah & Miyake, 1999). Its role is to temporarily store chunks of integrated information that have been constructed from a range of memory subsystems that have different basic memory codes. Its contents are assumed to be available to conscious awareness. A recent addition to the WM model, the episodic buffer has not yet been subject to as much research attention as the other WM components, particularly in developmental studies.

Finally, the central executive is argued to be responsible for directing and controlling attention, and allocating the finite resources of the WM system (Baddeley, 2007). Like the other WM components, the capacity of the central executive is limited. Central executive functioning is commonly measured using tasks that require both the retention and manipulation of information.

WM and Children's Sentence Comprehension

There is evidence that various components of WM account for aspects of language acquisition over and above the contribution of age. For instance, many studies have implicated phonological STM in vocabulary acquisition (Baddeley, Gathecole, & Papagno, 1998; Gathercole, Willis, & Baddeley, 1992; for a discussion, see Gathercole, 2006). Other studies have shown that children with Specific Language Impairment (SLI) perform worse on measures of phonological STM than do matched peers (Conti-Ramsden, Botting, & Faragher, 2001). Willis and Gathercole (2001) reported that phonological STM predicts children's ability to produce a range of sentence types. The role of the other WM components in language acquisition has not been studied in similar detail. This is problematic, because phonological STM must interact with other memory systems (e.g., the central executive, LTM) in order for both structure and meaning to be extracted from the speech stream.

Several past studies have investigated the role of WM processes other than phonological STM in children's sentence comprehension. Each has followed the adult literature in assuming that complex memory processes approximating something like Baddeley's (2007) central executive are implicated in the mastery and processing of complex structures. For instance, a common finding in the adult literature is that WM span, as measured by the Daneman and Carpenter (1980) listening span task, predicts the ease with which first-language adult speakers of English process center-embedded object relative clauses (e.g., *The senator that the assassin shot _ died later in hospital*) in comparison to the syntactically less complex subject relative clauses (e.g., Just & Carpenter, 1992; but see Waters & Caplan, 1996). By contrast, the available developmental studies carried out with children have produced fairly inconsistent results, mainly due to marked differences in the sentence types tested and the WM measures that have been used. Table 1 summarizes six key studies that investigated WM and sentence comprehension in young children.

Montgomery and colleagues have investigated the relationship between sentence comprehension and different components of WM. Montgomery, Magimairaj, and O'Malley (2008) used WM tasks that measure phonological STM (a nonword repetition task) and resource allocation and control (akin to central executive functioning) to investigate the relationship between WM and comprehension in typically developing 6- to 12-year-old children. The children were tested on sentences categorized as simple and complex. The simple sentences

Table 1 Overview of past studies of working memory (WM) and sentence comprehension by children

Authors & Population	WM Measures	Sentence Types	Results
Montgomery, Magimairaj, & O'Malley (2008) Typically developing (TD) 6- to 12-year-olds ($M = 8;8$)	1. NWRep (Dollaghan & Campbell, 1992) 2. ARAC (Martin & Schwartz, 2003)	1. Simple actives 2. Reflexive/ Pronominal sentences 3. Passives	ARAC → Complex sentences (Reflexive/ Pronominal sentences + passives)
Montgomery & Evans (2009) SLI children ($M = 9;1$) Aged matched ($M = 9;1$) and language matched ($M = 6;3$) TD controls	1. NWRep (Dollaghan & Campbell, 1992) 2. ARCC (Gaulin & Campbell, 1994)	1. Simple actives 2. Reflexive/ Pronominal sentences 3. Passives	NWRep → Simple actives (SLI) ARAC → Complex sentences (SLI, aged matched)
Montgomery, Evans & Gillam (2009) SLI children ($M = 8;5$) Aged matched TD controls ($M = 8;2$)	1. ARAC (Montgomery, 2000)	1. Simple sentences (double marking of number & extra verbiage) 2. Reduced subject relative clauses (one- & two-clause)	ARAC → Complex (Relative clauses) (SLI)
Booth, MacWhinney, & Harasaki (2000) 8- to 12-year-old TD children	1. (Forward) digit-span 2. WM/Listening span (Daneman & Carpenter, 1980; Swanson, 1996)	1. Conjoined clauses 2. Subject relative clauses 3. Object relative clauses	Forward digit span → Subject & object relative clauses (online) WM/Listening span → Off-line final interpretations

(*Continued*)

Table 1 Continued

Authors & Population	WM Measures	Sentence Types	Results
Arosio, Guasti, & Stucchi (2011) 9-year-old TD Italian children	1. (Forward) digit-span 2. WM/Listening span (Swanson, 1996)	1. Subject relative clauses 2. Object relative clauses	Forward digit span → Off line comprehension (object relative clauses)
Felser, Marinis, & Clahsen (2003) 6- to 7-year-old TD children	1. WM/Listening span (Daneman & Carpenter, 1980; Swanson, 1996)	1. Relative clause attachment ambiguities	WM/Listening span → Online ambiguity resolution

Note. NWRep = nonword repetition; ARAC = attentional resource allocation/control task; SLI = Specific Language Impairment; TD = typically developing.

were simple actives. The complex sentence category consisted of two sentence types: (i) passives and (ii) reflexive/pronominal sentences (see Table 1). The results showed no relationship between the WM measures and simple sentence comprehension, but a significant relationship was found between a measure of resource allocation/control and the children's processing of complex sentences. The researchers concluded that the comprehension of complex sentences is supported by complex WM processes in development over and above the influence of age.

In another study using the same sentence comprehension task, Montgomery and Evans (2009) investigated the contribution of WM to sentence comprehension in typically developing and SLI children. The results showed that, for the SLI children (mean age = 9;1), nonword repetition scores significantly predicted simple sentence comprehension and performance on a different resource allocation and control task from the one used in Montgomery et al. (2008) significantly predicted complex sentence comprehension. A younger (mean age = 6;3) language- and WM-matched control group only showed the latter relationship, and a typically developing age-matched group showed no relationship between WM and comprehension at all. Finally, Montgomery, Evans, and Gillam (2009) largely replicated the findings of Montgomery and Evans (2009), showing that a resource and allocation/control task predicted complex sentence comprehension in SLI children aged 8;5 years, but found no relationship between WM and sentence comprehension in typically developing age-matched controls. This study operationalized sentence complexity differently to the two previous studies (see Table 1). Two types of simple sentences were used (Noun Phrase-Prepositional Phrase construction and actives). The complex sentences were reduced subject relative clauses that varied in the number of clauses they contained.

In sum, studies that have investigated WM and sentence comprehension in typically developing and SLI children have yielded a mixed pattern of results, where the role of WM memory in sentence comprehension appears to interact with the age and clinical status of children. Two significant problems limit any conclusions that can be made from these data. First, in each study a different measure of complex memory span/central executive was used. Second, the operationalization of sentence complexity was inconsistent. In both Montgomery et al. (2008) and Montgomery and Evans (2009), the category of complex sentences included two sentence types: (i) passive constructions such as *Goldilocks was frightened by the Bear* and (ii) sentences that, for children, contain potentially ambiguous pronominal reference, such as *Papa Bear says Baby Bear is tickling him/himself*. The difficulty associated with the passive is most likely syntactic: The English full BE passive is low in frequency and is acquired late. In contrast, the difficulty associated with pronominal reference is at least partially discourse pragmatic (Matthews, Lieven, Theakston, & Tomasello, 2009; O'Grady, 2005). Therefore it is unclear whether the two sentence types should be treated as equivalent in terms of their complexity. Moreover, Montgomery and colleagues did not report the results for the individual sentence types, and it is therefore possible that only one of these sentence types is in actual fact associated with resource allocation and control.[1] Finally, the items used by Montgomery et al. (2009) were reduced subject relative clauses, such as *The boy standing is kissing the little girl sitting*. These sentences are likely to be complex because they contain temporary main clause/reduced relative clause ambiguities; that is, they are garden-path sentences (see Trueswell, Tanenhaus, & Garnsey, 1994). Typically developing children do not seem to reanalyze garden-path sentences as readily as do adults (Kidd, Stewart, & Serratrice, 2011; Trueswell, Sekerina, Hill, & Logrip, 1999). Therefore, the presence of temporary ambiguities in the test sentences suggests that the difficulty in these sentences derives from postinterpretative processes involved in the reanalysis of misparsed structure.

