Rapid Serial Visual Presentation Interacts with Ambiguity During Sentence Comprehension

Sarah E. Key-DeLyria  
Portland State University, s.keydelyria@pdx.edu

Todd Bodner  
Portland State University, tbodner@pdx.edu

Lori J. P. Altmann  
University of Florida

Follow this and additional works at: https://pdxscholar.library.pdx.edu/psy_fac

Citation Details

This Post-Print is brought to you for free and open access. It has been accepted for inclusion in Psychology Faculty Publications and Presentations by an authorized administrator of PDXScholar. Please contact us if we can make this document more accessible: pdxscholar@pdx.edu.
Rapid Serial Visual Presentation interacts with ambiguity during sentence comprehension

Sarah E. Key-DeLyria¹, Todd Bodner², and Lori J. P. Altmann³

¹Speech and Hearing Sciences Department, Portland State University, Portland, Oregon, USA
²Department of Psychology, Portland State University, Portland, Oregon, USA
³Department of Speech, Language, and Hearing Sciences, University of Florida, Gainesville, Florida, USA

October 2018

Funding: This study was not funded.

Conflict of Interest: The authors declare that they have no conflict of interest.

Corresponding Author:
Sarah Key-DeLyria
Speech and Hearing Sciences Department
P.O. Box 751 (SPHR)
Portland State University
Portland, OR 97207-0751

keydel@pdx.edu
Rapid Serial Visual Presentation interacts with ambiguity during sentence comprehension
Abstract

Conventional opinion about using Rapid Serial Visual Presentation (RSVP) for examining sentence comprehension maintains that RSVP taxes working memory (WM), which probably affects sentence processing. However, most RSVP studies only infer the involvement of WM. Other cognitive resources, such as cognitive control or vocabulary may also impact sentence comprehension and interact with RSVP. Further, sentence ambiguity is predicted to interact with RSVP and cognitive resources to impact sentence comprehension. To test these relationships, participants read ambiguous and unambiguous sentences using RSVP and Whole-Sentence presentation, followed by comprehension questions that were targeted to the ambiguous region of the sentences. Presentation type and ambiguity interacted to affect RT such that the effect of RSVP was exaggerated for ambiguous sentences. RT effects were moderated by WM and vocabulary. WM and cognitive control affected accuracy. Findings are discussed in light of depth of processing and the impact of cognitive resources on sentence comprehension.

Keywords
Rapid Serial Visual Presentation, sentence processing, shallow processing, working memory, executive function
Introduction

Rapid Serial Visual Presentation (RSVP) of sentences is a method in which the words in a sentence are briefly presented, one at a time. Conventional opinion about using RSVP for examining sentence comprehension maintains that RSVP taxes working memory (WM), which affects sentence processing (Caplan & Waters, 1999; Miyake, Carpenter, & Just, 1994; Waters & Caplan, 1996). WM is defined broadly as the ability to actively store and manipulate information according to situational demands (Caplan & Waters, 2005; Engle, Kane, & Tuholski, 1999; Salthouse, 1991). RSVP is used specifically to manipulate the availability of verbal information to WM (e.g. Schremm, Horne, & Roll, 2016). Though several studies have discussed the potential relationship of RSVP to cognitive processing and its differences from Whole-Sentence processing, few have directly explored the relationship (Lee & Newman, 2010; Schotter, Tran, & Rayner, 2014; Waters & Caplan, 1996). In this study, the sentence comprehension of matched ambiguous and unambiguous sentences is directly compared between RSVP and Whole-Sentence presentation methods and is examined with regard to interactions with WM, cognitive control, and vocabulary tasks.

The presumed relationship between RSVP and WM is theoretically straightforward. In RSVP, the reader is unable to look back at previous words in a sentence to make judgments about grammaticality or content. Therefore, WM would be required in order to maintain the words in short-term memory and perform any syntactic or semantic manipulations necessary. Early studies demonstrated that RSVP reading was affected by linguistic complexity and that recall for passages presented using RSVP was poor (Forster, 1970; Potter, 1980). More recent studies have used time manipulations to test the contributions of WM to different types of sentence processing. However, WM and other cognitive abilities are not frequently tested
separately from passage recall or sentence comprehension tasks (Lee & Newman, 2010; Potter, 1984). Therefore, there is no direct comparison of RSVP to whole-sentence processing and no evidence that the linguistic complexity effect observed during RSVP may be mediated by WM or other cognitive abilities.

A recent study highlights the need to validate the critical impact of WM on sentence comprehension during RSVP. Schremm et al. (2016) used RSVP to manipulate the time between sentence elements. They found that increasing the temporal distance between words (i.e. slowed rate) slowed acceptability, or acceptability judgments for object-relative sentences containing grammatical errors or semantic incongruences. The authors interpret their results as support for the notion that grammatical information decays in WM after 2-3 seconds. However, if grammatical information had decayed, accuracy would also be expected to decline, and it did not. Accuracy rates were all above 90% and did not differ between presentation rates. Finally, individual differences in goodness judgment RTs could be related to a number of interrelated or unrelated factors, including WM, conflict resolution, familiarity with object-relative structures, and other factors. Their methodology did not allow these possibilities to be addressed.

