



Reading span task performance, linguistic experience, and the processing of unexpected syntactic events

Thomas A. Farmer, Alex B. Fine, Jennifer B. Misyak & Morten H. Christiansen


To cite this article: Thomas A. Farmer, Alex B. Fine, Jennifer B. Misyak & Morten H. Christiansen (2017) Reading span task performance, linguistic experience, and the processing of unexpected syntactic events, *The Quarterly Journal of Experimental Psychology*, 70:3, 413-433, DOI: [10.1080/17470218.2015.1131310](https://doi.org/10.1080/17470218.2015.1131310)

To link to this article: <http://dx.doi.org/10.1080/17470218.2015.1131310>



Accepted author version posted online: 11 Dec 2015.
Published online: 16 Mar 2016.



[Submit your article to this journal](#) 



Article views: 212



[View related articles](#) 



[View Crossmark data](#) 

Reading span task performance, linguistic experience, and the processing of unexpected syntactic events

Thomas A. Farmer^{a,b}, Alex B. Fine^c, Jennifer B. Misyak^d and Morten H. Christiansen^e

^aDepartment of Psychological and Brain Sciences, University of Iowa, Iowa City, IA, USA; ^bDepartment of Linguistics, University of Iowa, Iowa City, IA, USA; ^cDepartment of Psychology, The Hebrew University of Jerusalem, Jerusalem, Israel; ^dBehavioural Science Group, Warwick Business School, University of Warwick, Coventry, UK; ^eDepartment of Psychology, Cornell University, Ithaca, NY, USA

ABSTRACT

Accounts of individual differences in online language processing ability often focus on the explanatory utility of verbal working memory, as measured by reading span tasks. Although variability in reading span task performance likely reflects individual differences in multiple underlying traits, skills, and processes, accumulating evidence suggests that reading span scores also reflect variability in the linguistic experiences of an individual. Here, through an individual differences approach, we first demonstrate that reading span scores correlate significantly with measures of the amount of experience an individual has had with written language (gauged by measures that provide “proxy estimates” of print exposure). We then explore the relationship between reading span scores and online language processing ability. Individuals with higher reading spans demonstrated greater sensitivity to violations of statistical regularities found in natural language—as evinced by higher reading times (RTs) on the disambiguating region of garden-path sentences—relative to their lower span counterparts. This result held after statistically controlling for individual differences in a non-linguistic operation span task. Taken together, these results suggest that accounts of individual differences in sentence processing can benefit from a stronger focus on experiential factors, especially when considered in relation to variability in perceptual and learning abilities that influence the amount of benefit gleaned from such experience.

ARTICLE HISTORY

Received 14 January 2015
Accepted 2 November 2015
First Published Online 16
March 2016



KEYWORDS

Language comprehension;
Language processing;
Individual differences;
Linguistic experience;
Reading span tasks

Considerable variability is frequently observed in measures of online sentence processing ability (e.g., Friedman & Miyake, 2004; King & Just, 1991; Kuperman & Van Dyke, 2011; MacDonald, Just, & Carpenter, 1992; Novick, Thompson-Schill, & Trueswell, 2008; Pearlmutter & MacDonald, 1995; Swets, Desmet, Hambrick, & Ferreira, 2007; Van Dyke, Johns, & Kukona, 2014—see Farmer, Misyak, & Christiansen, 2012, for a review), although both the sources and the nature of this documented variability are not well understood. Fast and accurate interpretation of an unfolding linguistic signal is made possible through the coordination of

multiple cognitive, perceptual, and motoric processes. Some or all of these processes are likely to vary across individuals for a variety of developmental, environmental, genetic, and other reasons. This observation strongly suggests that indices of processing difficulty elicited during online sentence comprehension tasks will necessarily reflect variability that stems from many different sources.

Historically, however, individual differences in online language comprehension have been attributed most heavily to variability in verbal working memory (vWM) capacity (Caplan & Waters, 1999; Just &

CONTACT Thomas A. Farmer  thomas-farmer@uiowa.edu  Department of Psychological and Brain Sciences, University of Iowa, 11 Seashore Hall E, Iowa City, IA 52242-1407, USA.

The authors would like to thank Caitie Hilliard, Dave Klienschmidt, and Shaorong Yan for useful conversations about the work presented here and for advice on data analysis, and Daniel Plebanek for assistance in data screening and coding. Additionally, we are grateful to Falk Huettig, Sharavan Vasishth, and one anonymous reviewer for detailed and cogent reviews that resulted in a much improved article.

Carpenter, 1992; Waters & Caplan, 1996a). Even in early explicit parsing models (e.g., Frazier & Fodor, 1978; Kimball, 1975), researchers speculated that memory constraints exerted a direct influence on the parsing process, creating pressure on the system to favour structural simplicity. Individual differences in working memory capacity remain a key contributor (if not an imperative, cf. Jackendoff, 2007) to most modern parsing models, from those emphasizing the effects of memory-related principles such as similarity-based interference (e.g., Gordon, Hendrick, & Johnson, 2004; Lewis, 1996; Lewis, Vasishth, & Van Dyke, 2006), to those that highlight the presumed necessity of working memory capacity for maintaining structural information across multiple intervening syntactic units (e.g., Gibson, 1998).

In this paper, we examine a modern measure of vWM capacity and argue that although variability in scores on these measures is likely to reflect individual differences in multiple underlying cognitive processes, some of the variance can be attributed to differences in linguistic experience. We find support for this conclusion by teasing apart the relationships between vWM (as gauged by a language-heavy reading span task), non-linguistic operation span (backward digit span), and the degree to which an individual experiences the garden-path effect. First, we find that two “proxy measures” of linguistic experience correlated positively with vWM, although the non-linguistic operation span did not. Additionally, we find that vWM exhibits a positive relationship with the magnitude of the garden-path effect. This positive relationship indicates that individuals with higher scores on a vWM task are more surprised by encountering a possible but highly unexpected resolution of a temporary syntactic ambiguity. Moreover, this relationship is present even after controlling for variability in an individual’s ability to simultaneously store and process non-linguistic information (as indexed by scores on the backwards digit span task). Taken together, these results suggest that linguistic experience is one determinant of scores on vWM span tasks, and we discuss the implications of this observation for accounts of individual differences in online language processing.

Before presenting our study, we review the background literature on the role of vWM in language processing, thus motivating three main questions that we address with respect to the data reported here.

Working memory in language processing

Working memory, in the Baddeley (1986) tradition, is viewed as a “cognitive workbench” used both for information storage and as the locus of processing. During reading, for example, a person must be able to incorporate incoming input into the developing representation of an author’s intended message as conveyed through previously encountered text. Early examinations of the relationship between working memory capacity and language comprehension abilities focused on links between memory capacity and scores on offline measures of language comprehension. Individual differences in memory were often gauged with digit span tasks that required memorization of increasingly longer lists of digits. The Backwards Digit Span task (Wechsler, 1981), for example, requires a series of numbers to be recalled in the order opposite to which they were presented, with the number of digits increasing as the task progresses. Performance on digit recall tasks often fails, however, to predict performance on offline comprehension measures (see Daneman & Merikle, 1996, for a review).

With an eye on issues related to ecological validity, Daneman and Carpenter (1980) noted that although digit span tasks (such as the Backwards Digit Span) involve simultaneous processing and storage of information, the processing component (i.e., remembering digits) does not map strongly onto the typical processing demands faced by readers or listeners. In response, they created a verbal working memory span task—often referred to as a “reading span task” or a “sentence span task”—that includes a substantially stronger language processing component. In their reading span task, participants read a set of sentences and are asked to remember the final word of each. Upon encountering a recall prompt, participants are then asked to recall the sentence-final word of each sentence in the set. As the task progresses, the number of sentences presented before the recall prompt increases incrementally, typically from two to six. A participant’s reading span score is quantified as the size of the largest set at which the participant can reliably recall all of the sentence-final words. Daneman and Carpenter’s initial studies (Daneman & Carpenter, 1980) demonstrated that scores on the reading span measure correlated significantly with measures of reading comprehension, such as the verbal section of the Scholastic Aptitude Test (see also Daneman & Merikle, 1996; Dixon, LeFevre, & Twilley, 1988; King & Just, 1991; MacDonald et al., 1992;

Rankin, 1993, for additional reports of relationships between scores on the Daneman and Carpenter span task and performance on a wide range of offline reading- and language-processing-related measures).

Variability in scores on the reading span task also accounts for variability in patterns of reading times (RTs) elicited during sentences that contain manipulations of syntactic complexity (e.g., Just & Carpenter, 1992).

- 1A The reporter that attacked the senator admitted the error. (subject relative)
- 1B The reporter that the senator attacked admitted the error. (object relative)

In (1), for example, sentences with a head noun (*the reporter*) that is the object of the embedded verb (*attacked*), as in (1B), are famously more difficult to process than sentences where the head noun is the subject of the embedded verb, as in (1A). This effect is evident through increased RTs on the main verb (*admitted*) of the object- as opposed to the subject-embedded relative clause sentences (King & Just, 1991—though see Reali & Christiansen, 2007). When encountering syntactically complex sentences such as those containing object-embedded relative clauses, King and Just (1991) found that participants with low reading span scores produced longer RTs on the difficult regions of these sentences than their high-span counterparts. They argued that the object–subject ordering of the object-embedded relative clause more quickly taxed the limited verbal working memory resources available to the lower span participants. Just and Carpenter (1992) interpreted these and similar results as evidence that the systems supporting syntactic processing rely upon a single pool of working memory resources, and that such a resource pool exists independently of linguistic knowledge. Under their account, a higher verbal working memory capacity fosters greater resilience to syntactically complex sentences during online language comprehension.

Evidence for the influence of linguistic experience on reading span task performance

Although reading span tasks contain a rote memory component (participants need to retain the sentence-final word of each sentence in the set), the task also engages perceptual, phonological, syntactic, and semantic processes. Based on this observation,

MacDonald and Christiansen (2002) proposed that reading span tasks are better conceptualized as measures of language processing skill, the development of which is driven by linguistic experience. Under their account, relationships between vWM scores and RTs on syntactically complex sentences arise as a result of shared variance attributable to language processing skill. Thus, instead of reflecting the size of a functionally independent verbal working memory resource pool, reading span scores are indirect indices of variability in linguistic experience.

To evaluate their experience-based claim, MacDonald and Christiansen (2002) trained a series of neural networks to predict the next word in syntactically simple versus complex sentences. They trained 10 simple recurrent networks (SRNs; Elman, 1990) on sentences from a context-free grammar with grammatical properties inherent to English such as subject–verb agreement, present and past tense verbs, and so on. Importantly, many of the training sentences contained simple transitive and intransitive constructions, and a small proportion (about 5%) of the training sentences contained embedded relative clauses, equally divided between subject (1A) and object (1B) relative constructions. To investigate the role of experience on the networks' abilities to learn, they examined the average network performance on novel test sentences containing object- and subject-embedded relative clause constructions after one, two, or three training cycles.