Studies that have investigated the online sentence processing of complex structures have also yielded mixed results (see Table 1). For example, Booth, MacWhinney, and Harasaki (2000) reported that 8- to 12-year-old children's forward digit-span performance, a task likely to measure phonological STM

(Alloway et al., 2004; Gathercole et al., 2004), predicted online processing of both subject and object center-embedded relative clause sentences. This effect held across two modalities: visual (i.e., reading) and auditory (i.e., listening). In contrast, the children's listening span (Swanson, 1996), a measure that is likely to be dependent on central executive functioning, predicted individual differences in the children's off-line final interpretations in the reading task, whereas digit span predicted the off-line interpretations in the auditory listening task. Arosio, Guasti, and Stucchi (2011) also reported that forward digit span predicted the offline comprehension of object relative clauses in 9-year-old Italian children, although they did not observe the same effect in the online data. These results contrast with findings reported by Felser, Marinis, and Clahsen (2003), who reported that 6- to 7-year-old children's listening span predicted their online resolution of sentences that contain ambiguity of relative clause attachment (e.g., *The husband of the actress <u>who was on the balcony</u>*).

The online studies therefore suggest some involvement of both phonological STM and the central executive in complex sentence interpretation, although the precise contribution of each component is far from clear. This may be once again due to differences in items across studies. All studies used relative clause structures; however, their materials are likely to be complex for different reasons. Restrictive relative clauses have been argued to be complex on the basis of various criteria, and their difficulty has been explained in a number of different ways. For instance, it has been attributed to: (i) difficulty with syntactic derivation (i.e., "movement"; e.g., Friedmann, Beletti, & Rizzi, 2009; Rizzi, 1990), (ii) perspective taking (e.g., MacWhinney, 1999), (iii) memory limitations on thematic role assignment (Gibson, 1998; O'Grady, 2011), and (iv) deviations from canonical word order (Bates, Devescovi, & D'Amico, 1999; Kidd & Bavin, 2002). Relative clause attachment ambiguities, as used by Felser et al. (2003), have been explained using a different set of theoretical concepts (e.g., see Fernández, 2003; Fodor, 1998).

One final set of results (not included in Table 1) that sheds further ambiguity on the role of WM in language acquisition comes from work on children with developmental disorders other than SLI. Working within the Baddeley WM model framework, Alloway and Gathercole (2005) reported that sentence repetition, a measure of the episodic buffer, was positively associated with a standardized measure of spoken language skills in 7- to 11-year-old children with learning difficulties, but a composite measure of the central executive (combining backward digit span, listening span, and counting recall) was not. While these results are preliminary because their outcome measure of language was broad, the data suggest that the episodic buffer may play an important role in spoken language comprehension and use (see also Marshall & Nation, 2003).

The Current Study

WM has long been argued to be implicated in child first language acquisition. However, to date its role has been obscured by inconsistent results. In part these inconsistencies are likely to be due to small sample sizes and differences in the ages of the children tested and in the WM measures used. Another contributing factor appears to be item selection and associated definitions of item complexity. In the current study we aimed to reassess the role of WM in children's sentence comprehension. We approached this task by reverting back to first principles: We took the most established and psychometrically validated model of WM in children, Baddeley's (2007) WM model, and tested whether each component predicted 4- to 6-year-old children's comprehension of sentences that were manipulated according to two principled definitions of sentence complexity.

On the one hand, complexity can be defined in the traditional linguistic sense, where any sentence that contains a dependent clause or a displaced element is considered to be complex (Lust, Foley, & Dye, 2009). On this definition, sentences (1) and (2), a simple active and a passive sentence, respectively, are simple sentences because they are monoclausal and do not contain displaced noun phrases (NPs):

(1) The dog chased the cat.
(2) The cat was chased by the dog.

In contrast, sentences (3) and (4), a subject relative clause and an object relative clause, are complex because they are multiclausal (i.e., they both contain a relative clause) and contain displaced NPs, as indicated by the underscore gap:

(3) The dog that __ chased the cat.
(4) The cat that the dog chased __.

Alternatively, complexity can be operationalized with reference to canonical word order relations (e.g., MacDonald & Christiansen, 2002). That is, sentences that contain the canonical Subject-Verb-Object (SVO) English word order where the sentential subject is an agent are less complex than those that contain noncanonical agent-patient relations. On this approach, sentences (1) and (3) above would be considered simple because they have canonical SVO word order where the subject and object map directly to the thematic roles of agent and patient. In contrast, (2) and (4) are complex because they have patient-first word order.

The present study was thus exploratory, and had two aims: (i) to determine one source of sentence complexity in children aged 4- to 6-years and (ii) to determine the components of WM that support the processing of comparatively easy and difficult structures. Following previous studies that have shown word canonicity to be the major determinant of sentence difficulty (e.g., Bever, 1970; Robertson & Joanisse, 2010; Slobin & Bever, 1982; Townsend & Bever, 2001), we hypothesized that children will experience most difficulty with sentences that contain noncanonical word order. That is to say, we expected that children's performance on sentences with canonical word order would be higher in comparison to their performance on sentences with noncanonical word order. Second, following the majority of past studies investigating WM and child first language acquisition, we hypothesized that different components of WM would support the processing of these different sentence types. Specifically, we hypothesized that the more difficult noncanonical sentences would be supported by central executive processes. In contrast, in line with past findings with typically developing children in studies similar to our own (Montgomery et al., 2008, 2009; Montgomery & Evans, 2009), we did not expect any component of WM to uniquely predict the performance on the comprehension of simple sentences.

Method

Participants

Fifty-six monolingual, typically developing children were recruited from kindergartens and primary schools in regional Australia. Six of these children were excluded from the final analyses because they either were not available for the second session of testing ($n = 2$), had articulation difficulties that were suggestive of language or speech impairment ($n = 3$), or had a developmental delay ($n = 1$). Data from the remaining participants ($n = 50$, 22 female, 28 male) were used in the data analysis. The age of the children in the final sample ranged from 48 months (4;0 years) to 78 months (6;6 years) ($M = 61.69$, $SD = 7.67$).

Materials

A battery of six tasks was used in the study. Three memory tasks were used: one task for each component of Baddeley's (2007) model that in principle is likely to have some involvement in sentence processing (i.e., central executive, phonological STM, and the episodic buffer). Two additional tasks measured general abilities in nonverbal intelligence and vocabulary knowledge, which were included so that any variance they explain in the relationship between memory and syntactic computation could be removed. Finally, the children completed a test of sentence comprehension. Each task is described in turn.

Raven's Colored Progressive Matrices (RCPM)

The RCPM (Raven, Court, & Raven, 1987) was used to measure children's nonverbal ability. Alloway et al. (2004) reported that nonverbal ability was significantly associated with performance on a range of WM measures even when age and demographic variables such as maternal education level are controlled (for associations in adults see Kane, Hambrick, & Conway, 2005). Therefore the RCPM was used so that it could be included as a covariate in our main analyses, ensuring that any association found between WM and language comprehension did not simply reflect a spurious association between nonverbal ability and language. In the task, the child is shown a series of 36 visual patterns that vary in complexity. Each pattern has a piece missing; the child's task is to select the missing piece from a choice of six possibilities, thereby completing the pattern. Children were asked to look carefully at the pattern, and to point to the picture that "fitted best" in the missing space. The children's raw scores were used in the statistical analyses (maximum score = 36). In a large Australian study by Reddington and Jackson (1981) the RCPM are reported to have a Cronbach's α reliability of $\alpha = .80$ in 5-and-a-half-year-olds. More recently, Cotton et al. (2005) examined the psychometric properties of the RCPM among Australian primary-school children, finding that it had good internal consistency and split-half reliability.