A number of prior studies have acknowledged the potential problems associated with using RSVP and have attempted to clarify how RSVP processing differs from Whole-Sentence processing and whether it relates to WM. Lee and Newman (2010) found indirect evidence for an interaction between presentation type and WM using fMRI. They contrasted activation for low complexity, conjoined active sentences and high complexity, object relative sentences using RSVP and Whole-Sentence presentation. They also measured accuracy and response times (RTs) for verification of probe sentences that described an event from a target sentence. Accuracy but not RT varied significantly between complexity conditions and presentation conditions, but a
significant interaction revealed that there was a larger effect of complexity in the RSVP condition. Based on the accuracy findings and differential activation patterns observed on fMRI, the authors suggested that RSVP prevented readers from showing processing benefits in the simpler sentences, because the serial order processing required by RSVP interfered with either syntactically-relevant word order processing or the recruitment of memory for deeper encoding of syntactic and semantic relationships. Unfortunately, a few methodological issues may have increased WM demands and influenced performance. Thus, there is still a need for a direct comparison of comprehension following RSVP and Whole-Sentence presentation, as well as a direct test of the relationship between comprehension and WM in these two conditions.

Waters and Caplan (1996) also explored the notion that RSVP might tap WM resources using garden-path and control sentences with semantic anomalies that contained sentential complements, reduced relatives, or embedded clauses, in older adults. They examined the effects of WM by splitting their group into three smaller groups based on reading span performance, which was the only cognitive measure reported. The authors reported that there was no differential effect of RSVP on garden-path structures and no differential effects of WM on processing. Therefore, while Waters and Caplan did find RSVP was more difficult overall, they also found that this difficulty was not exaggerated when sentences used required more processing, which was quite unexpected. Waters and Caplan did not explicitly test sentence comprehension but instead assessed acceptability judgments, like Shremm et al. (2016). Furthermore, specific details of the methods employed in Waters and Caplan may have impacted results, particularly with regard to RTs.

Most previous studies using RSVP have used goodness or acceptability judgments or sentence verification prompts that would not identify errors in comprehension. Schotter and
colleagues (2014) tested comprehension of sentences similar to the sentence types used in our
study and restricted the ability of readers to re-read previously read portions of a sentence by
using a trailing-mask paradigm. They found that readers had lower accuracy in the trailing-mask
condition than the whole-sentence condition, regardless of sentence ambiguity. The trailing-mask
paradigm is similar to RSVP in that it prevents re-reading or eye regressions.

A growing number of studies suggest that people may process sentences shallowly,
without determining the exact meaning of the sentence. Moreover, this happens more often than
previously believed (Christianson, Hollingworth, Halliwell, & Ferreira, 2001; Christianson,
Christianson and colleagues (2001), developed a set of sentences and comprehension questions
specifically targeting comprehension of garden path sentences. They tested the persistence of the
initial, incorrect interpretation with questions targeting the reinterpretation step. They found that
participants do partially reanalyze garden path sentences to come to the correct interpretation but
also maintain the ungrammatical initial interpretation. This finding has implications for the
interpretation of studies examining sentence processing using RSVP and critically, it suggests
the need to be testing comprehension rather than simply acceptability judgments.

In addition to differences in presentation methodology, individual differences in language
experience and cognition may also affect sentence processing and comprehension. Individual
differences in WM, as measured by reading span tasks (Daneman & Carpenter, 1980, Waters &
Caplan, 1996) or operation span tasks (Conway et al., 2005) may also impact the success of
sentence processing. The theoretical need for WM in sentence processing is straightforward
because the meaning and structure of a sentence must be constructed over time. Debate continues
whether there is a language-specific WM resource (Caplan & Waters, 1999; Kane & Engle,
2002) or whether language relies on a domain-general WM resource (Kane & Engle, 2002) (see Caplan & Waters 1999; 2013; Just & Carpenter, 1992; Martin, 2006). A number of studies have found a link between sentence processing using a variety of sentence measures and WM measures (e.g. Baddeley, 2003; Farmer, Misyak, & Christiansen, 2012; Kempler, Almor, Tyler, Andersen, & MacDonald, 1998; MacDonald & Christiansen, 2002; Swets, Desmet, Hambrick, & Ferreira, 2007). Increased language experience, particularly with more complex or rare syntactic structures, may lead to more efficient sentence processing of difficult structures (MacDonald & Christiansen, 2002; Waters & Caplan, 1996; Wells, Christiansen, Race, Acheson & MacDonald, 2009).

Cognitive control is a more recent cognitive component that has been explored as one potential difference between individuals that may affect sentence processing (Fedorenko, 2014; Key-Delyria & Altmann, 2016; Martin, Yan, & Schnur, 2014; Novick, Hussey, Teubner-Rhodes, Harbison, & Bunting, 2013; Novick, Trueswell, & Thompson-Schill, 2005; Nozari, Dell, & Schwartz, 2011; Thothathiri, Gagliardi, & Schwartz, 2012; Ye & Zhou, 2009), though the specifics of the relationship are not well understood (see Key-DeLyria & Altmann, 2016 for a review). One possibility is that cognitive control or other executive functions are recruited when sentence comprehension is more difficult as during ambiguous sentence processing. Thus, better cognitive control ability should lead to faster and more accurate ambiguous sentence processing and comprehension. In a study using stimuli similar to Christianson et al. (2001), Vuong and Martin (2014) compared sentence processing performance (grammaticality judgments) to verbal and nonverbal Stroop tasks which are purported to measure inhibition as well as a number of other executive functions (Key-DeLyria & Altmann, 2016). They found that participants who made more grammaticality judgment errors also had more interference errors on the verbal
Stroop. This suggests that people with poor verbal inhibition may have more difficulty recognizing grammatical errors.

