Sentence Processing in Context: The Impact of Experience on Individual Differences

Thomas A. Farmer (taf22@cornell.edu) and Morten H. Christiansen (mhc27@cornell.edu)
Department of Psychology, Cornell University, Ithaca, NY 14853 USA

Karen A. Kemtes (kkemtes@unlv.nevada.edu)
Department of Psychology, University of Nevada-Las Vegas, 4505 Maryland Parkway MS 5030
Las Vegas, NV 89154 USA

Abstract

There exists considerable variation, at the level of the individual, in human sentence processing performance. Here, we aim to illuminate the degree to which experience with language can account for these individual differences. In Experiment 1, we demonstrate that subtle interactions between specific verbs and preceding linguistic context can drive reading times on complex sentences, but only in participants with a high amount of reading experience. Experiment 2 demonstrates, psychometrically, that traditional reading span tasks seem to measure language processing skill, heavily influenced by experience with language, instead of a verbal working memory capacity. In combination, these results support the idea that reading span measures and sentence processing tasks are tapping into the same underlying skill, and crucially, that this skill is determined, primarily, by experience.

Keywords: Ambiguity Resolution; Individual Differences; Language Experience; Span Tasks

Introduction

From what factors do individual differences in sentence processing arise? One proposal is that performance on language comprehension tasks varies as a function of verbal working memory capacity (Just & Carpenter, 1992). Proponents of the capacity argument often note that on complex sentences, such as those containing relative clauses, high span individuals elicit patterns of reading times distinctly different from those elicited by low span individuals. Indeed, these distinct processing patterns are attributed to differences in the capacity of high versus low span individuals to simultaneously store and process information.

Alternatively, MacDonald and Christiansen (2002), arguing for an experience-driven comprehension system, proposed that individual differences in language comprehension are, in part, a product of differences in language experience. Crucially, the authors proposed that reading span tasks, traditionally thought to measure verbal working memory capacity, actually measure differences in language experience; given the highly linguistic nature of these tasks, people with more language experience have better language-related skills, and as such, exhibit superior performance.

Although these two perspectives on individual differences in sentence processing overlap considerably in terms of the kinds of predictions they make for behavioral data, the experienced-based approach often predicts more subtle interactions between particular structural elements and specific lexical items (see below for an example). In this paper, we capitalize on such fine-grained predictions to explore the manner in which individual differences in reading experience influence on-line sentence processing performance. In Experiment 1, we demonstrate that performance on a traditional reading span task is predictive of the degree to which individuals, during the processing of complex sentences, are sensitive to subtle interactions between specific verbs and preceding linguistic context, as would be predicted by the experienced-based approach. In Experiment 2, we present direct psychometric evidence that reading span tasks do seem to be measuring language experience instead of a verbal working memory capacity.

Experiment 1

Although some theories of sentence processing maintain that syntactic information is the primary factor influencing an initial first-pass parse of a sentence (Frazier & Fodor, 1978), other researchers have found that non-syntactic information can also influence first-pass reading time patterns on complex sentences.

Altmann, Garnham, and Dennis (1992) investigated the manner in which discourse-context influenced processing of the Sentential Complement/Relative Clause ambiguity (1).

1 (a) **SC-Resolved:** He told the woman / that he'd misunderstood / the nature / of her / question.

(b) **RC-Resolved:** He told the woman / that he'd misunderstood / to repeat / her last / question.

(c) **Unambiguous Control:** He asked the woman / that he'd misunderstood / *to repeat* / her last / question.

The fragment *...that he'd misunderstood*… contains a syntactic ambiguity because *told* can be followed by either an NP + sentential complement (1a) or a relative clause (1b). In the first case, *that* becomes a complementizer, thus resulting in a sentential complement (SC) interpretation. In the second case, *that* becomes a pronoun leading to a relative clause (RC) interpretation. Disambiguation occurs in the segment of the sentence occurring after *misunderstood*.

Research has demonstrated that when participants read ambiguous sentences of this type, they experience an increase in reading times (RTs) at the point of disambiguation when the ambiguity is resolved in accordance with the more complex RC interpretation (Kemtes & Kemper, 1997; MacDonald, Just, & Carpenter, 1992). This increase in RTs is typically referred to as the garden-path effect.