British Picture Vocabulary Scale-II (BPVS- II)

The second test of children's general abilities was the BPVS- II (Dunn, Dunn, Whetton, & Burley, 1997). A measure of verbal ability was considered an essential addition to the study; because we were interested in how WM supported the comprehension of sentences that differed in structure, including a measure of vocabulary knowledge allowed us to remove any variance vocabulary level may explain in syntactic processing. This ensured that any significant positive association found between WM and sentence comprehension could then be interpreted to mean that WM at least partially supports syntactic computation.

The BPVS-II is suitable for use for English-speaking children aged from 3 to 15 years. In this test children are orally presented with a word (i.e., *ladder*) and are asked to identify the picture that matches the word from an array of four. The test consists of 14 blocks of 12 items (maximum score = 168); testing discontinues when children make eight or more errors in any one block. The children's raw scores were used in the statistical analyses as a covariate. The authors of this scale report a Cronbach's α internal consistency reliability of $\alpha = .96$ for preschoolers (3 to 5 years) and $\alpha = .94$ for children in their first year of school (5 to 6 years).

Phonological STM: Children's Test of Nonword Repetition (CNRep)

The CNRep (Gathercole, Willis, Baddeley, & Emslie, 1994) was used to measure phonological STM. The CNRep is intended for use with children aged between 4 and 8 years. The test consists of 40 nonsense words with an equal number of two-, three-, four-, and five-syllable words. The CNRep is a well-attested and well-validated measure of phonological STM (see Alloway et al., 2004; Gathercole et al., 2004). Aspects of the test depend on lexical knowledge, for instance, subparts of the test correspond to words or morphemes of English (e.g., *pen, -ing*), the items follow the prosodic contour of English words of equivalent length, and some of the stimuli contain consonant clusters. Such features were not considered to be problematic for the current research, because, although successful performance might involve lexical mediation (Archibald & Gathercole, 2006), the inclusion of a vocabulary measure as a covariate ensured that any lexical involvement could be partialled out.

Children were asked to wear a set of headphones with an attached microphone. The test was prerecorded by a male speaker of Australian English so that it would be suitable for our sample, following the instructions from the test manual. The children were told that they would hear a man talking, who would say some "funny, pretend words—not real words, made-up words," and that they should listen carefully and then try to copy exactly what the man has said. Gathercole et al. (1994) report a test-retest reliability of $r = 0.77$ for 5-year-olds. The children's raw scores were used in the analyses (maximum score = 40).

Episodic Buffer Measure: Sentence Repetition (SRep)

A SRep task was used to measure the episodic buffer component of WM. We followed Alloway et al. (2004) in the choice of SRep as a measure of the episodic buffer because the task involves the integration of information from phonological STM with long-term linguistic knowledge into chunks of linguistic information that should be available to conscious awareness. Alloway et al. (2004) provided empirical evidence for the existence of a separate episodic buffer component in children aged 4 to 6 years using sentence repetition. They tested 633 children on multiple measures of phonological STM, the episodic buffer (two types of sentence repetition), and the central executive. Using confirmatory factor analysis, they showed that the Baddeley (2000) WM model best fitted the data. Crucially, they showed that the inclusion of sentence repetition as a measure of the episodic buffer significantly increased the model's fit to the data, providing strong empirical evidence for: (i) the existence of the episodic buffer in children aged 4 to 6 years, and (ii) the use of sentence repetition as a measure of this construct. Rohl and Pratt (1995) have also reported that sentence repetition constitutes a statistically different construct to complex verbal WM span (i.e., the central executive). Specifically, repeating sentences involves holding the serial order of words in verbatim memory, accessing their syntactic and semantic information from LTM and reassembling them given the syntactic constraints of the language, implicating the language processing system (for evidence that sentence repetition taps into language processing ability, see Potter & Lombardi, 1990, 1998). Moreover, memory for sentences has been consistently shown to be superior to memory for lists of unrelated words (the so-called sentence superiority effect), a process that is argued to reflect chunking (Baddeley, Hitch, & Allen, 2009; Miller, 1956). Finally, recent experimental and computational work presented by Cowan, Rouder, Blume, and Scott Saults (2012) has convincingly argued that this chunking mechanism interfaces with an activated LTM component that is unlimited in capacity. As such, our choice of SRep is also consistent with the most recent conceptualization of the episodic buffer as a multidimensional passive store that holds chunks of information derived from LTM.

The current study used the sentence repetition subtest from the Clinical Evaluation of Language Fundamentals—Preschool 2, Australian (Wiig, Secord, & Semel, 2001), which is suitable for children aged 3;0 to 6;11. The task comprises 13 sentences, ranging in length from 3 words (3 syllables) to 13 words (17 syllables). On the first two (easiest) sentences children can score a total of two points for each; they received two points for a verbatim repetition, one point for a repetition that contained one error, and zero points if they made two or more errors. On the remaining 11 sentences children can score three points (verbatim repetition), two points (one error), one point (for two to three errors), and zero points if four or more errors are made. The sentence represented a range of structural types, including intransitives (2), active transitives (4), passives (4), subject relatives (2), and adverbial clauses.[2] Children are required to listen to sentences and immediately repeat each as they remember them. The task was coded online, as the children repeated the sentence, and was also audio recorded so that the online coding could be checked against the recordings. Any discrepancies between the two were corrected accordingly. Children's raw scores were used in the analyses (maximum score = 37).

Central Executive Measure: Backwards Digit Span (BDig)

A Bdig task was used as a measure of central executive functioning. The task was a subtest of the Working Memory Test Battery for Children (Pickering & Gathercole, 2001). The BDig task requires children to listen to a series of digits and recall it in backwards order. For example, if the child hears the series "1,4,7," they need to repeat "7,4,1." The necessity to both retain, manipulate, and recall the number sequence makes this task a suitable measure of central executive functioning (Baddeley, 2007). Gathercole et al. (2004) report a split-half reliability for the backwards digit span task of $r = .83$ in their sample of 46 6- to 11-year-olds.

To ensure children understood what was meant by "backwards," picture training aids were used to demonstrate and explain the concept before undertaking test items. As is specified in the task manual, two practice digit sequences were administered along with feedback and encouragement before moving on to the test items. These practice sequences of two numbers (e.g., 2, 3) were followed by test items. The test items are grouped into blocks depending on the number of digits in each item, beginning with sets of two and becoming increasingly longer as the test progresses, up to a maximum of six digits. Testing discontinues when children make three errors on sequences of the same block of items (total 6 items per block). Children are scored for the number of sequences that they correctly recall. Items were presented by a female researcher in an even monotone, and at the rate of one digit per second. The test was considered the most appropriate measure of Central Executive for the current study, because Gathercole et al. (2004) reported that 4- and 5-year-old children found the task demands for alternative measures (i.e., Listening Span & Counting Recall) too difficult.