In the current study, we predict that both WM and cognitive control will relate to sentence comprehension. Higher WM performance should moderate ambiguous sentence comprehension due to the need to remember and reconsider the early part of the sentence once the ambiguity is encountered. On the other hand, higher WM may not be related to comprehension of sentences that are not ambiguous because there is less of a need to update memory over time. Further, ambiguous sentence comprehension, but not unambiguous sentence comprehension, is expected to relate to cognitive control as measured by a verbal Stroop task. Moreover, we expect that these relationships will interact with presentation method, such that WM will be more necessary during RSVP than during Whole Sentence presentation. On the other hand, the existing literature does not address whether cognitive control would play in role in sentence comprehension during RSVP. There are not obvious overlaps in task demands between RSVP and the inhibition/revision processes of cognitive control beyond perhaps increasing the difficulty of revising sentence interpretations when encountering an ambiguity.

In summary, this study aims to examine how differences in sentence presentation method and cognition relate to comprehension. We specifically test comprehension using comprehension questions and do not use acceptability judgments as in most previous studies using RSVP. We hypothesize that participants would respond to comprehension questions following sentences presented using RSVP more slowly and less accurately than when using Whole-Sentence presentation and that this effect would be exaggerated for ambiguous sentences. Furthermore, we hypothesize that sentence presentation and ambiguity would interact with measures of WM,
cognitive control, and vocabulary and that these cognitive resources would moderate sentence comprehension performance.

Methods

Twenty-three undergraduates (ages 18-23, 4 males) participated for course credit at the University of Florida. There were more female than male participants due to a similar imbalance in enrollment in the courses where students were recruited. Participants reported no significant neurological, developmental, or uncorrected visual impairment. Participants completed WM, cognitive control, vocabulary and experimental sentence tasks.

A computerized Operation-Span task (O-Span; Conway et al., 2005) tested complex WM. First, 2-5 letters are presented individually. Next, participants verify whether simple mathematical operations (e.g. \(8 + 7 = 12\)) are accurate. Accuracy must remain above 85% or scores are not used (four participants did not maintain this accuracy level resulting in their scores not being included in further analyses). Finally, participants rank an array of letters on the screen in the order they were presented prior to the intervening math verification. A score is computed that is the total number of accurate items in the correct rank.

Other tasks were chosen that balanced a need to measure a range of different cognitive abilities but also did not add significantly to the time requirements for participation in the study. Tasks chosen are commonly included in neuropsychological screening batteries. Digit Span-Forward (DSF) and Digit Span-Backward (DSB) (Wechsler, 1997) tested short-term memory and attention. Participants receive a point for each perfect repetition of increasingly long lists of digits verbatim or in reverse order.

The Trail Making Task (Reitan, 1958) and the Stroop task (Stroop, 1935) tested cognitive control. The Trail Making Task consists of two timed subtasks, Trails A and Trails B, and is
thought to measure visual processing speed and ability to shift sets (an aspect of executive function) (Key-DeLyria & Altmann, 2016). In Trails A, participants connect numbers 1-25, in sequence, randomly distributed across a paper. In Trails B, participants switch between numbers 1-13 and letters A-L in sequence (e.g. (1)-(A)-(2)-(B)). The difference in completion time between Trails A and Trails B is thought to uniquely measure shifting ability. The Stroop task used consists of two timed subtasks, XX’s and Color Words. In both subtasks, participants name the color of the ink used to print the XX’s or the words. In the XX’s subtask, the number of sets of colored XX’s named in 45 seconds is measured. In the Color Words subtask, the printed color of the word and color are incongruent (e.g. the word Red printed in green ink). The number of colors named in 45 seconds is again measured. The difference between the number named in each subtask represents cognitive control.

In the Shipley Vocabulary test, participants choose the best synonym out of four choices for each of 40 words (Shipley, 1940). Sample items for the test are in Appendix 1. Scores are a total number of a correct responses. Participants read aloud words of increasing phonological complexity and decreasing lexical frequency in the North American Adult Reading Test (NAART; Nelson, 1982). Higher scores typically reflect higher vocabularies. Scores on all cognitive tasks are shown in Table 1.

--- Table 1 ---

**Sentence comprehension tasks**

Sentence tasks used 120 sentences, 9-11 words long, half ambiguous and half unambiguous following Christianson et al. (2001). In addition to 42 sentence pairs from Appendix 1 in Christianson et al. (2001), 18 additional sentence pairs using novel verbs and nouns were designed to meet the same criteria. Ambiguous sentences were designed to lead
participants to make an incorrect initial interpretation that required revision, and unambiguous sentences were created by adding a disambiguating comma. Filler sentences were not used due to time constraints in participant testing. In order to account for the development of strategies, order of presentation was included in the linear mixed effects models.

(a) Ambiguous: *While the woman bathes the baby spits up on the bed.*

(b) Unambiguous: *While the woman bathes, the baby spits up on the bed.*

Comprehension questions were designed to test correct interpretation of the sentence and targeted different portions of sentence content. The comprehension questions in Appendix 1 of Christianson et al. (2001) followed the pattern of question type (c) below to target the incorrect interpretation of the ambiguity and were used. Additional questions were constructed that followed the pattern of the other question types to target correct resolution of the ambiguity, as in (d), and to target comprehension of the sentence-final clause, as in (e). Questions for a sentence were the same regardless of presentation type or ambiguity condition.

(c) ‘No’ Comprehension question: *Did the woman bathe the baby?*

(d) ‘Yes’ Comprehension question: *Did the woman bathe?*

(e) Yes’ Comprehension question: *Did the baby spit up on the bed?*

Half of the questions presented to each participant required a ‘yes’ response. Sentence and question types were counterbalanced. Sentences were divided into four lists total. There were two lists each of 60 items, one which was presented in the Whole Sentence condition, the other in RSVP. Within each presentation condition, half of the sentences were disambiguated with a comma. This means that while participants saw the ambiguous and unambiguous version of each sentence pair, they did not see each version of the sentence within the same presentation condition. These four lists were counterbalanced across participants. Question types were also
counterbalanced such that any participant did not see the same question twice, and each participant answered half ‘Yes’ and half ‘No’ questions.