Of interest, Altmann et al. found that the nature of the referents contained within the discourse-context (full context sentences can be seen in Table 1) could influence a reader's susceptibility to the garden-path effect. When the discourse-context contained two similar referents (*the two women*), the garden-path effect on RC-resolved sentences was attenuated. Additionally, when the discourse-context contained two distinct entities (*the man and the woman*), the SC interpretation was facilitated. The attenuation of the garden-path effect associated with the more complex RC-resolved sentences was attributed to the fact that encountering two very similar entities within a discourse sets up an expectation that the entities will be differentiated, and a relative clause is one primary way for that differentiation to occur.

Table 1: Complete contexts for example (1).

Sentential Complement-Supporting Context
A bank manager was giving financial advice to **a man** and **a woman**. They were asking about the benefits of a high-interest savings account. The bank manager had misunderstood **the woman's** question about the account but understood **the man** perfectly.

Relative Clause-Supporting Context
A bank manager was giving financial advice to **two women**. They were asking about the benefits of a high interest savings account. The bank manager had misunderstood **one of the women's** questions about the account but understood **the other** perfectly.

In Experiment 1, we administered the SC/RC ambiguous sentences from Altmann et al., along with a reading span task. Both the experience-based and the capacity-based perspectives predict an effect of reading span on the ability to utilize information contained within discourse-context during syntactic ambiguity processing; namely, high span individuals would be more garden-pathed than low span individuals in situations where a mismatch existed between the context and the ambiguity resolution (i.e., an SC-resolved sentence occurring in a context that contains two related entities, or vice versa). To further differentiate the two perspectives, we exploit an interesting aspect of the stimuli used by Altmann et al. All but one of the target sentences used the verb *told* to introduce the SC/RC structural ambiguity. Crucially, Spivey and Tanenhaus (1994) conducted a corpus analysis in which they found that when *told* creates an SC/RC ambiguity, it is always resolved with an SC continuation. This kind of distributional asymmetry would be predicted by the experience-based approach to interact with linguistic context. Specifically, high span subjects, due to their greater (distributional) experience, should show stronger biases toward SC continuations, overall, than low span subjects. There is no a priori reason to assume that differences in working memory capacity, in and of itself, would result in a similar prediction.

Method

Participants Fifty-three undergraduates (mean age = 18.79 years, *SD* = .93) from a medium-sized Mid-Atlantic university participated in this study

Materials The experimental sentences were adapted from Altman et al. (1992), and were used because of the noted distributional biases exhibited by the sentences. They were constructed from 36 sentence frames. Each experimental frame was altered in order to include an SC-resolved sentence, an RC-resolved sentence, and an unambiguous control sentence. Additionally, two different contexts, the SC-supporting context (2 distinct NPs, such as *The man and the woman* in (1)) and the RC-supporting context (2 related NPs, such as *The two women* in (1)), were created for each sentence frame. All sentence types within each frame were crossed with all possible contexts to form six possible combinations from each sentence frame.

The experimental sentences were counterbalanced across different presentation lists. Each list contained four instances of each possible condition, but only one version of each sentence frame. Additionally, eight unrelated practice items and 22 filler items were incorporated into each list.

Procedure All sentences were randomly presented in a non-cumulative, word-by-word moving window format (Just, Carpener, & Woolley, 1982) using Psyscope version 1.2.5 (Cohen, MacWhinney, Flatt, & Provost, 1993). Participants were randomly assigned to one of the presentation lists.

Participants initially viewed a tutorial designed to acquaint them with the task. Participants were then instructed to press the 'GO' key to begin the task. The entire test item appeared on the center (left-justified) of the screen 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 (msec) were recorded for each word. Following each sentence, participants responded to a Yes/No comprehension question, and upon another key press, the next item appeared.