After the first training epoch—and, thus, early in training—the networks exhibited processing difficulty on the critical region of the object- but not the subject-embedded relative clause sentences. This pattern is consistent with the pattern of RTs produced by low-span participants in King and Just (1991). After additional training, however, the difference in processing difficulty between the two sentence conditions decreased. More experience with distributional patterns embedded in language yielded performance progressively approximating the performance of individuals with high reading span scores.

Another approach to examining the effects of linguistic experience on the processing of complex sentences is to train participants on infrequent sentence types, such as object-extracted relative clauses. Wells, Christiansen, Race, Acheson, and MacDonald (2009) systematically manipulated participants' exposure to relative clause constructions over the course of three 30–60-minute experimental sessions

spanning multiple weeks. During the three training sessions, an experimental group of participants read an equal number of subject and object relatives. A control group, however, read the same number of sentences, but did not encounter embedded relatives (i.e., they read complex sentential complements and conjoined sentences). Both groups were matched beforehand on reading span scores. Importantly, on a post-test administered after training, the two groups' processing of relative clauses diverged. RTs from the experimental group resembled the pattern for high-span individuals, whereas the control group produced the low-span RT profile. This experiment provides a compelling example of how variability in the linguistic experiences of an individual influences their ability to process complex syntactic structures at some later point in time (see Christiansen & Chater, 2016, for further discussion of experience-related effects on relative clause processing).

The psychometric properties of reading span tasks

Examination of the psychometric properties of reading span tasks provides additional evidence that increasing the language processing component of the task contributes to an increase in the degree to which task scores reflect variability in linguistic experience. Waters and Caplan (1996a, 1996b), for example, evaluated the reliability and validity of the Daneman and Carpenter reading span task (and derivatives thereof). First, Waters and Caplan (1996b) argued that the processes engaged by the sentence-reading portion of the reading span task are unrelated to the types of syntactic computations generally carried out during sentence processing. They also noted that the reading span task requires "controlled" processing (explicit recollection of temporarily stored information), in contrast to general language comprehension tasks that are more implicit in nature. Their final point of contention was that the Daneman and Carpenter reading span task has many forms, and that assessments of test-retest reliability and equivalence across forms were lacking. In response, Waters and Caplan (1996b) examined the relationships between several working memory measures, including the Daneman and Carpenter reading span task, and various measures of global verbal ability, such as receptive and reading vocabulary, reading comprehension, and reading rate, measured mostly by means of the Nelson-Denny Reading Test (Nelson &

Denny, 1960). Overall, Waters and Caplan found a low retest reliability for the Daneman and Carpenter reading span measure (.41). In fact, a number of individuals actually changed span categories. Some low-span individuals were reclassified as high span, and vice versa.

In light of these noted weaknesses, Waters and Caplan (1996b) created a new version of the reading span task. It too requires participants to recall sentence-final words from incrementally increasing sentence sets. The Waters and Caplan version of the task differs from the Daneman and Carpenter span task in that participants read the sentences to themselves on a computer display (instead of reading them out loud). Additionally, the task incorporates sentence types of varying syntactic complexity and also requires participants to interpret and evaluate the semantic acceptability of each sentence. They argued that these additions produce a task that better accounts for the concurrent processing component of the verbal working memory construct. Task performance was scored by taking the highest set-level (number of sentences presented before recall prompt) at which the participant accurately and reliably performed. Waters and Caplan found that their working memory span task exhibited greater test-retest reliability than the Daneman and Carpenter reading span task. They also created a composite working memory score by summing the standardized scores for the speed, (semantic judgment) accuracy, and word recall components of the task. They found that measures of reading comprehension were most highly associated with the composite score (see also Friedman & Miyake, 2004, for another example of increased predictive ability upon adding processing times to the calculation of sentence span scores).

In pursuit of ecological validity, reading span tasks became progressively more imbued with task components that tap into language processing skill. Scores on the more language-heavy versions of these tasks may engender a higher degree of variance that is shared with measures of language processing skill, and thus at least in part with measures of the quantity and quality of an individual's linguistic experiences.

The Present Study

Although evidence suggests that variability in linguistic experience contributes to an individual's score on reading span tasks, much of this evidence is indirect.

In the human training experiment detailed above, for example, linguistic experience was manipulated but the experimental and control groups were matched on span scores. An examination of the psychometric properties of different reading span measures also suggests that linguistic experience may be one contributor to variability in reading span scores. Reading span scores, however, were not systematically examined in relation to other indices of linguistic experience. Moreover, the measures of reading comprehension typically utilized to assess the contribution of span scores to reading comprehension have tended to be global offline metrics of language ability, rather than measures of online comprehension. In this paper, we analyse data collected from over 70 participants on five psychometrically and theoretically relevant tasks. We use these data to evaluate three hypotheses, discussed below, which stem from the claim that reading span scores capture, at least in part, variability in linguistic experience.

Hypotheses

Goal 1: Correlational evidence for the contribution of linguistic experience to vWM span task performance

First, we aimed to determine whether reading span scores correlate with individual differences in indices of linguistic experience. As such, we administered five individual difference measures. Three measures were administered in an attempt to gauge variability in the amount of an individual's exposure to linguistic input. We focus our efforts here on variability in exposure to written language, given that infrequent vocabulary words and complex syntax are more likely to occur in written language (Biber, 1986; Hayes, 1988; Roland, Dick, & Elman, 2007). It is nearly impossible to reliably count an individual's exposures to specific sentence types (i.e., there are as of yet no person-specific corpora of written or spoken language). Therefore, following previous work, we quantify individual participants' linguistic experience using a variety of measures (1–3 below) that provide proxy estimates of an individual's exposure to printed material.

Individual differences measures

1. *Author Recognition Task* (ART; Stanovich & West, 1989; West, Stanovich, & Mitchell, 1993)—a measure of the amount of text to which someone has been exposed. ART is a questionnaire that lists potential author names. Some of the names belong to actual, well-known authors, and some of the names are foil (false) non-author names. Participants are instructed to read the list and place a checkmark next to the names they believe to be real authors. By assumption, people who spend more time reading should also be better at distinguishing actual author names from false ones. In support of the task's validity, people who were observed reading in public had significantly higher scores on the task than did people who were not (West et al., 1993), and scores on the ART have been shown to correlate significantly with measures of various reading-related processes (Acheson, Wells, & MacDonald, 2008; James & Watson, 2013; Stanovich & West, 1989).
2. *Vocabulary Task* (VOCAB; Shipley, 1940)—a measure of vocabulary size, which is often argued to be a strong indicator of the amount of time an individual spends reading. Hayes (1988), for example, analysed the lexical richness of natural conversations, language used in TV programmes, and a variety of written text sources. Written sources contained more infrequent words than other sources of language input. He argued that exposure to text is likely to be a key predictor of the acquisition of words that are not heavily redundant in non-written sources.
3. *Need for Cognition* (NEEDCOG) scale (Cacioppo, Petty, & Kao, 1984)—a personality-based variable that indexes the degree to which an individual prefers cognitively engaging activities—such as reading—to activities that require less cognitive engagement (Cacioppo & Petty, 1982). We reasoned that NEEDCOG might serve as a plausible “motivational” proxy to linguistic experience, under the assumption that individuals with higher need for cognition will be more likely to engage with printed materials and thus to possess a higher degree of print exposure.

We also administered a reading span and a digit span measure:

1. *Waters and Caplan (1996b) span task* (vWM)—as per the discussion above, we reasoned that scores on this task reflect, at least in part, variability in linguistic experience. We chose this version of a vWM span task because it contains the largest

language processing component of any available vWM span task.

2. *Backward Digit Span* (BDS; Wechsler, 1981)—requires a series of numbers to be recalled in the order opposite to which they were presented. Given the relatively non-linguistic nature of this task, its inclusion provides us the ability to quantitatively partition out variance associated with a non-language-heavy working memory (or, operation span) measure and variance associated with the language-related processing-skill component of the vWM task.

Should some proportion of variability in vWM span scores reflect variation in processing skill driven by differences in linguistic experience, we predict that scores on our proxy measures of linguistic experience will correlate positively with vWM span task scores.

Goal 2: The relationship between vWM and online language processing skill

The linguistic input that individuals are exposed to on a daily basis is highly structured, and individuals are sensitive to this structure during comprehension. For example, readers are sensitive to conditional probabilities between adjacent words, such that reading times on the second word of a two-word pair (bigram) decrease in proportion to the probability of those two words occurring together in natural language (McDonald & Shillcock, 2003). Participants have even demonstrated sensitivity to frequency differences in the probability of occurrence of four-word (4-gram) phrases (Arnon & Snider, 2010; see Caldwell-Harris, Berant, & Edelman, 2012, for a review of various lexical-level frequency effects).

The processing of syntactic structures is similarly sensitive to the frequency with which they occur, such that less frequent structures take longer to process. For example, the varying frequencies of different relative clause types are directly reflected in the ease with which adults process such constructions (Gennari, Mirkovic, & MacDonald, 2012; Jäger, Chen, Li, Lin, & Vasishth, 2015; Real & Christiansen, 2007; see also Kidd, Brandt, Lieven, & Tomasello, 2007), and comprehenders demonstrate sensitivity to the probability with which a specific verb occurs with different structures (e.g., Garnsey, Pearlmutter, Myers, & Lotocky, 1997; MacDonald, Pearlmutter, & Seidenberg, 1994). The well-established relationship between online processing times and frequency manipulations is often taken as evidence that

indices of processing difficulty reflect the degree of expectation for a linguistic event (e.g., Jurafsky, 1996). For example, “surprisal”—or, the negative log probability of a word given preceding context (Hale, 2001; Levy, 2008)—strongly predicts indices of processing difficulty. Words with higher surprisal values elicit more processing difficulty (e.g., Boston, Hale, Vasishth, & Kliegl, 2011; Demberg & Keller, 2008; Jäger et al., 2015). Expectancy-dependent frameworks often assume that the strength of an expectancy is derived from the cumulative effects of exposure to linguistic input (e.g., Hale, 2001; Husain, Vasishth, & Srinivasan, 2014; Levy, 2008).

Consider, for example, the main verb/reduced relative clause (MV/RC) ambiguity, as expressed in Example 2, taken from materials provided by MacDonald et al. (1992).

- 2a. The experienced soldiers/warned about the dangers/**before the midnight**/raid.
- b. The experienced soldiers/spoke about the dangers/**before the midnight**/raid.
- c. The experienced soldiers/warned about the dangers/**conducted the midnight**/raid.
- d. The experienced soldiers/who were told about the dangers/**conducted the midnight**/raid.