Presentation type (Whole-Sentence vs. RSVP) was blocked and order counterbalanced. Stimuli were presented on a computer screen in a light-colored, 16-point font on a dark background. Sentence trials began with a 1000 ms blank screen with a centered fixation cross. In the Whole-Sentence condition, sentences were then presented in their entirety slightly above the fixation cross for 5000 ms. Some sentences were split between two lines on the monitor, and these lines were both above the fixation cross. In the RSVP condition, single words were each presented slightly above the centered fixation for 300 ms followed by a 200 ms blank screen. Presentation rate reflects a slightly slower than normal reading rate for young adults (Hagoort & Brown, 2000) while matching timing between conditions. Next, in both conditions, a 500 ms blank screen with a centered fixation cross appeared. A comprehension question then appeared slightly below the centered fixation until the participant’s response. Finally, a 300 ms blank screen appeared after which participants pressed a mouse button to begin a new trial. Prior to starting the experiment, participants were instructed to “read to understand” each sentence and to answer comprehension questions as accurately and quickly as possible. Six practice sentences with comprehension questions were given to confirm that participants understood the instructions. Practice sentences were the same for each participant and followed the same structure as the experimental sentences (half were unambiguous, half were ambiguous).

Results

Prior to analysis, participant RTs for incorrect responses and RTs less than 500ms were excluded from further analyses. Participant RTs less than 500 ms were assumed to be machine error. This cutoff was determined prior to data inspection, and only 2 ambiguous and 2
unambiguous trials out of 2381 total accurate trials were below this cutoff. No upper limit for RT was set to allow participants to fully read comprehension questions and respond accurately. An upper limit of 2.5 SD above the mean was inspected but did not change the main interaction pattern, probably because few trials were above this cutoff. These RTs were not removed from the main analyses. Christianson et al. (2001) do not report a lower or upper limit for RTs and report means between 2.14 and 3.19 sec and SDs between .44 and 2.67. Descriptive statistics for the dependent variables, RT and accuracy, appear in Table 2 and are comparable to previous research.

--- Table 2 ---

Following the recommendations of Locker, Hoffman, and Bovaird (2007), inference for the experimental and moderated experimental effects on these outcomes were conducted with generalized linear mixed effects models that included both participant and trial random effects. An identity link function was used in the models for the RT outcome; a logit link function was used in the models for the accuracy outcome. All analyses were conducted using SPSS 22 (IBM Corp., 2013). To improve parameter interpretability, RT was rescaled from milliseconds to seconds.

**Accuracy**

Table 3 provides the model results for experimental and moderated experimental effects on the accuracy outcome. As appears in the first column of Table 3, RSVP resulted in a non-significant reduction in the log-odds of an accurate response compared to Whole Sentence presentation ($b = -0.05, p = .79$). Similarly, stimulus ambiguity resulted in a non-significant reduction in the log-odds of an accurate response compared to stimulus unambiguity ($b = -0.10, p = .61$). The interactive effect of these experimental factors on the log-odds of an accurate
response, however, approached statistical significance ($b = -0.52$, $p = .056$). Descriptively, RSVP reduced the log-odds of an accurate response, relative to Whole Sentence presentation, to a greater extent when stimuli were ambiguous than when stimuli were unambiguous, but performance in the Whole Sentence conditions were unaffected by ambiguity. This interactive effect, transformed to the probability metric, is displayed in Figure 1.

---

Table 3 about here   Figure 3 about here
---

The rightmost columns in Table 3 provide the results of models testing whether individual differences in cognitive performance moderate the effect of presentation type or ambiguity on the log-odds of accurate responding. Due to the greater estimation precision when controlling for the effects of these cognitive performance variables, the Presentation Type X Ambiguity interaction shown in Figure 1 that approached statistical significance in the base model reached statistical significance in the models that included Stroop Interference ($b = -.547$, $p = .050$), Shipley vocabulary ($b = -.547$, $p = .045$), NAART ($b = -.543$, $p = .046$), and DSF ($b = -.550$, $p = .044$) when evaluated at the grand mean of these cognitive performance variables.

Whether cognitive performance measures moderated experimental effects was also examined. Only the Trails Difference score significantly moderated the effect of Ambiguity on accuracy ($b = .013$, $p = .049$). This interaction is shown in Figure 2. People with lower Trails Difference scores (i.e., more able to suppress interference) were more accurate on unambiguous sentences than unambiguous sentences, while those with greater Trails Difference scores showed lower performance on both types of sentence with a much reduced effect of ambiguity. Although
outside the primary focus of this study, we also noted that higher overall accuracy rates on the experimental sentence task corresponded to lower Trails Difference scores ($b = -.022, p = .023$), lower Stroop Interference scores ($b = -.032, p = .043$), and higher DSB scores ($b = .272, p = .049$). The correspondence between higher Operation Span scores and higher accuracy rates approached but did not reach significance ($b = .024, p = .057$). No other significant effects were observed.

--- Figure 2 about here ---

**Reaction Time**

Table 4 provides the model results for experimental and moderated experimental effects on the RT outcome. As appears in the first column of Table 4, responses following RSVP presentation were significantly slower than those following Whole Sentence presentation ($b = 0.32, p = .020$). In contrast, the ambiguity manipulation did not result in a significant mean RT differences ($b = -0.09, p = .546$). A significant Presentation Type X Stimulus Ambiguity interaction, however, qualified these simple effects ($b = 0.43, p = .038$). Responses following RSVP were significantly slower following ambiguous sentences than unambiguous sentences; however, ambiguity of stimuli had no significant effect on responses following Whole Sentence presentation. This interaction is shown in Figure 3.