Sentence Comprehension Task

The final measure used in this study was designed to assess children's ability to comprehend a range of sentences that varied in complexity. In all, 64 test sentences were formulated for the task: 16 active sentences, 16 passives, 16 subject relative clauses, and 16 object relative clauses. Each child was tested on half of these sentences (i.e., 8 of each sentence type), which were chosen for each child using a Latin-square design. There were 16 fillers (intransitive sentences such as *The girl is sitting*). The sentences were chosen in order to test

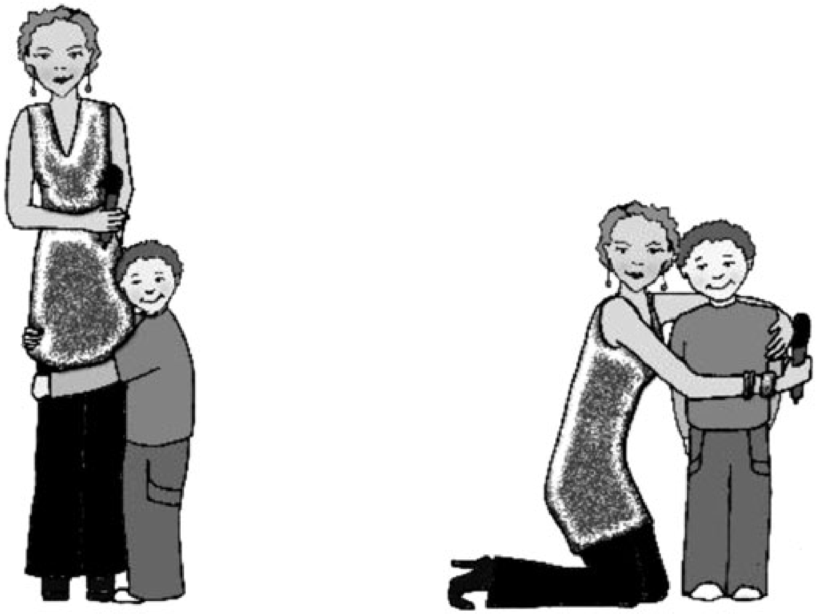

Figure 1 Example of the reversible picture set used for the verb *hug*.

the two possible sources of sentence complexity outlined in the introduction. On the one hand, complexity (and processing difficulty) could be attributable to the traditional linguistic notion of subordination, which in the case of relative clauses also means that nouns are displaced and co-referential with missing elements within the relative clause. Such an account predicts that relative clauses are more difficult than simple monoclausal sentences. Alternatively, complexity (and processing difficulty) could be attributable to noncanonical forms, in which case passives and object-relative clauses are predicted to be more difficult than actives and subject-relative clauses.

Eight verbs were used in total (*wake, comb, bang, push, hug, kiss, follow,* and *splash*), which were chosen because they can be used to depict reversible actions and are familiar to young children (see Figure 1). For any one child each verb was used once only in each sentence type (i.e., once each in an active, passive, etc.); thus each verb was used 4 times for each child. Given that the NP animacy has been shown to affect comprehension (e.g., Brandt, Kidd, Lieven, & Tomasello, 2009; Corrêa, 1995; Goodluck & Tavakolian, 1982), all test items contained animate nouns only.

A picture-pointing task was used. For each test sentence there were two test pictures, one that depicted the sentence and another that depicted the reverse (see Figure 1). Each picture was A4 size, in color. To ensure that any variance in responses was not due to sentence-length effects, both simple and complex sentences in each set had the same average number of syllables across the entire list of 64 items. Sentence length ranged between 8 and 12 syllables, depending on the picture pair. The combined number of syllables in the two canonical sentences was equal to that of the noncanonical sentences. For instance, for Figure 1 the test items were: Active: *The little boy is hugging the mum*, Passive: *The mum is being hugged by the boy*, Subject Relative: *The nice boy that is hugging the mum*, Object Relative: *The mum that the nice boy is hugging*.[3] In this instance all sentences have 9 syllables. Sentences were matched in length by the addition of adjectives that are well known to children of this age (e.g., *little, nice*).

Children were randomly assigned to receive one of four sets of sentences. The task was divided into two halves, which were presented to children over two testing sessions. The experimenter showed the children the two pictures, and then read out the test sentence in a slow and deliberate manner, taking care to ensure that there were no prosodic breaks that might provide a cue to interpretation. Children were asked to listen to each sentence and to point to the picture they believed matched the sentence. The location of the target picture (i.e., left or right) and the position of the agent and patient were counterbalanced across all lists.

Procedure

Children whose parents had given written informed consent were invited to take part in some "remembering and matching activities." Children who agreed to be involved were tested over two to three sessions, each lasting between 15 and 30 minutes. Each child's subsequent testing session(s) were conducted within 2 weeks of the initial testing session. The children completed all six tasks, with the sentence comprehension task being split into two equal parts, each presented in separate sessions. In most cases, four tasks were presented to children in Session A (sentence comprehension part 1, CNRep, RCPM, and BDig) and three tasks in Session B (sentence comprehension part 2, SRep, and BPVS). The within-session order of tasks was counterbalanced. Half of the participants were first tested on the testing schedule in Session A, and half were first tested on the testing schedule in Session B. An age-appropriate explanation and several practice examples were given to children before the beginning of

Table 2 Percentage correct (and *SD*s) for each sentence type on the comprehension task

	Simple	Complex	Total
	Active	*Subject RC*	
Canonical	91 (28.7)	92 (27.1)	91.5 (27.9)
	Passive	*Object RC*	
Non-canonical	79 (40.8)	58.9 (49.3)	69.1 (46.2)
Total	85 (35.7)	76 (42.7)	80.5 (39.6)

Note. RC = relative clause.

Table 3 Descriptive statistics for general ability and working memory measures

	Mean	SD	Min-Max	Skewness	Kurtosis
Vocab[a]	42.52	13.07	19–68	0.15	−0.99
NV IQ[b]	13.56	5.37	4–30	0.65	0.76
Ph. STM[c]	23.22	5.32	12–33	−0.13	−1.02
Ep. Buffer[d]	21.68	7.05	7–34	−0.34	−0.78
Cent. Exec.[e]	8.36	5.53	0–20	0.30	−0.56

[a]test = BPVS-II, Max Score = 168, [b]test = RCPM, Max Score = 36, [c]CNRep, Max Score = 40, [d]test = Sentence Repetition, Max Score = 37, [e]test = BDig, Max Score = 42, Min-Max = Minimum-Maximum score.

each task to ensure that they understood the procedure. The CNRep and SRep tasks were audio recorded to ensure accuracy of coding.

Results

The children's responses in the language comprehension task were marked as either correct or incorrect. A correct answer required the child to select the correct picture. The mean percentage of correct answers (and SDs) for each sentence type are provided in Table 2.

Table 2 shows that the children performed similarly on the simple active sentences and subject relative clauses (RCs). Their performance was reduced on passive and object RCs, markedly so for the latter. Table 3 displays the descriptive statistics for the general ability and verbal WM measures. One child refused to complete the CNRep; her score was replaced with the mean. All variables were normally distributed.

Pearson's bivariate correlations were used to analyze relationships between the three measures of WM ability and the two general ability measures. Partial correlations between the three WM measures that controlled for age, nonverbal

Table 4 Bivariate (lower triangle) and partial correlations (upper triangle) between age, nonverbal IQ, vocabulary, and WM measures

	1	2	3	4	5	6
1. Age	—					
2. NV IQ	.33*	—				
3. Vocab.	.61**	.36**	—			
4. Ph. STM	.49**	.35*	.49**	—	.355*	.001
5. Ep. Buffer	.61**	.36*	.71**	.58**	—	.433**
6. Cent. Exec.	.61**	.32*	.67**	.36*	.73**	—

Note. $^{*}p < .05$, $^{**}p < .01$ level. Partial correlations between WM measures after controlling for age, nonverbal IQ, and vocabulary.

IQ, and vocabulary were also conducted. These correlations are shown in Table 4.

Table 4 shows simple correlations in the lower triangle and partial correlations in the upper triangle. The simple correlations revealed low to moderate statistically significant relationships between almost all variables. The partial correlations between the three WM variables are lower than the simple correlations, suggesting that the three measures tap into at least partially distinct cognitive constructs once variance associated with age, nonverbal ability, and vocabulary knowledge was removed.