After having completed the sentence comprehension task, a modified version of the Waters and Caplan (1996) composite Z-score task was used to measure working memory capacity. Participants first saw a sentence. After they read it, they first had to memorize the final word of the sentence. Then they had to make an acceptability judgment of the semantic properties of the sentence by pressing the "YES" key if the sentence they had just read made sense or the "NO" key if it did not. Another sentence appeared after the semantic judgment was made, and participants were asked to repeat the process. An asterisk then appeared on the screen and participants were requested to recall the last word of each sentence in the set.

The number of words the participant had to maintain in memory while making semantic judgments was increased incrementally. Three sets of each level appeared in such a way that participants had three attempts at the two-word level, three attempts at the three-word level, and so on until the final six-word level. Participants were instructed to keep going all the way until the end of the task, even if they were not able to remember some of the words.

Results and Discussion

The score on the modified version of the Waters and Caplan (1996) span task was the number of levels for which participants were able to recall all of the words from at least two of the three sets for each level. Participants were also given a half of a point if they got one of the sets correct from the level appearing after the highest level fully completed. This scoring procedure deviates from the method advocated by Waters and Caplan (1996). Instead of creating a composite score based on several different aspects of the task, as advocated by the authors, we simply scored performance in accordance with the method used to score the more traditional Daneman and Carpenter (1992) reading span task. This was done in order to ensure comparability with the results of other studies investigating the relationship between reading span and language comprehension. The Daneman and Carpenter span task was not used here because the Waters and Caplan task, even without the composite scoring method, has been shown to be more reliable (Waters and Caplan, 1996).

RTs on each word were length-adjusted according to a procedure described by Ferriera and Clifton (1986). First, using the raw RTs on all words in both the experimental and filler items, we computed a regression equation predicting each participant's overall RT per word from the number of characters in each word. The equation was used in order to generate an expected RT on each word *given* its length. Expected RTs on each word were then subtracted from the observed RTs, and the resulting difference score was used for all analyses.

Experimental target sentences were divided into five different regions (see segment delimitation, indicated by a " / " in (1a-c)). The second segment constituted the point of ambiguity, segment three was the point of disambiguation, and segment four consisted of the remaining words up to, but not including, the sentence-final word. Segment four will be referred to as the carry-over segment because difficulty in ambiguity processing may not end in segment three; the effect of the ambiguity may be so strong that it exerts downstream effects.

A 2 (SC vs. RC-supporting context) X 3 (SC-resolution vs. RC-resolution vs. unambiguous) X 3 (ambiguity vs. disambiguiation vs. carry-over) repeated-measures ANOVA yielded a statistically reliable three-way interaction, $F_1(4, 208)=5.97$, $p<.0005$, $F_2(4, 120)=5.9$, $p<.0005$. As evident in Figure 1(a and b), the garden-path effect on the RC-resolved sentences appearing in the RC-supporting context was reduced, although not completely attenuated. Interestingly, when the ambiguity was resolved in accordance with the SC interpretation, and the context supported the SC resolution, the SC-resolved sentences were read significantly more quickly than the unambiguous control sentences at disambiguation, $t(52)=2.33$, $p=.024$.