--- Table 3 about here ---

The rightmost columns in Table 3 provide the results of models testing whether individual differences in cognitive performance moderate these experimental effects. In these models, we note that the results discussed above are largely the same, which is expected given that the cognitive performance measures are grand-mean centered. The non-significance of the effects of presentation and ambiguity in the Operation Span model may be due to a reduction in
statistical power due to the smaller number of participants with valid Operation Span measurements (i.e., N = 19 rather than N = 23); the non-significance of the Presentation Type X Ambiguity interaction in the model with the Stroop measure (i.e., $p = .054$) just misses the criterion significance level of $\alpha = .05$. Of primary interest was whether cognitive performance moderated the experimental effects. Operation Span significantly moderated the effect of Presentation Type on RT ($b = 0.02, p = .023$). As shown in Figure 4, the effect of RVSP compared to Whole Sentence presentation resulted in significantly slower mean RTs for participants with higher than lower Operation Span scores. NAART significantly moderated the effect of Presentation Type on RT ($b = 0.06, p = .011$). As shown in Figure 5, people with lower NAART scores showed little difference in RT between presentation types, whereas those with higher NAART scores responded significantly faster in the Whole Sentence condition than in RSVP. Interestingly, RTs following RSVP were similar for participants with both higher and lower NAART scores. Although not the focus of the present study, we note that higher Stroop interference scores corresponded with significantly slower RTs overall ($b = 0.04, p = .025$), while higher Shipley vocabulary scores corresponded with significantly faster RTs overall ($b = -0.15, p = .008$). No other significant effects were observed.

Discussion

The goal of this study was to examine how sentence presentation methods and cognition relate to comprehension of ambiguous versus unambiguous sentences. We predicted that responses to comprehension questions following RSVP would be slower and less accurate than
following Whole-Sentence presentation. Further, we predicted that these effects would be exaggerated for ambiguous versus unambiguous sentences and that there would be interactions with WM, cognitive control, and vocabulary. Our predictions were partially supported and the results related to each prediction are discussed in turn below.

As predicted, and consistent with most of the reviewed RSVP research, responses to comprehension questions following RSVP were significantly slower overall than Whole-Sentence presentation. This finding is consistent with previous studies (Lee & Newman, 2010; Schremm et al., 2016; Waters & Caplan, 1996) that report that presentation methodologies affect sentence comprehension, even though the current study employed more in depth comprehension questions than most previous studies. Importantly, in the current study ambiguity and presentation type interacted for RT, such that the effect of RSVP was exaggerated for ambiguous sentences. The interaction effect was also present for accuracy with marginal significance without the cognitive performance measures included in the model and with statistical significance when Stroop Interference, Shipley vocabulary, NAART, or DSF included in the model. This is a novel finding and suggests that RSVP does in fact exacerbate the effects of local ambiguity on comprehension. Accuracy findings suggest that many participants did not fully reanalyze a substantial portion of the ambiguous sentences during RSVP, consistent with shallow processing studies (e.g. Christianson et al., 2001). On the other hand, our findings are inconsistent with previous research by Waters and Caplan (1996) which did not find an interaction between ambiguity and presentation type for accuracy. However, the lack of an interaction effect in that study could have been due to a number of factors as discussed above, especially task differences. Waters and Caplan asked participants to confirm whether sentences
were grammatical, whereas we asked participants to answer comprehension questions directed at the correct resolution of the ambiguity.

Our hypothesis that WM and cognitive control would moderate sentence comprehension performance was supported. Overall RTs and accuracy were predicted by a number of cognitive moderators. Overall faster RTs were predicted by better cognitive control, as measured by lower Stroop interference. Better accuracy was predicted by better cognitive control, as measured by higher Stroop and Trails scores, and better working memory, as measured by the DSB.

Additionally, better WM was associated with faster RTs in the Whole Sentence presentation condition but slower RTs in the RSVP condition, revealing perhaps a sensitivity to the higher task demands and potential need for reanalysis in this condition. This interpretation is consistent with accuracy also being high in participants with better WM. In contrast, lower WM was associated with similar RTs in both presentation conditions and lower accuracy overall. These findings are consistent with claims of the importance of WM for sentence comprehension (e.g. Baddeley, 2003; Farmer, Misyak, & Christiansen, 2012; Kempler, Almor, Tyler, Andersen, & MacDonald, 1998; MacDonald & Christiansen, 2002; Swets et al., 2007) and add further support to claims regarding the importance of cognitive control for sentence comprehension (Key-Delyria & Altmann, 2016).

Our hypothesis that WM would specifically interact with presentation type and ambiguity to moderate comprehension was partially supported. Disambiguating information is not immediately available during RSVP unlike in Whole-Sentence presentation. During RSVP, each word was available for 500 ms without variation, whereas mean fixation time in silent reading is approximately 225 ms (Rayner, 1998), thus requiring sentence-initial words to be held in WM for longer. Further, eyes can regress during Whole-Sentence reading because previous text
remains available, reducing the need for WM (Rayner, 1998). Therefore, understanding sentences presented in RSVP has been hypothesized to require reliance on WM (Lee & Newman, 2010). Our findings that Operation Span moderated the effect of presentation on RT support this view and are somewhat consistent with findings of Lee and Newman (2010). Lee and Newman found that accuracy was more affected by complexity during RSVP than during Whole-Sentence presentation and noted a relationship between accuracy and WM-related neural activity. However, they did not observe an interaction between presentation type and sentence complexity in RTs, which could have been attenuated by the 6 second delay between the sentence and the comprehension question in their study.