For Sentences 2a and 2c, the syntactic role of the verb *warned* is ambiguous. It could act either as the main verb (MV) of the sentence (as in 2a), or as the beginning of a reduced relative clause (RC) that modifies the participant (as in 2c). Although readers cannot resolve the ambiguity before encountering the disambiguating region (**bolded** in Example 2), they exhibit a strong bias in favour of the MV reading. This bias can be attributed to the fact that for the verbs utilized in the experiment reported below, the probability of an MV/RC ambiguity-producing verb occurring in an MV structure is approximately .7 in natural language. The probability of the verb being used as the beginning of the RC, however, is less than .01 (as estimated from corpus data reported by Roland et al., 2007). The point of disambiguation contains the information necessary to arrive at the ultimately intended interpretation of the ambiguity. Given participants’ strong pre-existing bias towards MV disambiguation, little to no evidence of processing difficulty is typically detected during the disambiguating region of ambiguous sentences like (2a), relative to an unambiguous control sentence (2b, where the verb *spoke* cannot

head an RC, thus eliminating the potential for ambiguity). When the ambiguity is resolved in accordance with the RC interpretation (2c), and thus in a manner that is inconsistent with the reader's expectations, processing difficulty in the form of increased RTs at the point of disambiguation is observed, relative to an unambiguous RC baseline (1d, where the inclusion of "who were" eliminates the ambiguity). The tendency for participants to experience processing difficulty upon encountering an unexpected resolution of a temporary syntactic ambiguity is typically referred to as the "garden-path effect".

If vWM scores capture variability in linguistic experience, and thus in the strength of syntactic expectations possessed by an individual, then scores on the vWM span task should correlate significantly, and positively, with the magnitude of the garden-path effect.

We note here, however, that the logic underlying this prediction is based on an assumption—namely, that more linguistic experience results in stronger expectancies. First, we note that the strong link between surprisal values and indices of processing difficulty indicates that expectancies are tightly yoked to conditional probability of occurrence in natural language (as per the discussion above). Additionally, much recent work on anticipatory processing lends support to the guiding role of expectancies in online language processing. For example, Mishra, Singh, Pandey, and Huettig (2012) demonstrated that literate participants used semantic and syntactic knowledge about language in order to anticipate the identity of a target referent well before the noun denoting the target referent became available. Participants with low literacy levels, on the other hand, did not fixate the target noun until slightly after its onset. In a similar anticipatory looking paradigm, Huettig and Brouwer (2015) found that both dyslexic and control participants utilized grammatical information to anticipate a target referent, although anticipatory looks to it were initiated significantly later in the dyslexic group. These results are consistent with recent observations that literacy onset exerts profound effects on language comprehension (e.g., Mani & Huettig, 2014; Montag & MacDonald, 2015) and indicate that exposure to written language is a key determinant of anticipation during language processing (see also James & Watson, 2014, for established links between ART scores and anticipatory looking behaviour during spoken language comprehension, and Rommers, Meyer, & Huettig, 2015, for evidence that

an individual's vocabulary size is strongly linked to the strength of expectancies during online comprehension).

Goal 3: Differential effects of BDS- and vWM-span scores on online comprehension

As explained in the introduction, both BDS and vWM require participants to process some information and to recall some portion of it. The primary difference between the two measures is that vWM requires extensive linguistic processing (of phonological, lexical, semantic, and syntactic information), while the BDS task requires only phonological processing (participants must subvocally rehearse digits that are to be recalled in the reverse order in which they were encountered). Administering both of these tasks in conjunction with the sentence materials that contain a syntactic ambiguity provides us with the opportunity to explore the independent effects of each variable on the processing of syntactically unexpected events. As expressed above, if variability in susceptibility to the garden-path effect is primarily associated with the language processing task demands embedded in the vWM task, we predict a positive relationship between vWM and the garden-path effect on RC sentences. This positive relationship should, however, remain significant, and positive, after statistically controlling for the effect of BDS on individual differences in susceptibility to the garden-path effect.

Method

Participants

Seventy-two native English-speaking ($M = 18.89$ years, $SD = 0.99$) undergraduate students participated in this study for credit in an introductory psychology course. One participant's data were excluded due to a self-reported auditory processing deficit.

Materials

An updated version of the Author Recognition Test (West et al., 1993) was used as a measure of print exposure. Participants were presented with a list of 82 potential author names; 41 were real authors, and 41 were foil (false) names. The foil names were presented in order to correct for guessing. Participants were instructed to read the list and place a checkmark next to the names they believed to be real authors. Scores on this task reflect the proportion of real

author names checked by a participant minus the proportion of foil names that the participant checked.

Vocabulary was measured with the Shipley (1940) vocabulary task. Participants were presented with a target word and were required to choose the closest synonym from a list containing four potential synonyms. The task contained 40 target words, and VOCAB scores denote the number of items for which the participant chose the correct synonym.

Need for cognition (NEEDCOG) was measured using a revised 18-item version of the Need for Cognition (NCS) scale (Cacioppo et al., 1984). Participants rated the relevance of each item to themselves (e.g., *I would prefer complex to simple problems*) on a 9-point Likert-type scale (−4 = extremely inaccurate, 4 = extremely accurate). NEEDCOG scores were created by summing responses to all items, with negative polarity items reverse scored. Higher scores thus reflect higher levels of need for cognition.

The backward digit span task (BDS) was taken from the Wechsler Adult Intelligence Scale–Revised (WAIS–R; Wechsler, 1981). It consisted of 14 strings of digits, with two strings occurring at each set size (i.e., the number of digits appearing before the recall prompt). At the beginning of the task, participants saw two digits presented rapidly one after another. After both were presented, an asterisk appeared, and participants were instructed to recall the digits in the order opposite to the one in which they appeared. The number of digits at each set level increased by one at each new set level, starting with two and ending with eight. Participants completed two trials at each set level. Scores on the BDS task reflect the number of consecutive trials for which participants correctly recalled all digits in the correct (reversed) order.

Verbal working memory (vWM) span was measured by a modified version of the Waters and Caplan (1996b) span task. Participants were presented with a sentence and were asked to make a semantic acceptability judgment by pressing the “YES” key if the sentence was semantically felicitous, or the “NO” key if it was not. Another sentence appeared immediately after the semantic judgment was made, and participants were asked to repeat the process. After all sentences in each sentence group were presented, an asterisk appeared on the screen, and participants were requested to write down the final word of each sentence in the sentence group. The number of words the participant had to maintain in memory (i.e., the number of sentence-final words to be recalled)

while making semantic judgments increased incrementally. Three items—or sentence groups—occurred at each set level, such that participants had three attempts at the two-sentence level, three attempts at the three-sentence level, and up through the final six-sentence level. Participants were instructed to keep going all the way until the end of the task, even if they were unable to remember some of the words. Scores on the Waters and Caplan (1996b) vWM span task reflected the highest level at which participants were able to recall all words, in the correct serial order, from at least two of the three sentence groups. Participants were also given a half of a point if they correctly answered one of the sentence groupings from the level occurring after the highest set-level that was successfully completed.

Online comprehension measure

The sentence materials (Example 2, above) consisted of a modified version of those used in MacDonald et al. (1992). In their experiment, 24 items were created from triplets of verbs. For instance, the verb triplet *warned, spoke, and who were told* would correspond to an item with four possible conditions, as in (2). In MacDonald et al. (1992), eight MV/RC-ambiguous verbs—such as *warned*—were used to create eight such triplets. Three items were derived from each triplet by varying the lexical content of the sentences. In order to extend the original MacDonald et al. sentence set, we introduced four more triplets (taken from Kemtes & Kemper, 1997) and constructed three items from each triplet. This added 12 items to the 24 from MacDonald et al. (1992), thus yielding a total of 36 experimental items.

The 144 sentences from the 36 experimental items were counterbalanced across four presentation lists such that each participant only saw one version of each item, but an equal number of trials per each condition produced by this 2 × 2 (Sentence Type × Ambiguity Status) design. Each list also contained 50 unrelated filler items along with eight practice items.

Online comprehension was assessed with a self-paced reading task. Participants were randomly assigned to one of the four presentation lists, and the order of item presentation was randomized for each participant. All sentences were presented in a non-cumulative, word-by-word moving window format (Just, Carpenter, & Woolley, 1982) using Psy-Scope Version 1.2.5 (Cohen, MacWhinney, Flatt, & Provost, 1993). At the beginning of each trial, an

entire test sentence appeared across the centre of the screen (left-justified) in such a way that dashes preserved the spatial layout of the sentence, but masked the actual characters of each word. As the participant pressed the “GO” key, the word that was just read disappeared, and the next one appeared. RTs (ms) were recorded for each word, reflecting the amount of time that each individual word was present on the display. After the final word of each sentence was read, participants answered a yes/no comprehension question, included to encourage the reading of the sentence materials for meaning.

Procedure

All tasks were administered in the same order to all participants. Participants first completed the vocabulary task, followed by the Waters and Caplan reading span task, the online language comprehension task, the Need for Cognition task, and the Author Recognition Task. The order of task administration was held constant across participants to avoid introducing variability into performance (on any of the tasks) that could be attributed to different administration orderings.

Results

Goal 1: Reading span score correlations with proxy measures of linguistic experience

The means and standard deviations for each individual difference measure appear in Table 1, and the correlations among the measures are presented in Table 2. vWM correlated significantly, and positively, with VOCAB and ART, demonstrating that participants with higher amounts of print exposure and vocabularies also have higher vWM scores. These relationships are consistent with previously reported significant positive relationships between vWM and either ART or VOCAB (e.g., Payne, Gao, Noh, Anderson, & Stine-Morrow, 2012). We detected no relationship, however, between vWM and NEEDCOG. BDS scores did not

correlate with any other measure. Thus, two of our proxy measures of linguistic experience correlated positively with vWM scores, whereas BDS scores—designed to measure working memory but without a strong language processing component—did not.

Goal 2: Increased experience with linguistic input increases sensitivity to violations of statistical regularities

We segmented sentences into the same regions originally used by MacDonald et al. (1992), as indicated by the forward slashes in (2) above. The first segment contains no manipulation of interest. For sentences in the ambiguous sentence condition, the second region, or “the point of ambiguity”, begins with the ambiguity-producing verb and terminates before any disambiguating information appears. In the unambiguous sentence condition, Region 2 begins at the onset of the word that eliminates the ambiguity and ends at the same location as that specified in the ambiguous sentence condition. The third region, or “the point of disambiguation”, begins with the first word that could be used to arrive at one interpretation of the temporary ambiguity. It also includes all subsequent words in the sentence except for the final word. Region 3 included the same words for sentences in the unambiguous sentence condition. In all sentence conditions, Region 4 included only the final word of each sentence. The sentence-final word was excluded from the disambiguating region due to sentence “wrap-up” effects, in which increases in RTs frequently occur due to extra processing before participants progress to a comprehension question.