We next conducted our main analysis. Because the dependent measure was a categorical response variable, the data were analyzed using Generalized Linear Mixed Models (Baayen, Davidson, & Bates, 2008; Jaeger, 2008), which were calculated using the *lme4* package for Linear Mixed Effects (Bates & Maechler, 2010) in *R* (version 2.14.2, R Core Development Team, 2008). The independent variables were: (i) Complexity, (ii) Canonicity, (iii) Phonological STM, (iv) Episodic Buffer, and (v) Central Executive. For the three continuous WM variables we removed variance associated with age, nonverbal IQ, and vocabulary by regressing these variables onto each WM measure and saving the standardized residuals using simple linear regression. The (zero-centred) standardized residuals were used in the analysis. This enabled us to control for these variables as well as reduce the potential complexity of our model. Only two-way interactions that involved either Complexity or Canonicity and one WM measure, and three-way interactions involving both Complexity and Canonicity and one WM measure were included in each model.[4] Participants and items were treated as random effects in order to accommodate by-participant and by-item variation in one model. By-participant and by-item random slopes were

Table 5 Final model predicting sentence comprehension

	Estimate	Std. Error	Wald-*Z*	*p*-value
Intercept	2.85	.26	10.99	$<.001^{***}$
Complexity	−.04	.31	−.129	.90
Canonical	−1.17	.29	−4.05	$<.001^{***}$
Ph. STM	−.12	.29	−.40	.69
Ep. Buffer	.14	.31	.46	.64
Cent. Exec.	.39	.29	1.36	.17
Complexity:Canonicity	−1.24	.35	−3.59	$<.001^{***}$
Complexity:Ph. STM	.03	.36	.09	.93
Canonical:Ph. STM	−.003	.322	−.01	.99
Complexity:Ep. Buffer	.08	.36	.23	.82
Canonical:Ep. Buffer	.77	.34	2.22	$.026^{*}$
Complexity:Cent. Exec.	−.20	.35	−.59	.55
Canonical:Cent. Exec.	−.18	.31	−.57	.57
Comp:Canon:Ph. STM	−.41	.39	−1.06	.29
Comp:Canon:Ep. Buffer	−.62	.4	−1.55	.12
Comp:Canon:Cent. Exec.	−.13	.37	−.36	.72

Note. $^{***}p < .001$, $^{*}p < .05$. log likelihood = −657.1. IVs: Sentence Complexity, Canonicity, Phonological STM, the Episodic Buffer, and the Central Executive Components of WM. Reference Levels (intercept terms) for Fixed-Effects Predictors: Complexity: Simple; Canonicity: Canonical.

also included to ensure that the effects observed for the fixed-effects predictor variables reflected the slopes for these effects and not between-participant and between-item variance (Baayen et al., 2008; Barr, 2008). A series of models were run and compared using the *anova* function in *R*. Including items as a random effect and by-item random slopes did not significantly improve the model, but including by-participant random slopes significantly improved model fit. The results from the model that best fit the data are shown in Table 5 (see the Appendix for *R* code).

There were three significant effects. First, the children performed significantly better on canonical than noncanonical sentences (estimate = −1.17, $z = -4.05$, $p < .001$). Second, the Canonicity × Complexity interaction was significant (coefficient = −1.24, $z = -3.59$, $p < .001$), which reflected the fact that children performed comparatively worse on object RCs in comparison to subject RCs than the same comparison between passives and actives. Finally, the Canonicity × Episodic Buffer interaction was significant (coefficient = 0.77, $z = 2.22$, $p = .026$). This reflected the fact that higher scores on the measure of the Episodic Buffer were associated with higher performance on the noncanonical sentences.

The results therefore suggest that only the episodic buffer is significantly implicated in sentence comprehension; specifically, in the comprehension of noncanonical sentences. Because our analyses were quite different from those conducted in previous studies, we decided to run some additional models so that more direct comparisons could be made. In particular, we ran a series of simple models where we used Phonological STM, the Episodic Buffer, and the Central Executive measures to predict noncanonical sentence comprehension. These are shown in Appendix S1 of the Supporting Information online. Consistent with past research that has not included a measure of the episodic buffer, children's performance on the central executive task significantly predicted comprehension when no Episodic Buffer measure was included in the analysis (coefficient $= .24$, $z = 2.1$, $p = .036$). However, once the Episodic Buffer measure was included in the model the contribution of the Central Executive measure was no longer significant (coefficient $= -.002, z = -.02, p = .98$), whereas the contribution of the Episodic Buffer measure was significant (coefficient $= .51, z = 4.16, p < .001$). For full details see Appendix S1 in the online Supporting Information.

Discussion

Our first hypothesis was supported. Following similar investigations into sentence complexity (e.g., Bever, 1970; Robertson & Joanisse, 2010; Slobin & Bever, 1982), our participants experienced most difficulty comprehending sentences with noncanonical word order. This effect reflects the well-documented preference for children (and adults) to process canonical word order (Bates & MacWhinney, 1982; Ferriera, 2003; MacDonald & Christiansen, 2002; Slobin & Bever, 1982; Townsend & Bever, 2001). This result is typically explained by appealing to the frequency of sentence frames. That is, because in English canonical SVO word order is also the most frequent configuration, the induction and application of a Noun-Verb-Noun sentence schema which maps onto canonical agent-patient relations makes active transitives and subject RCs easiest to process. While children performed equivalently on both canonical sentence types, the complexity by canonicity interaction showed that they performed better on passives than on object RCs. This is likely to be due to morphological differences across these two sentence types. The English full BE passive used in the current study had two morphological cues to interpretation: (i) the present progressive auxiliary *being*, and (ii) the *by*-phrase. In contrast, the object RCs contained only word order cues to interpretation. Children's relative difficulty with object RCs is likely to be due to this comparative lack of surface cues to grammatical role assignment, and the fact that object RCs are likely to be initially analyzed as subject RCs until children hear the second noun (e.g., *The man that the woman*. . .). Children at this age are typically poor at reanalyzing initial parses (Kidd et al., 2011; Trueswell et al., 1999), which is likely to explain their poorer performance on object RCs.

Our predictions regarding the role of WM in sentence comprehension were only partially borne out. The comprehension of the more difficult noncanonical sentences was not supported by the central executive component of WM, but by the episodic buffer. This is inconsistent with studies that have reported central executive involvement in sentence comprehension (e.g., Booth et al., 2000; Felser et al., 2003; Montgomery et al., 2008, 2009). However, these studies did not include a measure of the episodic buffer; our analyses showed that the central executive was implicated in comprehension if the measure of the episodic buffer was removed from an analysis that best approximated analyses reported in prior studies (see Appendix S1 in the online Supporting Information). Therefore, while we were able to replicate past results, our more comprehensive battery of WM measures yielded a different result. The significant contribution of the episodic buffer to (spoken) language comprehension is consistent with the preliminary data presented in Alloway and Gathercole (2005), who observed that the episodic buffer, as measured by sentence repetition, predicted performance on a standardized measure of spoken language comprehension and production in a sample of children with learning difficulties. As in the current study, Alloway and Gathercole also measured the central executive, but did not find that it significantly predicted spoken language.[5]

We found that the episodic buffer was only implicated in the processing of noncanonical sentences, narrowing down the effect observed by Alloway and Gathercole (2005). This result raises some important questions for the traditionally held belief that language comprehension is supported by a capacity-limited processing store that supports the storage and analysis of linguistic stimuli (e.g., Just & Carpenter, 1992). In acquisition research there has been a general assumption that capacity-limit models are in principle correct; research efforts have been concerned with establishing that WM predicts individual differences in language comprehension in developmental populations (see Table 1). Capacity limits almost certainly affect language processing, because spoken language is a serial auditory signal that must be parsed in such a way that dependencies between lexical elements can be established. These processes are likely to cause a processing bottleneck, which is formalized in models such as Gibson's (2000) Dependence Locality Theory, where sentence complexity is operationalized according to the resource cost associated with storing and integrating verbs and their arguments. Our data suggest that the chunking mechanism characteristic of the episodic buffer could be involved in the temporary storage of unassigned arguments, although our finding that the episodic buffer is implicated equally in the comprehension of both passives and object RCs is not entirely consistent with this suggestion, given that a passive contains no incomplete dependencies.

An alternative explanation is that these data say more about the children's long-term linguistic knowledge than they say about capacity limits. As shown by Cowan et al. (2012), models of verbal WM must include access to a capacity-unlimited LTM store in order to be explanatory. How WM interfaces with long-term knowledge of language is a rarely considered topic, but one which our episodic buffer by canonicity interaction forces us to tackle head on.