Our findings are also consistent with previous research reporting that people with lower WM span process ambiguity, and possibly incoming language in general, differently than those with a higher WM span (Swets et al., 2007). Swets and colleagues found that participants with low WM spans seemed to use different strategies to process ambiguous sentences than participants with high WM spans. The authors suggested that readers with lower spans are more likely to segment text due to their limited WM, leading to higher attachment of the ambiguous clause.

We also predicted that cognitive control would moderate the effects of ambiguity and presentation on accuracy and RT based on the theory that the need to inhibit and revise an incorrect interpretation are parallel to the construct of cognitive control (Fedorenko, 2014; Key-DeLyria & Altmann, 2016; Martin, Yan, & Schnur, 2014; Novick et al., 2005; Novick et al., 2013; Nozari, Dell, & Schwartz, 2011; Thothathiri et al., 2012; Vuong & Martin, 2014; Ye & Zhou, 2009). Cognitive control performance, as measured by the lower interference on the Trail Making Test, significantly moderated the effect of ambiguity on accuracy, as well as overall
accuracy. Interestingly, it was the participants who were successful at managing interference that experienced the greatest ambiguity effect on accuracy, rather than those with more limited cognitive control. Thus, this finding is consistent with our expectations but the overall moderating effect of cognitive control on ambiguity is weaker than predicted.

Finally, while we predicted that vocabulary would interact with ambiguity and presentation type, we did not make strong predictions as to the direction or strength of the relationship. We found that faster RTs corresponded, in general, to better vocabulary and that performance on the NAART moderated the presentation effect. The NAART and Shipley vocabulary scores likely relate most closely to general experience with language and may serve as a proxy for language proficiency (Braze et al., 2016; Nation & Coady, 1988). In general, the previous literature on RSVP and ambiguity processing do not include language proficiency measures. However, the current findings are consistent with the one important exception to this trend: Engelhardt, Nigg and Ferreira (2017) found that vocabulary, or other measures of general language proficiency, may be better explanatory factors for individual differences in language processing than individual measures of cognition. Due to the small number of participants in the current study, it is not possible to conduct more complex modeling to determine the relative contributions of vocabulary versus WM or cognitive control. However, our findings suggest that vocabulary is an important factor to consider in observing language comprehension performance.

A few limitations were present in this study. Our measures focused on comprehension and were only a secondary measure of sentence processing. Using a paradigm to combine processing and comprehension measures could provide a better measure of RSVP processing to relate to cognitive abilities. Also, participants reported that Whole-Sentence stimuli seemed to stay on the screen for a very long time because we used a fixed timing of 5 seconds to match
RSVP total time. However, while participant responses were quite fast in this condition, participants still did not perform at ceiling for accuracy. We did not use filler sentences in this study due to time constraints in testing participants, which could have resulted in different strategy use during reading the different sentence types. However, order of presentation was included as a random effect in the linear mixed effects model in order to minimize any effects of strategizing. Finally, and most importantly, our participant pool was quite small, so these results are preliminary and should be followed by larger studies. In our study, we offset the small sample size using a within-subjects experimental design; the moderated experimental effects using the cognitive performance measures, a between-subjects variable, will have lower statistical power when all else is equal. Despite this, we are encouraged that we found some experimental effects that were significantly moderated by cognitive performance measures. We view these results as interesting and potentially important for the field. We encourage replication and extension of these findings in future research.

This study provides an interesting look into the differences in reading ambiguous sentences during Whole-Sentence presentation versus RSVP using current best practice for linear mixed effect models that include random effects for both participants and trial order. RSVP is widely used in sentence processing research for a number of reasons, such as time-locking stimuli to an EEG recording without confounding eye movements. However, presentation methods appear to interact with ambiguity, and cognitive demands may moderate this interaction. RSVP appears to induce shallower processing in normal readers. Consequently, RSVP may offer a means to study critical issues such as predictive processing, sentence comprehension, and information retention but should be compared to typical reading constraints only with an understanding of the limitations specific to this methodology.
References


Figure 1. Interactive effect of presentation type and sentence ambiguity on response accuracy. Standard errors are represented by the error bars on each column.
Figure 2. Interactive effect of sentence ambiguity and Trails Difference scores on response accuracy. Standard errors are represented by the error bars on each column.
Figure 3. Interactive effect of presentation type and sentence ambiguity on reaction time (in ms). Standard errors are represented by the error bars on each column.
Figure 4. Interactive effect of presentation type and Operation Span on reaction time (in ms). Standard errors are represented by the error bars on each column.
Figure 5. Interactive effect of presentation type and NAART on reaction time (in ms). Standard errors are represented by the error bars on each column.
Table 1

Scores on cognitive tasks

<table>
<thead>
<tr>
<th>Task</th>
<th>Mean</th>
<th>SD</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td>NAART (max 50)</td>
<td>35.30</td>
<td>6.07</td>
<td>23</td>
</tr>
<tr>
<td>Shipley (max 40)</td>
<td>32.48</td>
<td>3.78</td>
<td>23</td>
</tr>
<tr>
<td>DSF (max 14)</td>
<td>8.30</td>
<td>1.55</td>
<td>23</td>
</tr>
<tr>
<td>DSB (max 14)</td>
<td>6.39</td>
<td>1.73</td>
<td>23</td>
</tr>
<tr>
<td>O-Span (max 50)</td>
<td>40.42</td>
<td>20.55</td>
<td>19</td>
</tr>
<tr>
<td>Trails A (seconds)</td>
<td>73.22</td>
<td>19.37</td>
<td>23</td>
</tr>
<tr>
<td>Trails B (seconds)</td>
<td>85.70</td>
<td>23.52</td>
<td>23</td>
</tr>
<tr>
<td>Trails B-A</td>
<td>12.48</td>
<td>23.72</td>
<td>23</td>
</tr>
<tr>
<td>Stroop Difference Score¹</td>
<td>27.91</td>
<td>13.93</td>
<td>23</td>
</tr>
</tbody>
</table>

¹Number of X’s read in 45 seconds — number of colors named when word and color are incongruent.