First, we asked whether the self-paced reading experiment replicated the classic garden-path effect. All RTs less than 100 ms or greater than 2000 ms were removed. The remaining RTs were then log-transformed to increase the normality of the distribution of residuals (Box & Cox, 1964). Linear mixed-effects models were used for all analyses and were implemented with the lme4 package (Bates, Maechler, & Bolker, 2012) in the R environment (R Development Core Team, 2014). Sentence type was effect coded (−1 = main verb, 1 = relative clause), as was ambiguity status (−1 = unambiguous, 1 = ambiguous). In these and all models reported below, the maximal random-effects structures were utilized (Barr, Levy, Scheepers, & Tily, 2013), including a random intercept for both subjects and items, as well as random slopes for the full factorial of Sentence Type × Ambiguity Status (the two within-subjects variables) on both

Table 1. Descriptive statistics for each individual difference measure.

Variable	Possible range	Observed range	Mean	SD
vWM	0 to 6	2 to 6	4.43	1.09
VOCAB	0 to 40	25 to 40	31.32	3.14
ART (%)	0 to 100	14 to 57	31.00	11.00
NEEDCOG	−72 to 72	−45 to 59	10.68	22.76
BDS	0 to 14	4 to 14	9.47	2.48

Note: vWM = verbal working memory; VOCAB = Vocabulary Task; ART = Author Recognition Task (in percentages); NEEDCOG = Need for Cognition; BDS = Backward Digit Span.

Table 2. Correlations among scores on each individual difference measure.

Variable	vWM	VOCAB	ART	NEEDCOG	BDS
vWM	—				
VOCAB	.30*	—			
ART	.29*	.44**	—		
NEEDCOG	.10	.34**	.21	—	
BDS	.13	.05	-.11	.005	—

Note: vWM = verbal working memory; VOCAB = Vocabulary Task; ART = Author Recognition Task; NEEDCOG = Need for Cognition; BDS = Backward Digit Span.

*Correlation significant at the .05 level (two-tailed). **Correlation significant at the .01 level (two-tailed).

the subject and item terms. In the event that a model would not converge, maximal random effects structures were reduced in a step-wise manner by removing the term on the random effects structure to which the least amount of variance was attributed until the model did converge (following the recommendations put forth by Barr et al., 2013).

In order to determine the relationship between sentence type and ambiguity status across each region, four separate models were conducted, one for each segment. Log-transformed RTs at each region were regressed onto the main effects of sentence type and ambiguity status, as well as their two-way interaction. Any *t*-value with an absolute value exceeding 1.96 was considered statistically significant at an alpha level of .05. The results of the models are summarized in Table 3.

No significant effects of sentence type or ambiguity status occurred at Region 1, nor was there evidence of an interaction between the two variables. Only a significant effect of ambiguity status occurred at Region 2 ($\beta = 0.01$, $SE = 0.01$, $t = 2.03$), such that sentences containing an ambiguity were read longer than their unambiguous counterparts. At Region 3, where participants encountered disambiguating information, we observed significant effects of sentence type ($\beta = 0.04$, $SE = 0.02$, $t = 2.53$) and ambiguity status

($\beta = 0.03$, $SE = 0.01$, $t = 3.92$): The disambiguating region was read more slowly for RC sentences than for MV sentences, and more slowly for ambiguous than for unambiguous sentences. Additionally, these two variables significantly interacted ($\beta = 0.02$, $SE = 0.01$, $t = 3.66$). As is evident in Figure 1, the RT difference between ambiguous and unambiguous sentences held only for RC sentences. This result replicates the classic garden-path effect on RC sentences previously elicited with different versions of this sentence set (Kemtes & Kemper, 1997; MacDonald et al., 1992; Pearlmutter & MacDonald, 1995). We note here that the same garden-path effect occurred on the final word of the sentence ($\beta = 0.02$, $SE = 0.01$, $t = 2.31$), consistent with the observation that differential amounts of processing difficulty can “spill over” to the final word of a sentence, even when readers have encountered sufficient disambiguating information.

Next, we explored the relationship between the magnitude of the garden-path effect and each of the five individual differences variables. Log-transformed RTs at the disambiguating region were regressed onto the main effects of sentence type, ambiguity status, all five of the individual difference variables, and all possible interactions among the sentence-level and individual difference variables (but not including interactions among the individual

Table 3. Regression coefficients and test statistics from linear mixed-effects models for the Sentence Type \times Ambiguity Status interaction at each region.

	Preamble (Region 1)			Point of ambiguity (Region 2)			Point of disambiguation (Region 3)			Sentence-final word (Region 4)		
	Est. β	SE	<i>t</i>	Est. β	SE	<i>t</i>	Est. β	SE	<i>t</i>	Est. β	SE	<i>t</i>
Intercept	5.92	0.02	251.04	5.91	0.02	258.27	5.91	0.02	269.55	6.37	0.03	199.62
Sentence type(-1 = main verb)	-2.6×10^{-4}	0.01	-0.02	-1.3×10^{-3}	0.02	-0.08	0.04	0.02	2.53	0.02	0.03	0.81
Ambiguity status(-1 = unambiguous)	3.6×10^{-3}	0.01	0.66	0.01	0.01	2.03	0.03	0.01	3.92	0.04	0.01	2.87
Sentence Type \times Ambiguity Status	-1.7×10^{-3}	0.01	-0.37	0.01	0.01	1.09	0.02	0.01	3.66	0.02	0.01	2.31

Note: $|ts| > 1.96$ are considered statistically significant at an alpha level equal to .05 and are shown in bold. Est = estimated.

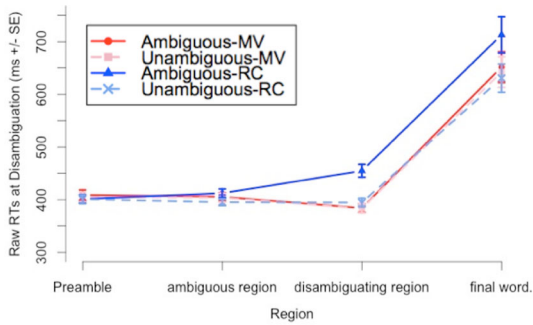


Figure 1. Mean raw reading times (RTs) at each sentence region (indicated in Example 2) for each sentence condition. MV = main verb; RC = reduced relative clause. Error bars represent 95% confidence intervals on the mean. To view this figure in colour, please visit the online version of this Journal.

differences measures, which are not theoretically motivated). Each individual difference variable was centred in order to reduce multi-collinearity between interaction terms and the lower order effects. The maximal possible random effects structure was identified as described above. The results of the model are summarized in Table 4.

In this full model, the effects of sentence type and ambiguity status—along with their two-way interaction—remained significant. Both vWM ($\beta = 0.06$,

$SE = 0.02$, $t = 3.09$) and BDS ($\beta = -0.03$, $SE = 0.002$, $t = -3.34$) exerted significant effects on log RTs. The positive vWM effect indicates that individuals with higher vWM scores exhibited more processing difficulty at the disambiguating region. Conversely, the negative effect of BDS on log RTs indicates that processing difficulty decreases (RTs go down) at disambiguation as scores on the BDS task increase. No significant effects of ART, VOCAB, or NEEDCOG were observed.

We focus now on the three-way interactions in order to determine whether any of the individual difference variables influenced the magnitude of the garden-path effect on RC-resolved ambiguous sentences. A significant Sentence Type \times Ambiguity Status \times vWM interaction was observed ($\beta = 0.01$, $SE = 0.01$, $t = 2.63$), and the corresponding three-way interactions were marginally significant for both BDS ($\beta = -0.003$, $SE = 0.002$, $t = -1.80$) and VOCAB ($\beta = -0.003$, $SE = 0.002$, $t = -1.72$).

We predicted that individuals with higher vWM scores would exhibit an increased ambiguity effect (defined as the RT difference on ambiguous as opposed to unambiguous sentences), specifically for RCs. To test this prediction, we assessed the simple effect of the vWM \times Ambiguity Status interaction at each level of sentence type. We found a positive,

Table 4. Regression coefficients and test statistics from the linear mixed-effects model including all individual difference variables at the disambiguating region.

	Est. β	SE	t-value
Intercept	5.91	1.96×10^{-2}	302.20
Sent (-1 = main verb)	3.97×10^{-2}	1.55×10^{-2}	2.56
Amb (-1 = unambiguous)	2.72×10^{-2}	6.61×10^{-3}	4.12
vWM	5.53×10^{-2}	1.79×10^{-2}	3.09
BDS	-2.48×10^{-2}	7.43×10^{-3}	-3.34
ART	1.49×10^{-2}	1.87×10^{-1}	0.08
VOCAB	-4.12×10^{-3}	6.78×10^{-3}	-0.61
NEEDCOG	7.29×10^{-4}	8.41×10^{-4}	0.87
Sent \times Amb	2.27×10^{-2}	5.83×10^{-3}	3.89
Sent \times vWM	2.60×10^{-3}	4.04×10^{-3}	0.64
Amb \times vWM	9.94×10^{-3}	4.49×10^{-3}	2.23
Sent \times BDS	6.84×10^{-4}	1.68×10^{-3}	0.41
Amb \times BDS	-9.55×10^{-4}	1.85×10^{-3}	-0.52
Sent \times ART	5.65×10^{-2}	4.28×10^{-2}	1.32
Amb \times ART	6.30×10^{-3}	4.79×10^{-2}	0.13
Sent \times VOCAB	5.14×10^{-4}	1.54×10^{-3}	0.33
Amb \times VOCAB	6.28×10^{-4}	1.71×10^{-3}	0.37
Sent \times NEEDCOG	1.32×10^{-4}	1.90×10^{-4}	0.70
Amb \times NEEDCOG	2.54×10^{-4}	2.09×10^{-4}	1.21
Sent \times Amb \times vWM	1.16×10^{-2}	4.40×10^{-3}	2.63
Sent \times Amb \times BDS	-3.29×10^{-3}	1.83×10^{-3}	-1.80
Sent \times Amb \times ART	7.67×10^{-3}	4.73×10^{-2}	0.16
Sent \times Amb \times VOCAB	-2.89×10^{-3}	1.68×10^{-3}	-1.72
Sent \times Amb \times NEEDCOG	5.75×10^{-5}	2.06×10^{-4}	0.28

Note: $|ts| > 1.96$ are considered statistically significant at an alpha level equal to .05 and appear in bold. Sent = sentence; amb = ambiguity; vWM = verbal working memory; BDS = Backward Digit Span; ART = Author Recognition Task; NEEDCOG = Need for Cognition; VOCAB = Vocabulary Task; Est. = estimated.

significant two-way interaction between ambiguity status and vWM for RCs ($\beta = 0.02$, $SE = 0.001$, $t = 3.23$), but no evidence for a corresponding two-way interaction for MVs, ($\beta = -0.002$, $SE = 0.001$, $t = -0.28$). As illustrated in Figure 2, individual differences in vWM scores influence RTs in the RC sentence condition. More specifically, the magnitude of the garden path effect on RC sentences increases as vWM scores increase. Higher span individuals are more susceptible to the garden-path effect than are their lower span counterparts.