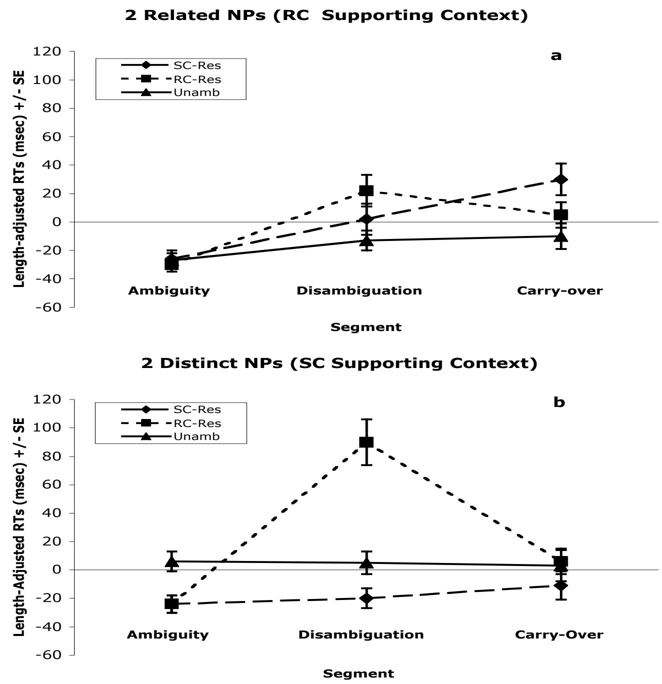

Figure 1: The significant three-way interaction, before accounting for span. Discourse-context does seem to be influencing RTs on ambiguous sentences given the fluctuations in RTs per target sentence across both contexts.

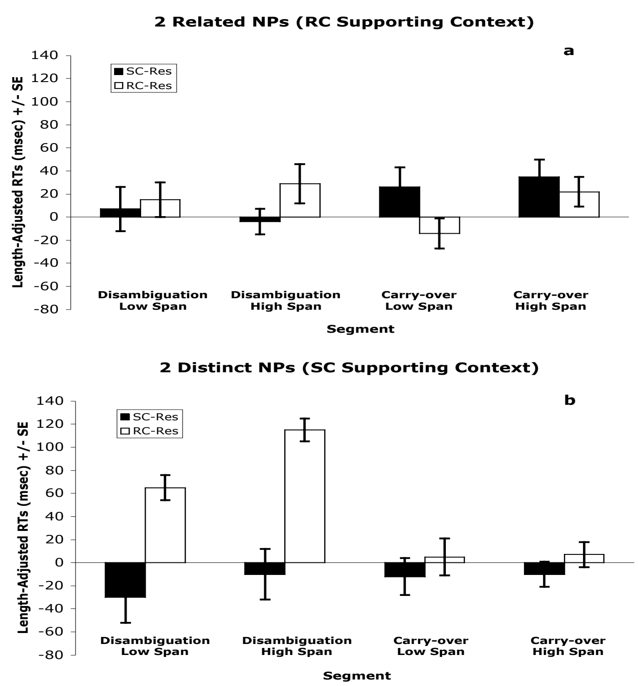

Figure 2: Span differences across segment and context.

Additionally, when the context supported the SC interpretation, but the ambiguity was resolved with the RC interpretation, participants were severely garden-pathed, $t(52)=5.35$, $p<.0005$. Participants also garden-pathed at the carry-over segment when the context supported the RC interpretation but the ambiguity was resolved with the SC interpretation, $t(52)=3.158$, $p=.003$.

More important to the goal of this study, however, are the results after accounting for span. Regression analyses were conducted with the continuous span scores. However, for illustrative purposes, participants were placed in the high span or the low span groups based on a median-split of span scores. Span scores significantly predicted the difference between the RC-resolved and unambiguous sentences at disambiguation in the SC-supporting context, $t(52)=2.04$, $p=.047$. As predicted, high span individuals exhibited a preference for the SC-resolved sentences across both conditions whereas low span individuals did not. Figure 2b illustrates that high span individuals are more sensitive to the mismatch created by the RC-resolved sentences in the SC-supporting context at disambiguation. Interestingly, in the RC-supporting context, high span individuals were slower to read RC-resolved sentences at disambiguation. Although not a significant difference at disambiguation, the effect of the ambiguity appears to have carried over into segment four, where high span participants have significantly longer RTs than do low span participants, $t(51)=2.02$, $p=.049$.

Given the high frequency with which ambiguities arising from *told* are continued with an SC in naturally occurring language, high span individuals appear to be biased toward them when presented with SC/RC ambiguous sentences. Moreover, this bias seems to be robust across both context conditions, and is especially pronounced in the context that favors SC resolution. Alternatively stated, the context manipulation seems to work for low span, but not high span, individuals. It is difficult for capacity-based theories to account for this result given that there exists no substantial reason why high span, but not low span, individuals would possess such a bias. Indeed, these results can be seen as a product of a more refined comprehension system that is more experientially attuned to naturally-occurring language patterns.