According to Baddeley and colleagues (e.g., Baddeley, Allen, & Hitch, 2010), the episodic buffer is a passive store for information integrated from other memory systems; it integrates information from WM and LTM into meaningful chunks that are available to conscious awareness. In language comprehension the contribution from LTM is likely to be lexico-semantic and syntactic knowledge (Alloway & Gathercole, 2005). The suggestion here is that the contribution of the episodic buffer to language comprehension might not be direct, but might reflect differences in children's long-term linguistic knowledge. Rather than identifying capacity limits as the source of individual differences, this alternative explanation explains individual differences as the outcome of an interaction in: (i) the amount of experience a speaker has with structural patterns and (ii) the endogenous ability to detect structural patterns in the input, the confluence of which ultimately leads to individual differences in attainment.

According to this *experience-dependent* explanation, which is based upon functionalist and connectionist approaches to language (e.g., Bates & MacWhinney, 1982, 1989; Christiansen & MacDonald, 2009; Tomasello, 2003), acquisition is conceptualized as identifying form-function correlations in the input. Frequency of occurrence is argued to be a major driving force behind acquisition, because the more available a structure is in the input the more opportunities children have to learn its formal and functional properties. This emphasis on the frequency of structural patterns naturally explains the general effect of canonicity in our data. Passives and object relatives are low in frequency compared to sentences that have canonical word order. Roland, Dick, and Elman (2007) showed that subject relative clauses are approximately

2.5 times more common in spoken English than are full (i.e., unreduced) object relative clauses. Similarly, Brown (1973) reported that full passives constitute less than 1% of all grammatical forms present in child-directed speech, whereas simple active sentences are commonplace.

However, the experience-dependent approach also predicts that there should be individual differences in a learner's ability to seize upon grammatical patterns in the input (Christiansen & MacDonald, 2009). The canonicity by episodic buffer interaction is consistent with this prediction. That is, those children who scored higher on the measure of the episodic buffer might be more adept at identifying and extracting structural regularities from the input. What might be the mechanism that detects regularities? Recent research suggests that an *implicit statistical learning* mechanism might provide such neurocognitive support for language acquisition and processing.

Implicit statistical learning[6] describes the largely or wholly unconscious process of inducing structure and regularity following exposure to repeated exemplars (Conway & Christiansen, 2006; Perruchet & Pacton, 2006; Reber, 1993). It has been most robustly demonstrated in speech segmentation and artificial grammar learning experiments in infants (see Gómez & Gerken, 2000; Romberg & Saffran, 2010). However, a growing body of research findings implicates an implicit statistical learning mechanism in the acquisition and processing of natural language. For instance, Misyak and Christiansen (2012) have shown that performance on artificial grammar learning tasks that tested participants' ability to learn adjacent and nonadjacent dependencies predicted the comprehension of structures involving these dependency types over and above the influence of complex WM span (see also Conway, Bauernschmid, Huang, & Pisoni, 2010; Wells et al., 2009). Similarly, Kidd (2012) showed that performance on an implicit statistical learning task predicted long-term use of the English passive structure following a syntactic priming task in a sample of typically developing English-speaking children aged 4 to 6 years. These recent studies suggest that a neurocognitive mechanism that tracks the distribution of elements across sequences of stimuli and induces underlying structure is directly associated with language acquisition in children and language processing in adults (see also Chang, Dell, & Bock, 2006; Chang, Janciauskas, & Fitz, 2012; Rebuschat & Williams, 2012). The suggestion we would like to make, based on our present findings, is that the canonicity by episodic buffer interaction might be indicative of underlying differences in children's ability to detect structural regularities via a mechanism like implicit statistical learning.

A crucial question concerns whether sentence repetition, our measure of the episodic buffer, captures individual differences in long-term linguistic knowledge driven by implicit statistical learning. Research by Potter and Lombardi (1990, 1998; Lombardi & Potter, 1992) has shown that sentence repetition involves the regeneration of sentence structure using sentence processing mechanisms, that is, it taps long-term syntactic and lexico-semantic knowledge that governs the comprehension and production of well-formed utterances. In the best articulated theory of implicit learning in language acquisition, Chang et al.'s (2006) model of sentence production, an implicit statistical learning mechanism serves as a sequencing mechanism which maps function onto form. Using a Simple Recurrent Network, it identifies form-function correlations, learning to sequence words into grammatical sentences. In the model, knowledge and process are intricately linked. Namely, an implicit statistical learning mechanism identifies and acquires structures from the input; and the same mechanism implements these constraints on word order during production. Our suggestion here is that the episodic buffer by canonicity effect could reflect more advanced syntactic knowledge in children with high scores on sentence repetition, which might in turn reflect, at least in part, superior implicit statistical learning ability. The net outcome of having superior implicit statistical learning capacity is that one requires less exposure to a grammatical pattern in order to learn it. Thus we see the effect only occurs for low frequency noncanonical forms, which children rarely experience and take much longer to acquire.

Conclusion

The current study points at how WM might be implicated in language acquisition, but the conclusion is different from traditional views. Unlike most past studies but consistent with some recent developmental and adults studies (Alloway & Gathercole, 2005; Misyak & Christiansen, 2012), we did not observe the central executive to be implicated in sentence comprehension. Instead, we found that the episodic buffer supported the comprehension of noncanonical sentences. This result could be interpreted to suggest that the episodic buffer is the source of capacity limits in language processing, but we have also suggested that it might reflect LTM involvement in sentence comprehension, a process supported by implicit statistical learning. Both interpretations are consistent with current models of verbal WM (e.g., Cowan et al., 2012) which rely on an activated (and unlimited) LTM component that contributes to verbal performance. Serious work on the role of implicit statistical learning as a mechanism for language learning has only just begun (Chang et al., 2012; Kidd, 2012), infant studies of statistical learning notwithstanding (Romberg & Saffran, 2010;

Saffran, 2003). Verbal WM and implicit statistical learning capacity are complementary processes; measures of each tap independent constructs (Kaufman et al., 2010). Investigations of how these two processes conspire to support first language acquisition are noticeably absent from the literature. This is an important research priority.

In this sense, our study should only be considered a preliminary investigation into verbal WM and sentence comprehension during language acquisition. There are several avenues for future research. First, the nature of sentence repetition as a measure of the episodic buffer needs further exploration. Sentence repetition is one potential measure of the episodic buffer, but others exist and their relationship to language needs to be explored (e.g., Baddeley, Allen, & Vargha-Khadem, 2010). Second, we need large-scale studies that chart the influence of neurocognitive processes such as WM and implicit statistical learning across broad age ranges in development and into adulthood. Such studies, combined with training studies that actively manipulate children's exposure to different structures (e.g., Vasilyeva, Huttenclocher, & Waterfall, 2006), will enable us to tease apart the influences of these processes on language learning.

Final revised version accepted 24 June 2012

Notes

1 Problematically, the two complex sentence types differed in length. The passive sentences were between 7–9 syllables in length ($M = 8.33$), and the pronominal/reflexive sentences were between 10–14 syllables in length ($M = 11.83$). This difference was significant and of a very large magnitude ($t(34) = 9.13$, $p < .001$, $d = 3.46$).

2 Although some of these sentences are the same structural types that we tested in the comprehension task, a direct comparison is unfortunately not possible, because the sentences in the repetition task also contained a number of additional grammatical devices that served to increase sentence length and complexity, such as NP coordination, negation, genitive NPs, and various types of adjuncts (e.g., adverbials).

3 The relative clause structures were not technically full sentences, which was necessary to control the length all test items. For ease of expression, we continue to refer to them as "sentences" throughout the paper.

4 A comparison between the final model in Table 2 and a full-factorial model including every possible interaction was not significant ($\chi^2 = 15.59$, $df = 16$, $p = .49$).

5 Alloway and Gathercole (2005) did find that the episodic buffer and central executive were *both* implicated in children's reading abilities, suggesting that the

central executive might be more strongly implicated in literacy. This may then explain the very consistent finding that complex verbal WM predicts language comprehension in adults, where comprehension has been typically tested using reading.