Additionally, the negative sign on the estimated beta coefficient for the three-way interaction term involving BDS indicates that the garden-path effect decreases as BDS scores increase. We assessed the simple effect of the Ambiguity Status \times BDS interaction for each level of sentence type. A negative, marginally significant two-way interaction occurred for RCs ($\beta = -0.004$, $SE = 0.003$, $t = -1.53$), although no evidence of the two-way interaction was detected for MV sentences ($\beta = 0.002$, $SE = 0.002$, $t = 0.96$). This pattern of results suggests that the Sentence Type \times Ambiguity Status \times BDS interaction was driven by a

decrease in the magnitude of the garden-path effect in the RC sentence type condition, a pattern illustrated in Figure 3.

The Sentence Type \times Ambiguity Status \times VOCAB interaction trended toward significance ($\beta = -0.003$, $SE = 0.002$, $t = -1.72$). The tests of simple main effects, however, indicate that vocabulary influenced the ambiguity effect on MV sentences ($\beta = 0.004$, $SE = 0.002$, $t = 1.58$) much more so than on the RC sentences ($\beta = -0.002$, $SE = 0.003$, $t = -0.89$). As indicated in Figure 4, as VOCAB increased, RTs increased in the RC sentence condition. But, the increase conferred by higher VOCAB scores was roughly equivalent for both the ambiguous and the unambiguous sentences. A negative relationship occurred, however, between VOCAB and the ambiguity status effect in the MV condition (see Figure 4). The models reported here, and below, were conducted while accounting for the effects of VOCAB on the Sentence Type \times Ambiguity Status interaction. Additionally, the goals of the present work are directed toward explanations of variability in the processing of garden-path sentences, and not in the processing of highly expected syntactic events such as MV

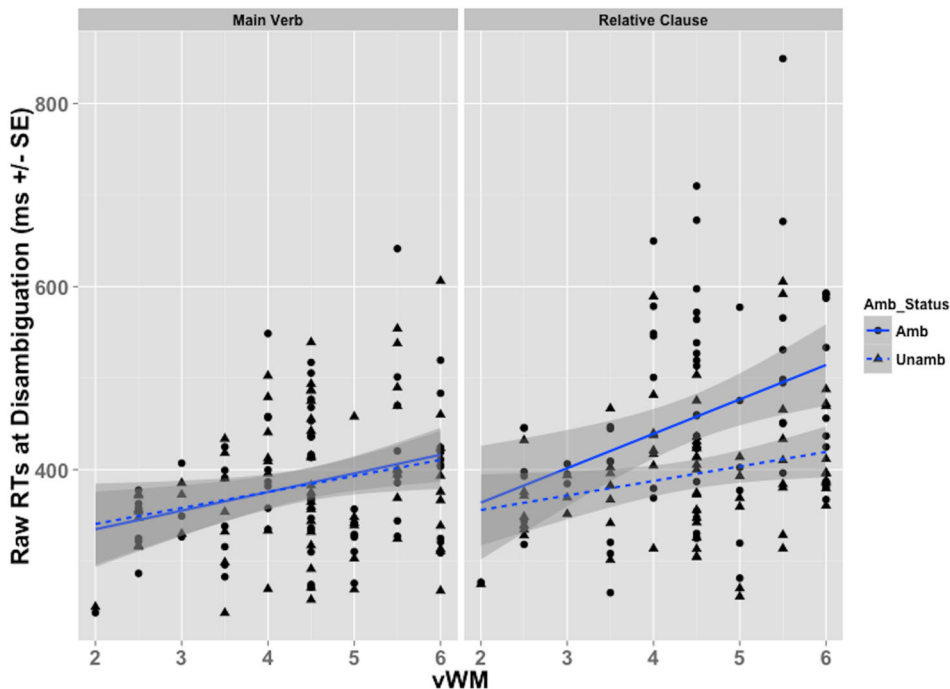


Figure 2. Raw reading times (RTs) for ambiguous (amb) and unambiguous (unamb) sentences in both the main verb (MV; left panel) and reduced relative clause (RC; right panel) sentence conditions plotted against verbal working memory (vWM) scores. The garden-path effect in the RC sentence condition increases in magnitude as vWM scores increase. Grey shaded regions represent 95% confidence intervals on the slopes. To view this figure in colour, please visit the online version of this Journal.

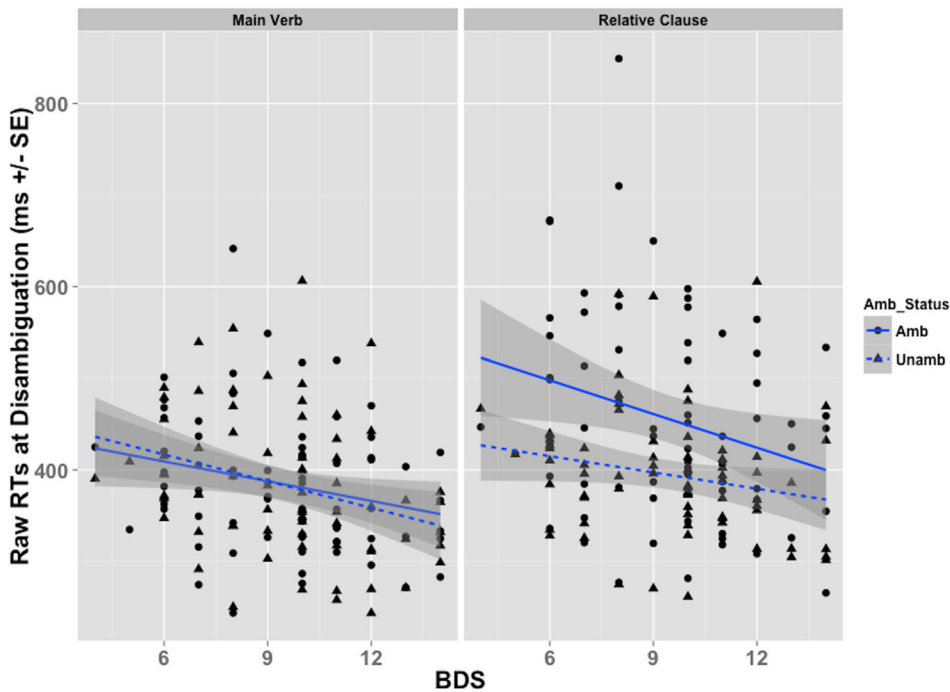


Figure 3. Raw reading times (RTs) for ambiguous (amb) and unambiguous (unamb) sentences in both the main verb (MV; left panel) and reduced relative clause (RC; right panel) sentence conditions plotted against Backward Digit Span (BDS) scores. These data suggest that the garden-path effect in the RC sentence condition decreases in magnitude as BDS scores increase. Grey shaded regions represent 95% confidence intervals on the slopes. To view this figure in colour, please visit the online version of this Journal.

resolved sentences. As a result, we do not interpret this marginal three-way interaction any further.

Goal 3: Differential relationships of vWM and BDS to the magnitude of the garden-path effect

With regard to the construct of verbal working memory, the central question we wish to address pertains to the effect of increased language processing-related task demands embedded in more modern span task measures. The analyses already reported demonstrate that out of the five individual difference variables, vWM is the only significant predictor of variability in the magnitude of the garden-path effect. BDS also exerted a marginally significant effect on the garden-path effect, as captured by the Sentence Type \times Ambiguity Status \times BDS interaction. Of interest, Figures 2 and 3 reveal that vWM and BDS exhibit fundamentally different relationships to RTs elicited by relative clause resolution of the temporary ambiguity. vWM exerts a positive effect on the garden-path effect—a pattern that would be readily predicted if vWM scores express variability in language processing skill. That is, individuals with more linguistic experience—and hence with expectancies that more accurately

reflect the statistics of the input—show greater processing difficulty when they encounter low-probability events such as relative clauses. BDS, on the other hand, exhibits a negative effect in susceptibility to the garden-path effect, such that individuals with better backward digit recall ability were *less* garden-pathed overall. This pattern would be readily accounted for under any number of frameworks highlighting the benefits of larger memory capacities for processing complex syntactic constructions (e.g., Gibson, 1998; Just & Carpenter, 1992).

Through a series of model comparisons, we next ask whether the effect of vWM on the magnitude of the garden-path effect (and, thus, the corresponding three-way interaction) increases model fit after statistically controlling for variability in BDS scores. In the analyses reported here we view BDS as a baseline measure, in the sense that BDS scores reflect an individual's ability to simultaneously process and store information (or perhaps an individual's rote storage ability), but without a heavy language processing component.

We first fitted a model that contained all terms (i.e., the model discussed in the previous section) except

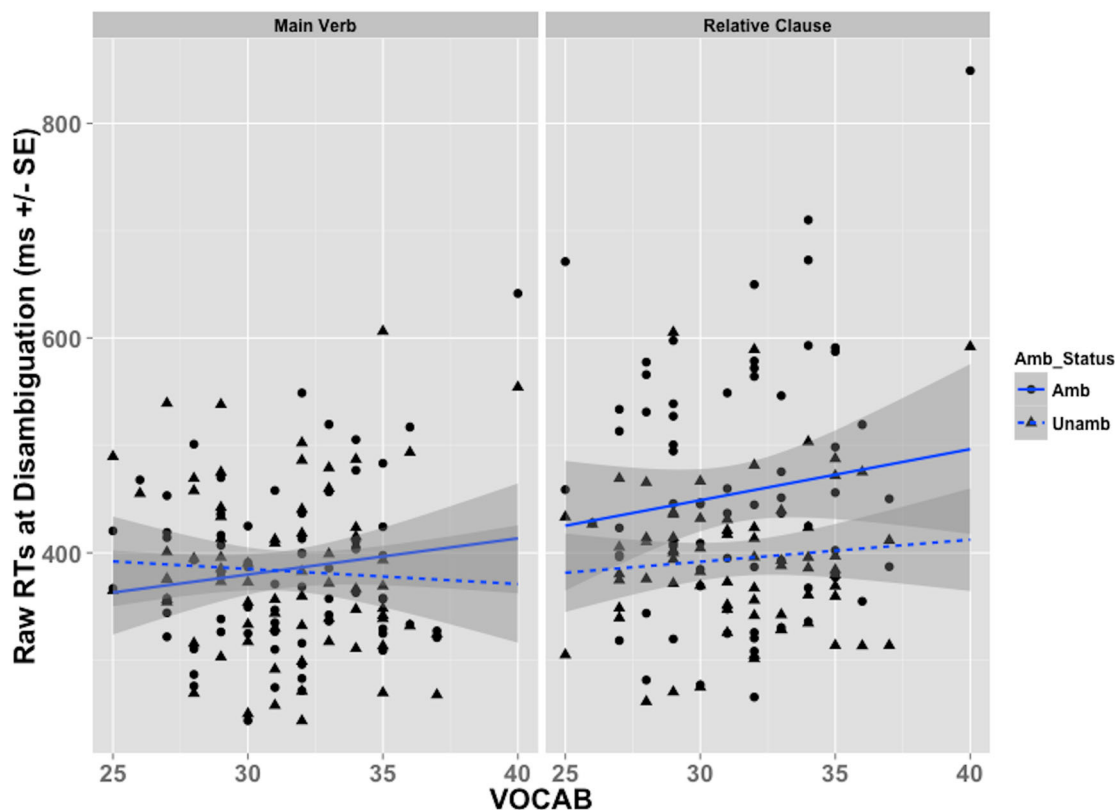


Figure 4. Raw reading times (RTs) for ambiguous (amb) and unambiguous (unamb) sentences in both the main verb (MV; left panel) and reduced relative clause (RC; right panel) sentence conditions plotted against Vocabulary Task (VOCAB) scores. Grey shaded regions represent 95% confidence intervals on the slopes. To view this figure in colour, please visit the online version of this Journal.

for the Sentence Type \times Ambiguity Status \times vWM interaction. In this model, no three-way interaction was statistically significant. When comparing this model to the full model reported above, we find that the inclusion of the Sentence Type \times Ambiguity Status \times vWM interaction significantly improved model fit, $\chi^2(1) = 6.58$, $p = .01$. This result suggests that after controlling for all other predictors, including BDS, the relationship between vWM and the garden path effect still accounts for unique variance in the data.