Experiment 2

The interpretation of the results obtained in Experiment 1 rests, in part, on the notion that the modified Waters and Caplan span task measures language comprehension skill. One might assume that participants who do well on reading span tasks are also participants who have had more experience with language than those who do poorly on them. In other words, we argue that reading span tasks measure individual differences in reading skill, but that these individual differences arise, to a large extent, via individual differences in reading experience.

In order to test this assumption, we administered five tasks we believed to measure either verbal working memory (vWM) capacity or language experience. As measures of language experience (or correlates thereof), we administered the *Author Recognition Task* (ART) (West, Stanovich, & Mitchell, 1993), a *Vocabulary Task* (VOCAB) (Shipley, 1940), and a *Need for Cognition* (COGNEED) scale (Cacioppo, Petty, & Kao, 1984). As a traditional measure of working memory, we administered the *Backward Digit Span* (BDS) task (Wechsler, 1981). Notably, we also administered the *Waters and Caplan (1996) span task* (vWM). Through exploratory factor analysis, we present some direct psychometric evidence indicating that reading span tasks measure individual differences in language comprehension skill, and that scores on an Experience factor (EX-Factor) are significant predictors of individual differences on garden-path relative clause sentences.

Method

Participants Seventy-two native English speakers (M=18.89 years, SD=.994) enrolled at a medium-sized Mid-Atlantic university participated in this study for extra course credit. One participant's data was excluded due to errors in data recording.

Materials The Author Recognition Test (ART) (West et al., 1993) was used as a measure of print exposure, and involved the presentation of 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; final scores on the task were penalized based on the number of foils checked. Participants were instructed to read the list and place a checkmark next to the names they believed to be real authors. One additional aspect of the Author Recognition Test was that two "effort probe" items were included. These effort probe items (Edgar Allen Poe and Stephen King) were items that, theoretically, every college student should be able to recognize.

The Shipley (1940) vocabulary task was given to participants as a measure of reading ability/reading experience. Participants were presented with a target word, and were required to choose the word most similar to it from a list containing four choices. The task contained 40 target words.

Need for cognition (COGNEED), a personality variable that is typically defined as the need to be cognitively engaged, was included under the assumption that people high in COGNEED would also be people who read more frequently. COGNEED was measured using a revised version of the Need for Cognition (NCS) scale (Cacioppo, Petty, & Kao, 1984), which contained 18 items that have been shown to be good predictors of Need for Cognition. Participants rated themselves on each item (e.g. *I would prefer complex to simple problems.*) on a nine-point Likert-type scale (-4=extremely inaccurate, 4=extremely accurate).

The WAIS-R (Wechsler, 1981) backward digit span task is similar to the vWM task, but without a strong grounding in language comprehension processes. It was composed of 14 sets of digits, with two sets at each level of complexity. Participants saw two digits presented rapidly. After all the digits in each set were presented, an asterisk appeared, and participants were instructed to recall the digits, not in the order in which they were presented, but in a backwards order. The number of digits that had to be recalled increased at each set-level, starting with two and ending with eight.

Verbal working memory (vWM) span was measured by the modified version of the Waters and Caplan (1996) span task (as described in Experiment 1).

On-line reading performance was assessed using the Main Verb/Reduced Relative Clause (MV/RRC) ambiguity materials from MacDonald, Just, and Carpenter (1992).

3 (a) 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 warned about the dangers / **conducted the midnight** / raid.

In example (3), the ambiguous sentences (3a and 3c) become ambiguous at segment two. The verb "warned" may be interpreted either as the main verb (MV) of the sentence (3a) or as the beginning of a relative clause (RC) (3c). Segment three, the point of disambiguation, contains the information necessary to arrive at the correct interpretation of the ambiguity. The MV unambiguous sentence (3b) is not ambiguous because the verb "spoke" does not produce an ambiguity. The unambiguous RC sentence is unambiguous because the inclusion of the relative pronoun (plus the past tense form of the verb "to be") eradicates any ambiguity. These sentence materials have been consistently shown to elicit a garden-path effect when the ambiguity is resolved in accordance with the RC interpretation (3c) (Kemtes & Kemper, 1997; MacDonald et al., 1992).