6 Although different permutations are used in the literature, *implicit learning*, *statistical learning*, and *implicit statistical learning* are all used to refer to the same phenomenon (Conway et al., 2010; Perruchet & Pacton, 2006). Here we use *implicit statistical learning*. It is a term that partly encompasses *implicit* or *procedural memory* components of LTM. Following Reber (1993), we assume implicit statistical learning and implicit memory to share particular patterns and characteristics; our emphasis on learning reflects our aim to explain language acquisition.

References

Alloway, T. P., & Gathercole, S. (2005). The role of sentence recall in reading and language skills of children with learning difficulties. *Learning and Individual Differences*, *15*, 271–282.

Alloway, T. P., Gathercole, S., & Pickering, S. (2006). Verbal and visuospatial short-term and working memory in children: Are they separable? *Child Development*, *77*, 1698–1716.

Alloway, T. P., Gathercole, S., Willis, C., & Adams, A. (2004). A structural analysis of working memory and related cognitive skills in young children. *Journal of Experimental Child Psychology*, *87*, 85–106.

Archibald, L., & Gathercole, S. (2006). Nonword repetition: A comparison of tests. *Journal of Speech, Language, and Hearing Research*, *49*, 970–983.

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.

Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modelling with crossed random effects for subjects and items. *Journal of Memory and Language*, *59*, 390–412.

Baddeley, A. (2000). The episodic buffer: A new component of working memory? *Trends in Cognitive Sciences*, *4*, 417–423.

Baddeley, A. (2007). *Working memory, thought, and action*. New York: Oxford University Press.

Baddeley, A., Allen, R., & Hitch, G. J. (2010). Investigating the episodic buffer. *Psychologica Beligica*, *50*, 223–243.

Baddeley, A. D., Allen, R. J., & Vargha-Khadem, F. (2010). Is the hippocampus necessary for visual and verbal binding in working memory? *Neuropsychologia*, *48*, 1089–1095.

Baddeley, A. D., Gathercole, S. E., & Papagno, C. (1998). The phonological loop as a language learning device. *Psychological Review*, *105*, 158–173.

Baddeley, A., & Hitch, G. (1974). Working memory. In. G. Bower (Ed.), *The psychology of learning and motivation* (Vol. 8, pp. 47–90). New York: Academic Press.

Baddeley, A. D., Hitch, G. J., & Allen, R. J. (2009). Working memory and binding in sentence recall. *Journal of Memory and Language*, *61*, 438–456.

Barr, D. J. (2008). Analyzing "visual world" eye-tracking data using multilevel logistic regression. *Journal of Memory and Language*, *59*, 457–474.

Barrouillet, P., & Gaillard, V. (Eds.). (2011). *Cognitive development and working memory: A dialogue between neo-Piagetian theories and cognitive approaches.* Hove, UK: Psychology Press.

Bates, E., Devescovi, A., & D'Amico, S. (1999). Processing complex sentences: A crosslinguistic study. *Language and Cognitive Processes*, *14*, 69–123.

Bates, E., & MacWhinney, B. (1982). Functionalist approaches to grammar. In. E. Wanner & L. Gleitman (Eds.), *Language acquisition: The state of the art* (pp. 173–218). New York: Cambridge University Press.

Bates, E., & MacWhinney, B. (1989). Functionalism and the competition model. In B. MacWhinney & E. Bates (Eds.), *The crosslinguistic study of sentence processing* (pp. 3–73). Cambridge, UK: Cambridge University Press.

Bates, D., & Maechler, M. (2010). *lme4: Linear mixed effects models using S4 classes. R package*. Version: 999375–33. Available online at http://lme4.r-forge.r-project.org/

Bever, T. (1970). The cognitive basis for linguistic structures. In J. R. Hayes (Ed.), *Cognition and the development of language* (pp. 279–362). New York: Wiley.

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

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

Brown, R. (1973). *A first language: The early stages*. Cambridge, MA: Harvard University Press.

Chang, F., Dell, G. S., & Bock, K. (2006). Becoming syntactic. *Psychological Review*, *113*, 234–272.

Chang, F., Janciauska, M., & Fitz, H. (2012). Language adaptation and learning: Getting explicit about implicit learning. *Language and Linguistic Compass*, *6*(5), 259–278.

Christiansen, M. H., & MacDonald, M. C. (2009). A usage-based approach to recursion in sentence processing. *Language Learning*, *59* (Suppl. 1), 126–161.

Conti-Ramsden, G. M., Botting, N. F., & Faragher, B. (2001). Psycholinguistic markers for specific language impairment (SLI). *Journal of Child Psychology and Psychiatry*, *42*, 741–748.

Conway, C. M., Bauernschmidt, A., Huang, S. S., & Pisoni, D. B. (2010). Implicit statistical learning in language processing: Word predictability is the key. *Cognition*, *114*, 356–371.

Conway, C. M., & Christiansen, M. H. (2006). Statistical learning within and between modalities: Pitting abstract against stimulus-specific representations. *Psychological Science*, *17*, 905–912.

Corrêa, L. (1995). An alternative assessment of children's comprehension of relative clauses. *Journal of Psycholinguistic Research*, *24*, 183–203.

Cotton, S., Kiely, P., Crewther, D., Thomson, B., Laycock, R., & Crewther, S. (2005). A normative reliability study for the Raven's Coloured Progressive Matrices for primary school aged children from Victoria, Australia. *Personality and Individual Differences*, *39*, 647–659.

Cowan, N. (2005). *Working memory capacity*. Hove, UK: Psychology Press.

Cowan, N., Rouder, J. N., Blume, C. L., & Scott Saults, J. (2012). Models of verbal working memory capacity: What does it take to make them work? *Psychological Review*, *119*, 480–499.

Dunn, L., Dunn, L., Whetton, C., & Burley, J. (1997). *British picture vocabulary scale* (2nd ed.). Windsor, UK: NFER-Nelson.

Ericsson, K. A., & Kintsch, W. (1995). Long-term working memory. *Psychological Review*, *102*, 211–245.

Felser, C., Marinis, T., & Clahsen, H. (2003). Children's processing of ambiguous sentences: A study of relative clause attachment. *Language Acquisition*, *11*, 127–163.

Fernández, E. (2003). *Bilingual sentence processing: Relative clause attachment in English and Spanish.* Amsterdam: John Benjamins.

Ferreira, F. (2003). The misinterpretation of non-canonical sentences. *Cognitive Psychology*, *47*, 164–203.

Fodor, J. D. (1998). Learning to parse? *Journal of Psycholinguistic Research*, *27*, 285–319.

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

Gathercole, S. E. (2006). Nonword repetition and word learning: The nature of the relationship. *Applied Psycholinguistics*, *27*, 513–543.

Gathercole, S., Pickering, S., Ambridge, B., & Wearing, H. (2004). The structure of working memory from 4 to 15 years of age. *Developmental Psychology*, *40*, 177–190.

Gathercole, S. E., Willis, C., & Baddeley, A. D. (1992). Phonological memory and vocabulary development during the early school years: Evidence from a longitudinal study. *Developmental Psychology*, *28*, 887–898.

Gathercole, S., Willis, C., Baddeley, A., & Emslie, H. (1994). The children's test of nonword repetition: A test of phonological short-term memory. *Memory*, *2*, 103–127.

Gómez, R. L., & Gerken, L. (2000). Infant artificial language learning and language acquisition. *Trends in Cognitive Sciences*, *4*, 178–186.

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

Goodluck, H., & Tavakolian, S. (1982). Competence and processing in children's grammar of relative clauses. *Cognition*, *11*, 1–27.

Jaeger, T. F. (2008). Categorical data analysis: Away from ANOVAs (transformation or not) and towards logit mixed models. *Journal of Memory and Language*, *59*, 434–446.

Just, M., & Carpenter, P. (1992). A capacity theory of comprehension: Individual differences in working memory. *Psychological Review*, *99*, 122–149.

Kane, M. J., Hambrick, D. Z., & Conway, A. R. A. (2005). Working memory capacity and fluid intelligence are strongly related constructs: Comment on Ackerman, Beier, and Boyle (2004). *Psychological Bulletin*, *131*, 66–71.