As Table 2 indicates, however, there are significant inter-relationships between ART, VOCAB, and vWM. Thus, to more directly address our question, we pursue the same model comparison as that detailed directly above, while excluding ART, VOCAB, and COGNEED, along with any interactions involving these three variables. We first fitted a model that included sentence type, ambiguity status, BDS, and vWM. This model also included all possible

interactions between these variables, excluding the Sentence Type \times Ambiguity Status \times vWM interaction. We then fitted the full model, including the crucial Sentence Type \times Ambiguity Status \times vWM three-way interaction. The inclusion of the Sentence Type \times Ambiguity Status \times vWM interaction still significantly improved model fit, $\chi^2(1) = 4.74$, $p = .03$.

These model comparisons demonstrate that after controlling for an individual's ability to simultaneously process and store information in an operation span task that involves little language—such as BDS—vWM still predicts a significant amount of the variability in the garden-path effect. The main difference between the two tasks involves the amount of language—and, thus, of language processing—required by each task. Thus, these results suggest that the language processing component of vWM tasks is the task component that significantly predicts variability in indices of online language comprehension. In other words, vWM and non-linguistic

operation-span style tasks (such as BDS) seem to contribute independent and differential effects on variability in susceptibility to the garden-path effect.

General Discussion

The study presented here was not designed to address issues pertaining to the existence of a verbal working memory resource capacity, and we note that our data do not speak to more general controversies regarding the role of memory in online language processing. Instead, our goal was to investigate the claim, articulated most clearly in MacDonald and Christiansen (2002), that reading span task scores capture variability in language processing skill—a skill that develops largely as a result of linguistic experience. Neural network simulations, human training experiments, and previous psychometric work on reading span tasks offer indirect evidence in support of this claim. Here, we provide a more direct assessment of the experiential hypothesis by examining the relationships among reading span scores, multiple experiential individual differences variables, and variability in sensitivity to the presence of syntactically unexpected events. Three aspects of our data implicate linguistic experience as one key determinant of reported relationships between reading span scores and online processing abilities.

First, vWM scores correlated with proxy measures of linguistic experience, although they did not correlate with scores on the BDS task (a non-linguistic operation span task).

Second, a positive relationship was observed between vWM span scores and susceptibility to the garden-path effect during online processing. Higher span individuals were substantially more garden-pathed than their lower span counterparts. Similar seemingly counterintuitive effects have previously been interpreted as evidence that high-span individuals possess enough working memory resources to maintain multiple interpretations of a structural ambiguity in parallel (MacDonald et al., 1992), whereas the low-span individuals do not. In another report of a positive relationship between reading span scores and garden-path strength, however, Pearlmutter and MacDonald (1995) hypothesized that high-span individuals may be more skilled language users and argued that they are better able to use constraints that support the online comprehension process.

Recent computational-level accounts of language processing attribute processing difficulty to

contextualized estimates of probability of occurrence (e.g., Hale, 2001; Jurafsky, 1996; Levy, 2008), and anticipatory looking experiments suggest that stored knowledge contributes to the generation of probabilistically weighted expectancies during online processing (e.g., Altmann & Kamide, 1999; Mishra, Singh, Pandey, & Huettig, 2012; see additional citations throughout). Thus, in line with Pearlmutter and MacDonald (1995), an experience-based interpretation of the positive relationship reported here is that individuals with higher span scores are likely to possess more robust knowledge about the distribution of syntactic constructions, especially in the written modality. As a result, they would be expected to experience greater processing difficulty upon encountering RC resolution of a temporary ambiguity, given the biases of the ambiguity-creating verbs in this experiment.

Third, and perhaps our most interesting finding, involves the differential relationship we observed between the magnitude of the garden-path effect when considered in relation to vWM and to BDS. One interpretation of digit-span task scores is that they reflect rote storage capacity. Under this interpretation of BDS span scores, one would predict that those with higher storage ability should be more resilient to processing complex or computationally demanding sentences (see Friedman & Miyake, 2004; Nicenboim, Vasishth, Gattei, Sigman, & Kliegl, 2015, for overviews of literature supporting a predicted negative relationship between digit and word span tasks and language processing ability). Indeed, this is the relationship observed here between BDS and the magnitude of the garden-path effect.

Alternatively, we note that the vWM and BDS tasks exhibit many overlapping task demands. Each task requires the processing of increasingly larger arrays of information, coupled with storage demands that increase alongside increases in these processing demands. In each task, the bi-directional effects of concurrent processing and storage demands are assessed by recall accuracy. The main difference between these two tasks, then, is the amount of linguistic processing that each task requires. Sentence-span tasks require a substantial amount of language processing, in addition to storage. BDS tasks require a relatively small amount of linguistic processing, in addition to similar storage demands. Thus, BDS scores may not reflect rote storage per se, but might reflect instead an individual's storage ability in the face of distracting processing demands (and, thus, a more unitary skill that takes into account both

processing and storage). In fact, although digit span tasks are typically construed as measuring memory capacity, recent work indicates that language exposure plays a role even in these tasks (Jones & Macken, 2015), though exactly how this might be reflected in individual differences in scores on such measures is yet unknown.

Regardless of one's preferred interpretation of digit span task scores, our model comparison results indicate that the positive relationship between vWM and the garden-path effect is still present after statistically accounting for the effect of BDS scores on the magnitude of the garden-path effect. In other words, increasing the language-processing component of typical operation span tasks (and of reading span tasks) may be central to increasing the predictive relationship between vWM span task scores and online language processing ability.

This observation has been articulated many times in the historical literature on vWM span tasks and language processing ability. Waters and Caplan (1996b) found, for example, that after accounting for variance associated with a sentence span task and a word comprehension measure, scores on a much less language-heavy span recall task (such as BDS) accounted for only a small amount of additional variability in offline indices of language comprehension. This and other similar observations led them to note that, "the predictive value of a working memory test for reading abilities is mostly a function of the extent to which the processing task in a working memory test requires verbal comprehension skills" (Waters and Caplan, 1996b, p. 76). In many respects, by including an index of online language processing ability, we provide here a direct test of this claim, and our data provide compelling support for it.

We note here that the degree to which reading span task scores correlate with indices of language comprehension has often served as a point of contention. As discussed in the introduction, scores on reading span tasks do correlate with global offline assessments of linguistic- and reading-related abilities. But, Waters and Caplan (1996c) demonstrated that verbal working memory span scores, as measured by the Daneman and Carpenter reading span task, did not influence acceptability ratings elicited in response to sentences containing syntactic ambiguities. Furthermore, Sprouse, Wagers, and Phillips (2012) reported no significant relationship between scores on non-linguistic working memory tasks (such as a serial recall task and an n-back task) and offline

acceptability judgments to sentences of varying syntactic complexity. When a reading span task was utilized in conjunction with similar materials, however, Hofmeister, Casasanto, and Sag (2012) did elicit the hypothesized relationship, but only for sentences that were moderately complex.

The results that motivate such exchanges are based on the relationship between offline acceptability judgments and (verbal) working memory measures. Offline acceptability judgments are susceptible to the influence of meta-linguistic processes that may not resemble the processes typically engaged during online comprehension. We note here, however, that debate even exists regarding the predictive utility of verbal working memory span measures with respect to measures of online language processing ability. In their critique of the Just and Carpenter (1992) account, for example, Waters and Caplan (1996a) rightfully note that King and Just (1991) never reported the results of the statistical analyses that would be necessary to demonstrate an interaction between span group and syntactic complexity. This observation was one factor that motivated Waters and Caplan's critique of the capacity-complexity trade-offs inherent to the capacity theory.

These inconsistent findings highlight an important point: Discrepancies in the nature, or even existence, of a link between scores on verbal working memory tasks and linguistic processing are likely to stem from across-experiment differences in the memory measure used, the type and complexity of the linguistic manipulation, the manner in which language processing or comprehension ability is assessed, the scoring procedure used to calculate span scores, and the analytic strategy pursued by the research team. Given our focus on the experiential components of sentence span task scores—as per its relative neglect within the span task literature—we did not administer the wide array of additional working memory, individual difference, and language comprehension measures necessary to address all of these inter-related issues (see Daneman & Merikle, 1996, for an extensive review of studies that have addressed the relationship between other operation span measures and sentence span task scores; and also A. R. Conway et al., 2005, for an assessment of the effects of different operation span scoring procedures).

Nonetheless, here, we did elicit a significant relationship between vWM span scores and sensitivity to syntactically unexpected events in an online

measure of language processing. We note, however, that we utilized a version of the Waters and Caplan span task. Given the semantic acceptability judgments and the complex syntax embedded in a subset of the sentences, this version of the reading span task is particularly language heavy. Most other experiments examining similar relationships have utilized the Daneman and Carpenter reading span task, thus reducing the comparability of our results to these previous experiments. Certainly, a meta-analysis on the relatively large body of work that seeks to specify the relationship between reading span task scores and online language comprehension ability is warranted. Such a meta-analysis would provide important insight into the effects of these factors on the statistical reliability and possible conceptual interpretations of this theoretically important relationship.