²Four participants were below the accuracy cutoff score so they did not complete the task and could not be scored.
Table 2

*Descriptive Statistics for Sentence Comprehension Measures as a Function of Experimental Condition*

<table>
<thead>
<tr>
<th></th>
<th>Whole Sentence Presentation</th>
<th></th>
<th>Rapid Serial Visual Presentation</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>SD</td>
<td>Mean</td>
<td>SD</td>
</tr>
<tr>
<td><strong>Accuracy (as a proportion)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Unambiguous Sentences</td>
<td>.89</td>
<td>.31</td>
<td>.89</td>
<td>.32</td>
</tr>
<tr>
<td>Ambiguous Sentences</td>
<td>.87</td>
<td>.34</td>
<td>.80</td>
<td>.40</td>
</tr>
<tr>
<td><strong>Reaction Time (in ms)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Unambiguous Sentences</td>
<td>2418.34</td>
<td>2049.61</td>
<td>2617.30</td>
<td>1830.24</td>
</tr>
<tr>
<td>Ambiguous Sentences</td>
<td>2217.52</td>
<td>1872.05</td>
<td>3075.93</td>
<td>2724.46</td>
</tr>
</tbody>
</table>

Note: N = 23.
<table>
<thead>
<tr>
<th>Predictor</th>
<th>Trails</th>
<th>Stroop</th>
<th>Operation Span</th>
<th>Shipley</th>
<th>NAART</th>
<th>DSF</th>
<th>DSB</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Intercept)</td>
<td>2.46</td>
<td>0.28</td>
<td>2.53</td>
<td>0.28</td>
<td>2.47</td>
<td>0.27</td>
<td>2.46</td>
</tr>
<tr>
<td>Presentation</td>
<td>-0.05</td>
<td>0.19</td>
<td>-0.09</td>
<td>0.20</td>
<td>-0.03</td>
<td>0.20</td>
<td>-0.11</td>
</tr>
<tr>
<td>(P)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ambiguity (A)</td>
<td>-0.10</td>
<td>0.20</td>
<td>-0.18</td>
<td>0.20</td>
<td>-0.11</td>
<td>0.20</td>
<td>-0.04</td>
</tr>
<tr>
<td>P*A</td>
<td>-0.52</td>
<td>0.27</td>
<td>-0.48</td>
<td>0.28</td>
<td>-0.55*</td>
<td>0.28</td>
<td>-0.53</td>
</tr>
<tr>
<td>Moderator (M)</td>
<td></td>
<td></td>
<td>-0.02*</td>
<td>0.01</td>
<td>-0.03*</td>
<td>0.02</td>
<td>0.02</td>
</tr>
<tr>
<td>M*P</td>
<td></td>
<td></td>
<td>0.01</td>
<td>0.01</td>
<td>0.00</td>
<td>0.01</td>
<td>0.01</td>
</tr>
<tr>
<td>M*A</td>
<td></td>
<td></td>
<td>0.01</td>
<td>0.01</td>
<td>0.00</td>
<td>0.01</td>
<td>0.07</td>
</tr>
<tr>
<td>M<em>P</em>A</td>
<td></td>
<td></td>
<td>-0.01</td>
<td>0.01</td>
<td>0.01</td>
<td>0.02</td>
<td>0.07</td>
</tr>
<tr>
<td>Random Effects</td>
<td>Est.</td>
<td>SE</td>
<td>Est.</td>
<td>SE</td>
<td>Est.</td>
<td>SE</td>
<td>Est.</td>
</tr>
<tr>
<td>Residual</td>
<td>1.00†</td>
<td>NA</td>
<td>1.00†</td>
<td>NA</td>
<td>1.00†</td>
<td>NA</td>
<td>1.00†</td>
</tr>
<tr>
<td>Subject</td>
<td>0.93*</td>
<td>0.32</td>
<td>0.87*</td>
<td>0.31</td>
<td>0.76*</td>
<td>0.28</td>
<td>0.78*</td>
</tr>
<tr>
<td>Trial Order</td>
<td>0.50*</td>
<td>0.17</td>
<td>0.51*</td>
<td>0.17</td>
<td>0.50*</td>
<td>0.17</td>
<td>0.51*</td>
</tr>
</tbody>
</table>

Notes: * p < .05. † SPSS constrains the residual variance to 1.00. Sentence presentation coded: 0 = Whole sentence; 1 = Rapid Serial Presentation. Sentence ambiguity coded: 0 = Unambiguous; 1 = Ambiguous. N = 19 for model including Operation Span; N = 23 otherwise.
Mixed Model Results for Main and Interactive Effects of Presentation Type, Sentence Ambiguity, and Cognitive Performance on Reaction Time