Kaufman, S. B., DeYoung, C. G., Gray, J. R., Jiménez, L., Brown, J., & Mackintosh, N. (2010). Implicit learning as an ability. *Cognition*, *116*, 321–340.

Kidd, E. (2012). Implicit statistical learning is directly associated with the acquisition of syntax. *Developmental Psychology*, *48*, 171–184.

Kidd, E., & Bavin, E. L. (2002). English-speaking children's understanding of relative clauses: Evidence for general-cognitive and language-specific constraints on development. *Journal of Psycholinguistic Research*, *31*, 599–617.

Kidd, E., Stewart, A. J., & Serratrice, L. (2011). Children do not overcome lexical biases where adults do: The role of the referential scene in garden-path recovery. *Journal of Child Language*, *38*, 222–234.

Logie, R. H. (2011). The functional organization and capacity limits of working memory. *Current Directions in Psychological Science*, *20*, 240–245.

Lombardi, L., & Potter, M. C. (1992). The regeneration of syntax in short term memory. *Journal of Memory and Language*, *31*, 713–733.

Lust, B., Foley, C., & Dye, C. (2009). The first language acquisition of complex sentences. In E. Bavin (Ed.), *The Cambridge handbook of child language* (pp. 237–257). New York: Cambridge University Press.

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

MacWhinney, B. (1999). The emergence of language from embodiment.In B. MacWhinney (Ed.), *The emergence of language* (pp. 213–256). Mahwah, NJ: Erlbaum.

Marshall, C., & Nation, K. (2003). Individual differences in semantic and structural errors in children's memory for sentences. *Educational and Child Psychology*, *20*, 7–18.

Matthews, D., Lieven, E., Theakson, A., & Tomasello, M. (2009). Pronoun co-referencing errors: Challenges for generativist and usage-based accounts. *Cognitive Linguistics*, *20*, 599–626.

Miller, G. A. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review*, *63*, 81–97.

Misyak, J. B., & Christiansen, M. H. (2012). Statistical learning and language: An individual differences study. *Language Learning*, *62*, 302–331.

Montgomery, J., & Evans, J. (2009). Complex sentence processing and working memory in children with specific language impairment. *Journal of Speech, Language a Hearing Research*, *52*, 269–288.

Montgomery, J., Evans, J., & Gillam, R. (2009). Relation of auditory attention on complex sentence comprehension in children with specific language impairment: A preliminary study. *Applied Psycholinguistics*, *30*, 123–151.

Montgomery, J., Magimairaj, B., & O'Malley, M. (2008). The role of working memory in typically developing children's complex sentence comprehension. *Journal of Psycholinguistic Research*, *37*, 331–354.

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

O'Grady, W. (2005). *Syntactic carpentry: An emergentist approach to syntax*. Mahwah, NJ: Erlbaum.

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: John Benjamins.

Perruchet, P., & Pacton, S. (2006). Implicit learning and statistical learning: One phenomenon, two approaches. *Trends in Cognitive Sciences*, *10*, 233–238.

Potter, M. C., & Lombardi, L. (1990). Regeneration in short-term recall of sentences. *Journal of Memory and Language*, *29*, 633–654.

Potter, M., & Lombardi, L. (1998). Syntactic priming in immediate recall of sentences. *Journal of Memory and Language*, *38*, 265–282.

Pickering, S., & Gathercole, S. (2001). *Working memory test battery for children*. London: Psychological Corporation.

R Core Development Team (2008). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing.

Raven, J., Court, J., & Raven J. (1987). *The raven's coloured matrices*. Oxford, UK: Oxford Psychologists Press.

Reber, A. S. (1993). *Implicit learning and tacit knowledge: An essay on the cognitive unconscious*. Hove, UK: Psychology Press.

Rebuschat, P., & Williams, J. N. (Eds.). (2012). *Statistical learning and language acquisition*. Berlin: Mouton de Gruyter.

Reddington, M., & Jackson, K. (1981). The Raven's coloured progressive matrices (1956): A Queensland standardization. *Australian Council for Educational Research Bulletin for Psychologists*, *30*, 20–26.

Rizzi, L. (1990). *Relativized minimality*. Cambridge, MA: MIT Press.

Robertson, E., & Joanisse, M. (2010). Spoken sentence comprehension in children with dyslexia and language impairment: The role of syntax and working memory. *Applied Psycholinguistics*, *31*, 141–165.

Rohl, M., & Pratt, C. (1995). Phonological awareness, verbal working memory and the acquisition of literacy. *Reading and Writing*, *7*, 327–360.

Roland, D., Dick, F., & Elman, J. (2007). Frequency of basic English grammatical structures: A corpus analysis. *Journal of Memory and Language*, *57*, 348–379.

Romberg, A. R., & Saffran, J. R. (2010). Statistical learning and language acquisition. *Wiley Interdisciplinary Reviews in Cognitive Science*, *1*, 906–914.

Saffran, J. R. (2003). Statistical language learning: Mechanisms and constraints. *Current Directions in Psychological Science*, *12*, 110–114.

Shah, P., & Miyake, A. (1999). Models of working memory. In A. Miyake & P. Shah (Eds.), *Models of working memory: Mechanisms of active maintenance and executive control* (pp. 1–27). New York: Cambridge University Press.

Slobin, D., & Bever, T. (1982). Children use canonical sentence schemas: A cross-linguistic study of word order and inflections. *Cognition*, *12*, 229–265.

Swanson, H. L. (1996). Individual and age related differences in children's working memory. *Memory and Cognition*, *24*, 70–82.

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

Townsend, D., & Bever, T. (2001). *Sentence comprehension: The integration of habits and rules*. Cambridge, MA: MIT Press.

Trueswell, J. C., Sekerina, I., Hill, N. M., & Logrip, M. L. (1999). The kindergartenpath effect: Studying on-line sentence processing in young children. *Cognition*, *73*, 89–134.

Trueswell, J. C., Tanenhaus, M. K., & Garnsey, S. (1994). Semantic influences on parsing: Use of thematic role information in syntactic ambiguity resolution. *Journal of Memory and Language*, *33*, 285–318.

Vasilyeva, M., Huttenlocher, J., & Waterfall, H. (2006). Effect of language intervention on syntactic skill levels in preschoolers. *Developmental Psychology*, *42*, 164–174.

Waters, G., & Caplan, D. (1996). The capacity theory of sentence comprehension: Critique of Just and Carpenter (1992). *Psychological Review*, *103*, 761–772.

Wells, J., Christiansen, M., Race, D., Acheson, D., & MacDonald, M. (2009). Experience and sentence processing: Statistical learning and relative clause comprehension. *Cognitive Psychology*, *58*, 250–271.

Wiig, E., Secord, W., & Semel, E. (2001). *Clinical evaluation of language fundamentals—preschool second edition (CELF Preschool 2)*. Toronto, Canada: The Psychological Corporation/A Harcourt Assessment Company.

Willis, C., & Gathercole, S. (2001). Phonological short-term memory contributions to sentence processing in young children. *Memory*, *9*, 349–363.

Appendix

R Code for Final Model

```
Final Model. lmer(Response ~ Complexity + Canonicity +
Complexity:Canonicty + Ph. STM + Ep. Buffer + Cent. Exec. +
Complexity:Ph. STM + Canonicty:Ph STM + Complexity:Ep. Buffer +
Canonicity:Ep. Buffer + Complexity:Cent. Exec. + Canonicty:Cent.
Exec. + Complexity:Canonicty:Ph. STM + Complexity:Canonicty:Ep.
Buffer + Complexity:Canonicty:Cent. Exec.)+(1|participants) + (1+
Complexity|participants) + (1+Canonicity|participants), data=data,
family=”binomial”).
```

Supporting Information

Additional Supporting Information may be found in the online version of this article at the publisher’s website:

Appendix S1. Additional *R* Models Pursuing Approximated Analyses Reported in Prior Studies.