Thirty-six sentence frames similar to the one above were counterbalanced across four lists. Each list was comprised of one sentence from each of the 36 sentence frames. As a result, each participant saw nine of each sentence-type, but only one sentence from each sentence frame. Fifty filler items, along with eight unrelated practice items, were incorporated into each list.

Procedure Participants completed the vocabulary task first, followed by the Need for Cognition scale. Then, the on-line language comprehension task was administered as described in Experiment 1. Participants subsequently completed the modified version of the Waters and Caplan span task, followed by the backwards digit span task. The Author Recognition Task was administered last.

Results and Discussion

The score on the Author Recognition Test was simply the proportion of real authors that were checked by each participant minus the proportion of foil names checked. This resulted in a mean score of .31 (interpreted as 31 percent) with a standard deviation of .11. All participants checked at least one of the effort probes.

The modified version of the Waters and Caplan span task was scored the same way as detailed in Experiment 1, eliciting a mean response of 4.43 (*SD*=1.09). Possible scores on the BDS task ranged from 0 to 14 and were taken to be the number of consecutive trials for which participants correctly recalled all digits in the correct order (*M*=9.47, *SD*=2.48). The score on the vocabulary task was simply the number of items for which the participant answered with the correct synonym (*M*=31.32, *SD*=3.14). The 18-item need for cognition scale was scored by summing each participant's responses. Given that participants responded to each item on a nine-point Likert-type scale ranging from –4 (low COGNEED) to +4 (high COGNEED), possible scores ranged from –72 to +72. The mean need for cognition score was 10.68 (*SD*=22.76).

Exploratory factor analysis was conducted on the scores derived from the five measures with a principal axis factoring extraction method and varimax rotation. Two factors were extracted accounting for 35.17% of the original variance (a number commensurate with many published EFA studies). Scores on the five tasks were considered in the interpretation of a factor only if the factor loading value was .3 or above. All rotated loading values for each factor can be seen in Table 1. It should be noted that the pattern of factor loadings was the same across all types of non-orthogonal rotation methods as well.

Table 2: Rotated factor loadings.

	vWM	ART	VOCAB	COGNEED	BDS
Factor 1 (EX-Factor)	**.413**	**.629**	**.745**	**.378**	.007
Factor 2	.168	-.155	.068	-.006	**.661**

Examining the values in Table 2 reveals that vWM, which many argue to be a measure of verbal working memory, actually appears to load on Factor 1, along with the three measures hypothesized to measure language experience (ART, VOCAB, COGNEED). More interestingly, scores on the BDS task, the task that measures working memory, but without being heavily grounded in language comprehension processes, do not load on Factor 1. Instead, the BDS task loaded by itself on Factor 2. All factor cross-loadings were quite low, indicating that two separate factors were identified.

Given that ART, VOCAB, vWM, and COGNEED all loaded together, Factor 1 was named and interpreted as the experience factor (EX-Factor). Unfortunately, given that only one task loaded on factor two, it was deemed unstable and was not named or interpreted. Scores on the EX-Factor tasks were converted to z-scores and those z-scores were summed, resulting in a score on the EX-Factor for each participant. Below, we demonstrate that scores on the EX-

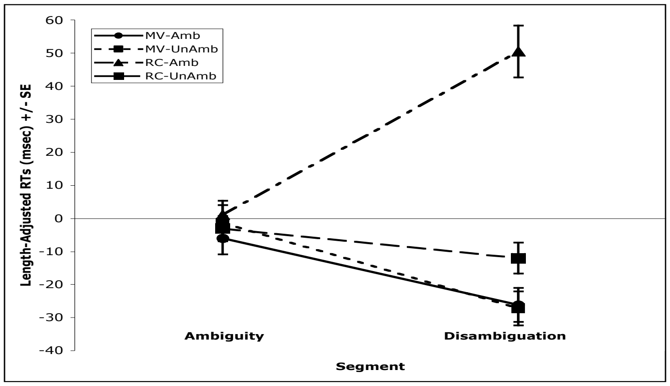

Figure 3: Length-adjusted RTs on the MV/RRC ambiguous sentences administered in Experiment 2.