One surprising aspect of our results is the lack of any hint of an effect of ART scores on RTs, and especially the lack of the Sentence Type \times Ambiguity Status \times ART interaction. ART has been shown to predict relatively coarse-grained reading-related variables. Acheson et al. (2008), for example, detected significant correlations between ART and other offline measures of language ability, such as scores on the verbal sections of the ACT. Misyak and Christiansen (2012) reported a moderately positive correlation between ART and language comprehension for a subset of syntactically complex sentences, but these were also offline measures. An emerging body of work strongly suggests, however, that ART does predict processing patterns during online language processing as well (e.g., James & Watson, 2014; Moore & Gordon, 2015; Payne et al., 2014).

One possibility for a lack of ART effects in this present study is that the version of ART implemented here was not appropriately constructed given the experiential backgrounds of our undergraduate participants. To create the version of the task utilized here, we obtained a version of ART from one of its original authors, who also provided us with data collected on the task from 1896 college undergraduates between 1997 and 1999.¹ Given that the popularity of authors changes over time, we decided to update the task before implementing it in the study reported here. Item analyses were conducted on the large data set provided to us. Calculation of Cronbach's alpha revealed that the test was reliable, yielding a reliability coefficient of .90. The item difficulties (the proportion

of participants selecting each name) for the foil items were .15 or less, indicating that no more than 15% of participants thought the names were real authors. This observation indicates that none of the foil names from the original author recognition task were unusually distracting to the participants, such that we did not modify them. After examining the item difficulties for the real author names, however, we found that 11 items appeared especially problematic. These items elicited item difficulties of .06 or less, indicating that fewer than 100 participants labelled them as real authors (got the items correct). Thus, to update the task, we replaced the 11 problematic real-author items with both fiction and non-fiction writers taken from multiple genres of a New York Times bestseller list. It is unclear, however, how well the item difficulty values from the original version of the ART apply to recent college students. Additionally, using author names from the New York bestseller list may not provide reasonable access to author names about which an average college student could reasonably be expected to know. Thus, it is possible that scores on our version of the ART under-represented participant knowledge of author names, or otherwise misrepresented the distribution of print exposure in our sample. We refer readers to other recently updated versions of the task that do appear to provide more accurate assessments of print exposure (e.g., Acheson et al., 2008; Moore & Gordon, 2015). The existence of a significant positive relationship between ART and variability in susceptibility to the garden-path effect thus remains an empirical question to be assessed in future work with more valid, reliable, and up-to-date measures of ART.

At the beginning of this paper, we noted that online language comprehension involves a large and diverse set of perceptual skills and cognitive processes. We also noted that individual differences are likely to exist in each of these. Thus, accounts of individual differences in online language comprehension ability will need to consider the interactive effects of a larger set of variables than typically assessed in much of previous work. In one recent example of a more encompassing individual difference approach to related questions, Van Dyke, Johns, and Kukona (2014) examined the interrelationships between scores on the Daneman and Carpenter version of the reading span task and scores on 23 additional individual difference tasks. The additional tasks were

¹Many thanks to Rich West at James Madison University for providing us with this valuable data source.

designed to gauge by-subject variability in multiple reading- and language-related processes such as phonological processing skill, receptive vocabulary, language experience, and IQ (as derived from the Vocabulary and Matrix Reasoning subtests that constitute the Wechsler Abbreviated Scale of Intelligence). They observed a high degree of multi-collinearity between reading span scores and many of their additional individual difference measures, especially phonological processing skill, word reading skill, and various global offline measures of reading comprehension ability. This observation is consistent with the notion that just like indices of online language processing ability, reading span scores probably reflect variability in multiple skills, processes, and knowledge bases that contribute to performance. Van Dyke et al. also found that reading span task scores correlated significantly with RTs on an online processing measure, but that after partialling out variability associated with IQ scores, the relationship between reading span scores and RTs was eliminated. This result suggests that variability in IQ may ultimately serve as an important variable in uncovering the nature of the relationship between reading span task performance and sentence processing skill. Unfortunately, we did not administer an IQ task in the study reported here, such that we cannot determine whether or not the differential relationship between BDS and vWM is maintained after controlling for IQ scores. Additionally, given the different linguistic manipulations utilized here versus those in Van Dyke et al., it is unclear whether statistically controlling for IQ would exert the same diminishing effect on the relationship between reading span scores and online processing patterns. Thus, the question of IQ effects—especially *non-verbal* IQ or “fluid intelligence”—on the results we report here remains a very important question for future research.

Conclusion

The results presented here suggest that reading span measures and sentence processing tasks are tapping into similar processing skills, and that they are determined, at least in part, by experience. Recent advances in work on statistical learning provide at least part of the solution to why linguistic experience is such a crucial component for the induction of grammatical and other linguistic knowledge. Statistical learning (see e.g., Misyak, Goldstein, & Christiansen, 2012; Romberg & Saffran, 2010, for reviews) refers to the

ability to learn about patterns in the environment from exposure to regularities in the input. A spate of recent research demonstrates a tight coupling between individual differences in statistical learning ability and variability in native language learning and processing, in both child and adult populations (e.g., C. M. Conway, Bauernschmidt, Huang, & Pisoni, 2010; Kaufman et al., 2010; Kidd, 2012; Kidd & Arciuli, in press; Misyak & Christiansen, 2012; Misyak, Christiansen, & Tomblin, 2010a, 2010b), and in adult second-language learner populations (e.g., Ettlinger, Morgan-Short, Faretta-Stutenberg, & Wong, in press; Frost, Siegelman, Narkiss, & Afek, 2013). Given these relationships, we suggest that statistical learning ability is likely to be a strong mediator of the relationship between reading span and the processing of complex syntactic structures (see also Misyak & Christiansen, 2012). It is possible that the ability to perform statistically based chunking of incoming input given linguistic experience might be a key factor in explaining variation in online language processing, as revealed by a recent preliminary study examining the relationship between serial recall of statistically defined sequences and online sentence processing (McCauley & Christiansen, 2015). Indeed, a key role of experience may be to facilitate the chunking and subsequent processing (as well as integration) of input occurring in rapid succession, thus enabling readers and listeners to better anticipate upcoming material (Christiansen & Chater, in press). In light of these results, we argue that a comprehensive account of individual differences in sentence processing will necessarily entail large-scale investigations of interactions among perceptual, cognitive, learning, and environmental factors at different points across an individual's life history.

References

- Acheson, D. J., Wells, J. B., & MacDonald, M. C. (2008). New and updated tests of print exposure and reading abilities in college students. *Behavior Research Methods*, *40*(1), 278–289.
- Altmann, G. T., & Kamide, Y. (1999). Incremental interpretation at verbs: Restricting the domain of subsequent reference. *Cognition*, *73*, 247–264.
- Arnon, I., & Snider, N. (2010). More than words: Frequency effects for multi-word phrases. *Journal of Memory and Language*, *62*, 67–82.
- Baddeley, A. D. (1986). *Working memory*. Oxford, UK: Clarendon Press.
- 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.

- Bates, D., Maechler, M., & Bolker, B. (2012). lme4: Linear mixed-effects models using Eigen and Eigen. *Journal of Statistical Software*, 65, 1-68.
- Biber, D. (1986). Spoken and written textual dimensions in English: Resolving the contradictory findings. *Language*, 62, 384-414.
- Boston, M. F., Hale, J. T., Vasishth, S., & Kliegl, R. (2011). Parallel processing and sentence comprehension difficulty. *Language and Cognitive Processes*, 26, 301-349.
- Box, G. E., & Cox, D. R. (1964). An analysis of transformations. *Journal of the Royal Statistical Society. Series B (Methodological)*, 26, 211-252.
- Cacioppo, J. T., & Petty, R. E. (1982). The need for cognition. *Journal of Personality and Social Psychology*, 42, 116-131.
- Cacioppo, J. T., Petty, R. E., & Kao, C. F. (1984). The efficient assessment of need for cognition. *Journal of Personality Assessment*, 48, 306-307.
- Caldwell-Harris, C. L., Berant, J. B., & Edelman, S. (2012). Measuring mental entrenchment of phrases with perceptual identification, familiarity ratings, and corpus frequency statistics. In S. T. Gries & D. Divjak (Eds.), *Frequency effects in cognitive linguistics (Vol. 1): Statistical effects in learnability, processing and change* (pp. 165-194). The Hague, The Netherlands: De Gruyter Mouton.
- Caplan, D., & Waters, G. S. (1999). WMM and sentence comprehension. *Behavioral and Brain Sciences*, 22, 77-126.
- Christiansen, M. H. & Chater, N. (2016). *Creating language: Integrating evolution, acquisition, and processing*. Cambridge, MA: MIT Press.
- Christiansen, M. H. & Chater, N. (in press). The Now-or-Never bottleneck: A fundamental constraint on language. *Behavioral & Brain Sciences*. doi:10.1017/S0140525X1500031X
- Cohen, J. D., MacWhinney, B., Flatt, M., & Provost, J. (1993). PsyScope: An interactive graphic system for designing and controlling experiments in the psychology laboratory using Macintosh computers. *Behavioral Research Methods, Instruments, and Computers*, 25, 257-271. Cambridge, UK: Cambridge University Press.
- Conway, A. R., Kane, M. J., Bunting, M. F., Hambrick, D. Z., Wilhelm, O., & Engle, R. W. (2005). Working memory span tasks: A methodological review and user's guide. *Psychonomic Bulletin & Review*, 12(5), 769-786.
- 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.
- Daneman, M., & Carpenter, P. A. (1980). Individual differences in working memory and reading. *Journal of Verbal Learning and Verbal Behavior*, 19, 450-466.
- Daneman, M., & Merikle, P. M. (1996). Working memory and language comprehension: A meta-analysis. *Psychonomic Bulletin and Review*, 3, 422-433.
- Demberg, V., & Keller, F. (2008). Data from eye-tracking corpora as evidence for theories of syntactic processing complexity. *Cognition*, 109(2), 193-210.
- Dixon, P., Lefevre, J., & Twilley, L. C. (1988). Word knowledge and working memory as predictors of reading skill. *Journal of Educational Psychology*, 80, 465-472.
- Elman, J. L. (1990). Finding structure in time. *Cognitive Science*, 14, 179-211.
- Ettlinger, M., Morgan-Short, K., Faretta-Stutenberg, M., & Wong, P. C. M. (in press). The relationship between artificial and second language learning. *Cognitive Science*.
- Farmer, T. A., Misyak, J. B., & Christiansen, M. H. (2012). Individual differences in sentence processing. In M. J. Spivey, K. McRae, & M. F. Joanisse (Eds.), *Cambridge handbook of psycholinguistics* (pp. 353-364). Cambridge, UK: Cambridge University Press.
- Frazier, L., & Fodor, J. D. (1978). The sausage machine: A new two-stage parsing model. *Cognition*, 6, 291-325.
- Friedman, N. P., & Miyake, A. (2004). The reading span test and its predictive power for reading comprehension ability. *Journal of Memory and Language*, 51, 136-158.
- Frost, R., Siegelman, N., Narkiss, A., & Afek, L. (2013). What predicts successful literacy acquisition in a second language? *Psychological Science*, 24, 1243-1252.
- Garnsey, S., Pearlmutter, N., Myers, E., & Lotocky, M. (1997). The contributions of verb bias and plausibility to the comprehension of temporarily ambiguous sentences. *Journal of Memory and Language*, 37, 58-93.
- Gennari, S. P., Mirkovic, 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.
- Gordon, P. C., Hendrick, R., & Johnson, M. (2004). Effects of noun phrase type on sentence complexity. *Journal of Memory and Language*, 51, 97-114.
- Hale, J. (2001). A probabilistic Earley parser as a psycholinguistic model. In *Proceedings of the second meeting of the North American Chapter of the Association for Computational Linguistics on Language technologies* (pp. 1-8). Stroudsburg, PA: Association for Computational Linguistics.
- Hayes, D. P. (1988). Speaking and writing: Distinct patterns of word choice. *Journal of Memory and Language*, 27, 572-585.
- Hofmeister, P., Casasanto, L. S., & Sag, I. A. (2012). How do individual cognitive differences relate to acceptability judgments? A reply to Sprouse, Wagers, and Phillips. *Language*, 88, 390-400.
- Huettig, F., & Brouwer, S. (2015). Delayed Anticipatory Spoken Language Processing in Adults with Dyslexia-Evidence from Eye-tracking. *Dyslexia*, 21, 97-122.
- Husain, S., Vasishth, S., & Srinivasan, N. (2014). Strong Expectations Cancel Locality Effects: Evidence from Hindi. *PLoS One*, 9(7), e100986.
- Jackendoff, R. (2007). A parallel architecture perspective on language processing. *Brain Research*, 1146, 2-22.
- James, A. N., & Watson, D. G. (2013). *Language experience and prediction in auditory sentence comprehension*. Poster presented at the 19th Annual Conference on Architectures and Mechanisms for Language Processing, Marseille, France.
- James, A. N., & Watson, D. G. (2014). *Language experience predicts eye movements in the visual world: An individual differences investigation*. Poster presented at the 28th Annual Meeting of the CUNY Conference on Human Sentence Processing, Los Angeles, CA.
- Jäger, L., Chen, Z., Li, Q., Lin, C. J. C., & Vasishth, S. (2015). The subject-relative advantage in Chinese: Evidence for expectation-based processing. *Journal of Memory and Language*, 79-80, 97-120.
- Jones, G., & Macken, B. (2015). Questioning short-term memory and its measurement: Why digit span measures long-term associative learning. *Cognition*, 144, 1-13.
- Jurafsky, D. (1996). A probabilistic model of lexical and syntactic access and disambiguation. *Cognitive Science*, 20, 137-194.