<table>
<thead>
<tr>
<th>Predictor</th>
<th>Trails</th>
<th>Stroop</th>
<th>Operation Span</th>
<th>Shipley</th>
<th>NAART</th>
<th>DSF</th>
<th>DSB</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Intercept)</td>
<td>2.42</td>
<td>0.23</td>
<td>2.44 0.24</td>
<td>2.44 0.23</td>
<td>2.54 0.26</td>
<td>2.42 0.21</td>
<td>2.42 0.24</td>
</tr>
<tr>
<td>Presentation (P)</td>
<td>0.32*</td>
<td>0.14</td>
<td>0.31* 0.14</td>
<td>0.31* 0.14</td>
<td>0.25 0.16</td>
<td>0.32* 0.14</td>
<td>0.32* 0.14</td>
</tr>
<tr>
<td>Ambiguity (A)</td>
<td>-0.09</td>
<td>0.14</td>
<td>-0.10 0.14</td>
<td>-0.09 0.14</td>
<td>-0.14 0.16</td>
<td>-0.09 0.14</td>
<td>-0.09 0.14</td>
</tr>
<tr>
<td>P*A</td>
<td>0.43*</td>
<td>0.21</td>
<td>0.43* 0.21</td>
<td>0.40 0.21</td>
<td>0.32 0.23</td>
<td>0.43* 0.21</td>
<td>0.43* 0.21</td>
</tr>
<tr>
<td>Moderator (M)</td>
<td>.</td>
<td>.</td>
<td>0.01 0.01</td>
<td>0.04* 0.02</td>
<td>-0.02 0.01</td>
<td>-0.15* 0.06</td>
<td>-0.07 0.04</td>
</tr>
<tr>
<td>M*P</td>
<td>.</td>
<td>.</td>
<td>-0.01 0.01</td>
<td>-0.02 0.01</td>
<td>0.02* 0.01</td>
<td>0.06 0.04</td>
<td>0.06* 0.02</td>
</tr>
<tr>
<td>M*A</td>
<td>.</td>
<td>.</td>
<td>-0.01 0.01</td>
<td>-0.01 0.01</td>
<td>0.01 0.01</td>
<td>0.03 0.04</td>
<td>0.04 0.02</td>
</tr>
<tr>
<td>M<em>P</em>A</td>
<td>.</td>
<td>.</td>
<td>-0.01 0.01</td>
<td>-0.02 0.01</td>
<td>0.01 0.01</td>
<td>-0.08 0.05</td>
<td>-0.03 0.03</td>
</tr>
</tbody>
</table>

Random Effects

<table>
<thead>
<tr>
<th></th>
<th>Est.</th>
<th>SE</th>
<th>Est.</th>
<th>SE</th>
<th>Est.</th>
<th>SE</th>
<th>Est.</th>
<th>SE</th>
<th>Est.</th>
<th>SE</th>
<th>Est.</th>
<th>SE</th>
<th>Est.</th>
<th>SE</th>
<th>Est.</th>
<th>SE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Residual</td>
<td>5.30*</td>
<td>0.16</td>
<td>5.26*</td>
<td>0.16</td>
<td>5.26*</td>
<td>0.15</td>
<td>5.61*</td>
<td>0.18</td>
<td>5.30*</td>
<td>0.16</td>
<td>5.28*</td>
<td>0.16</td>
<td>5.29*</td>
<td>0.16</td>
<td>5.30*</td>
<td>0.16</td>
</tr>
<tr>
<td>Subject</td>
<td>0.95*</td>
<td>0.30</td>
<td>0.99*</td>
<td>0.32</td>
<td>0.90*</td>
<td>0.29</td>
<td>1.03*</td>
<td>0.37</td>
<td>0.74*</td>
<td>0.24</td>
<td>0.97*</td>
<td>0.31</td>
<td>0.98*</td>
<td>0.32</td>
<td>0.96*</td>
<td>0.31</td>
</tr>
<tr>
<td>Trial Order</td>
<td>0.13*</td>
<td>0.05</td>
<td>0.13*</td>
<td>0.05</td>
<td>0.13*</td>
<td>0.05</td>
<td>0.12*</td>
<td>0.06</td>
<td>0.13*</td>
<td>0.05</td>
<td>0.13*</td>
<td>0.05</td>
<td>0.13*</td>
<td>0.05</td>
<td>0.13*</td>
<td>0.05</td>
</tr>
</tbody>
</table>

Notes: * p < .05. Sentence presentation coded: 0 = Whole sentence; 1 = Rapid Visual Serial Presentation. Sentence ambiguity coded: 0 = Unambiguous; 1 = Ambiguous. N = 19 for model including Operation Span; N = 23 otherwise.
Appendix 1

Sample Items from the Shipley Vocabulary Task

Instructions: In the test below, the first word in each item is printed in capital letters. Next are four other words. Circle the one word that means the same thing, or most nearly the same thing, as the word in all capital letters.

Example:

<table>
<thead>
<tr>
<th>Example:</th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>LARGE</td>
<td>red</td>
<td><strong>big</strong></td>
<td>silent</td>
<td>wet</td>
</tr>
<tr>
<td>(1)</td>
<td>TALK</td>
<td>draw</td>
<td>eat</td>
<td><strong>speak</strong></td>
</tr>
<tr>
<td>(17)</td>
<td>NARRATE</td>
<td>yield</td>
<td>buy</td>
<td><strong>associate</strong></td>
</tr>
<tr>
<td>(21)</td>
<td>SQUANDER</td>
<td>tease</td>
<td>belittle</td>
<td>cut</td>
</tr>
<tr>
<td>(29)</td>
<td>AMULET</td>
<td><strong>charm</strong></td>
<td>orphan</td>
<td>dingo</td>
</tr>
</tbody>
</table>