- Just, M. A., & Carpenter, P. A. (1992). A capacity-based theory of comprehension: New frontiers of evidence and arguments. *Psychological Review*, 103, 773–780.
- Just, M. A., Carpenter, P. A., & Woolley, J. D. (1982). Paradigms and processes in reading comprehension. *Journal of Experimental Psychology: General*, 111, 228–238.
- 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.
- Kemtes, K. A., & Kemper, S. (1997). Younger and older adults' on-line processing of syntactically ambiguous sentences. *Psychology and Aging*, 12(2), 362–371.
- Kidd, E. (2012). Implicit statistical learning is directly associated with the acquisition of syntax. *Developmental Psychology*, 48, 171–184.
- Kidd, E. & Arciuli, J. (in press). Individual differences in statistical learning predict children's comprehension of syntax. *Child Development*.
- Kidd, E., Brandt, S., Lieven, E., & Tomasello, M. (2007). Object relatives made easy: A crosslinguistic comparison of the constraints influencing young children's processing of relative clauses. *Language and Cognitive Processes*, 22, 860–897.
- Kimball, J. (1975). Predictive analysis and over-the-top parsing. In J. Kimball (Eds.), *Syntax and Semantics* (pp. 155–179). New York: Academic Press.
- King, J., & Just, M. A. (1991). Individual differences in syntactic processing: The role of working memory. *Journal of Memory and Language*, 30, 580–602.
- Kuperman, V., & Van Dyke, J. A. (2011). Effects of individual differences in verbal skills on eye-movement patterns during sentence reading. *Journal of Memory and Language*, 65, 42–73.
- Levy, R. (2008). Expectation-based syntactic comprehension. *Cognition*, 106, 1126–1177.
- Lewis, R. L. (1996). Interference in short-term memory: The magical number two (or three) in sentence processing. *Journal of Psycholinguistic Research*, 25, 93–115.
- 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., & Christiansen, M. H. (2002). Reassessing working memory: Comment on Just and Carpenter (1992) and Waters and Caplan (1996). *Psychological Review*, 109, 35–54.
- MacDonald, M. C., Just, M. A., & Carpenter, P. A. (1992). Working memory constraints on the processing of syntactic ambiguity. *Cognitive Psychology*, 24, 56–98.
- MacDonald, M., Pearlmutter, N., & Seidenberg, M. (1994). The lexical nature of syntactic ambiguity resolution. *Psychological Review*, 101, 676–703.
- Mani, N., & Huettig, F. (2014). Word reading skill predicts anticipation of upcoming spoken language input: A study of children developing proficiency in reading. *Journal of Experimental Child Psychology*, 126, 264–279.
- McCauley, S. M., & Christiansen, M. H. (2015). Individual differences in chunking ability predict on-line sentence processing. In D. C. Noelle, & R. Dale (Eds.), *Proceedings of the 37th Annual Conference of the Cognitive Science Society* (pp. 1553–1558). Austin, TX: Cognitive Science Society.
- McDonald, S. A., & Shillcock, R. C. (2003). Eye movements reveal the on-line computation of lexical probabilities during reading. *Psychological Science*, 14, 648–652.
- Mishra, R. K., Singh, N., Pandey, A., & Huettig, F. (2012). Spoken language-mediated anticipatory eye movements are modulated by reading ability: Evidence from Indian low and high literates. *Journal of Eye Movement Research*, 5, 1–10.
- Misyak, J. B., & Christiansen, M. H. (2012). Statistical learning and language: An individual differences study. *Language Learning*, 62, 302–331.
- Misyak, J. B., Christiansen, M. H. & Tomblin, J. B. (2010a). On-line individual differences in statistical learning predict language processing. *Frontiers in Psychology*, 1, 31. doi:10.3389/fpsyg.2010.00031
- Misyak, J. B., Christiansen, M. H., & Tomblin, J. B. (2010b). Sequential expectations: The role of prediction-based learning in language. *Topics in Cognitive Science*, 2, 138–153.
- Misyak, J. B., Goldstein, M. H., & Christiansen, M. H. (2012). Statistical-sequential learning in development. In P. Rebuschat & J. N. Williams (Eds.), *Statistical learning and language acquisition* (pp. 13–54). Berlin: Mouton de Gruyter.
- Montag, J. L., & MacDonald, M. C. (2015). Text exposure predicts spoken production of complex sentences in eight and twelve year old children and adults. *Journal of Experimental Psychology: General*, 144, 447–468.
- Moore, M., & Gordon, P. C. (2015). Reading ability and print exposure: Item response theory analysis of the author recognition test. *Behavior Research Methods*, 47(4), 1095–1109.
- Nelson, N. S., & Denny, E. D. (1960). *The Nelson-Denny Reading Test*. Boston, MA: Houghton Mifflin.
- Nicenboim, B., Vasishth, S., Gattei, C., Sigman, M., & Kliegl, R. (2015). Working memory differences in long-distance dependency resolution. *Frontiers in Psychology*, 6.
- Novick, J. M., Thompson-Schill, S. L., & Trueswell, J. C. (2008). Putting lexical constraints into the visual-world paradigm. *Cognition*, 107, 850–903.
- Payne, B. R., Gao, X., Noh, S. R., Anderson, C. J., & Stine-Morrow, E. A. L. (2012). The effects of print exposure on sentence processing and memory among older adults: Evidence for efficiency and reserve. *Aging, Neuropsychology, and Cognition*, 19, 122–149.
- Payne, B. R., Grison, S., Gao, X., Christianson, K., Morrow, D. G., & Stine-Morrow, E. A. L. (2014). Aging and individual differences in binding during sentence understanding: Evidence from temporary and global syntactic attachment ambiguities. *Cognition*, 130, 157–173.
- Pearlmutter, N. J., & MacDonald, M. C. (1995). Individual differences and probabilistic constraints in syntactic ambiguity resolution. *Journal of Memory and Language*, 34, 521–542.
- R Development Core Team. (2014). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. ISBN 3-900051-07-0, URL <http://www.R-project.org>.
- Rankin, J. L. (1993). Information processing differences of college-age readers differing in reading comprehension and speed. *Journal of Reading Behavior*, 25, 261–278.
- 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.
- 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.
- Romberg, A. R., & Saffran, J. R. (2010). Statistical learning and language acquisition. *Wiley Interdisciplinary Reviews: Cognitive Science*, 1, 906–914.

- Rommers, J., Meyer, A. S., & Huettig, F. (2015). Verbal and nonverbal predictors of language-mediated anticipatory eye movements. *Attention, Perception, & Psychophysics*, *77*, 720–730.
- Shipley, W. C. (1940). A self-administered scale for measuring intellectual impairment and deterioration. *Journal of Psychology*, *9*, 371–377.
- Sprouse, J., Wagers, M., & Phillips, C. (2012). A test of the relation between working-memory capacity and syntactic island effects. *Language*, *88*, 82–123.
- Stanovich, K. E., & West, R. F. (1989). Exposure to print and orthographic processing. *Reading Research Quarterly*, *24*, 402–433.
- Swets, B., Desmet, T., Hambrick, D. Z., & Ferreira, F. (2007). The role of working memory in syntactic ambiguity resolution: A psychometric approach. *Journal of Experimental Psychology: General*, *136*, 64–81.
- 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.
- Waters, G. S., & Caplan, D. (1996a). The capacity theory of sentence comprehension: Critique of Just and Carpenter (1992). *Psychological Review*, *103*, 761–772.
- Waters, G. S., & Caplan, D. (1996b). The measurement of verbal working memory capacity and its relation to reading comprehension. *The Quarterly Journal of Experimental Psychology Section A*, *49*, 51–79.
- Waters, G. S., & Caplan, D. (1996c). Processing resource capacity and the comprehension of garden path sentences. *Memory & Cognition*, *24*, 342–355.
- Wechsler, D. (1981). *The Wechsler Adult Intelligence Scale—Revised*. New York: Psychological Corporation.
- Wells, J. B., Christiansen, M. H., Race, D. S., Acheson, D. J., & MacDonald, M. C. (2009). Experience and sentence processing: Statistical learning and relative clause comprehension. *Cognitive Psychology*, *58*, 250–271.
- West, R. F., Stanovich, K. E., & Mitchell, H. R. (1993). Reading in the real world and its correlates. *Reading Research Quarterly*, *28*, 35–50.