Factor clearly predict RTs on RC garden-path sentences. Reading times on the MV/RRC sentences were length-adjusted as explained in Experiment 1. A 2 (MV vs. RRC) x 2 (ambiguous vs. unambiguous) x 2 (ambiguity vs. disambiguation) repeated-measures ANOVA revealed a significant three-way interaction, $F(1, 70)=18.60$, $p<.0005$.

The pattern of the interaction, illustrated in Figure 3, reveals that the garden-path effect did occur. There was a significant difference in RTs between the RC-resolved ambiguous sentences and the RC-unambiguous sentences at the point of disambiguation, $t(70)=6.72$, $p<.0005$.

Subsequently, EX-Factor scores were used to predict RTs at the point of disambiguation for each of the four sentence-types. EX-Factor scores significantly predicted RTs at disambiguation for the RC-resolved garden-path sentences, $t(70)=3.03$, $p=.003$, but not for the other three sentence types (all p's > .1). More impressively, EX-Factor scores predicted the difference in RTs between the RC-ambiguous and RC-unambiguous sentences at disambiguation, $t(70)=2.20$, $p=.031$. Additionally, although Factor 2 was considered unstable, it should be noted that memory (BDS scores) did not predict RTs at disambiguation for any of the four conditions (all p's > .1).

The factor analysis results suggest that the reading span task is grounded more in experience than memory. Furthermore, the predictive value of the EX-Factor scores in segments of the sentence where an individual difference effect would be expected offers some validity evidence in support of the EX-Factor—i.e., experience—in explaining individual differences in sentence processing.

General Discussion

Experiment 1 reveals that high span individuals seem to possess a bias toward the SC resolution of the SC/RC ambiguity. Given that there exists a strong bias in naturally-occurring language for an ambiguity created by *told* to be resolved with the SC interpretation, individuals with more language experience are also more likely to exhibit a preference for that resolution. As a result of the noted difficulty in explaining these results under a capacity-based view of individual differences in sentence processing, we argue that these results support an experience-based approach. Experiment 2, presenting additional support for an experience-based approach, demonstrates that span tasks are measuring processing skill instead of memory capacity.

In combination, these results support the idea that reading span measures and sentence processing tasks are tapping into the same underlying skill, and crucially, that this skill is determined, primarily, by experience. Investigations into the role that reading experience exerts on language processing are currently lacking. In light of these results, we argue that current conceptualizations of individual differences in sentence processing should be re-evaluated with a focus on the effects of experience with language.

References

Altmann, G. T. M., Garnham, A., & Dennis, Y. (1992). Avoiding the garden path: Eye movements in context. Journal of Memory and Language, 31, 685-712.

Cacioppo, J., Petty, R., & Kao, C. (1984). The efficient assessment of need for cognition. *Journal of Personality Assessment, 48*, 306-307

Cohen, J. D., MacWhinney, B., Flatt, M., & Provost, J. (1992). Psyscope: A new graphic interactive environment for designing psychology experiments. *Behavioral Research Methods, Instruments, and Computers, 25*(2), 257-271.

Daneman, M., & Carpenter, P. A. (1980). Individual differences in working memory and reading. *Journal of Verbal Learning and Verbal Behavior, 19*, 450-466.

Ferreira, F., & Clifton, C. (1986). The independence of syntactic processing. *Journal of Memory and Language, 25*, 348-368.

Frazier, L., & Fodor, J. D. (1978). The sausage machine: A new two-stage parsing model. *Cognition, 6*, 291-325.

Just, M. A., Carpenter, P. A., & Woolley, J. D. (1982). Paradigms and processes in reading comprehension. *Journal of Experimental Psychology: General, 111*, 228-238.

Just, M. A., & Carpenter, P. A. (1992). A capacity-based theory of comprehension: New frontiers of evidence and arguments. *Psychological Review, 103*, 773-780.

Kemtes, K. A., & Kemper, S. (1997). Younger and older adults' on-line processing of syntactically ambiguous sentences. *Psychology and Aging, 12*(2), 362-371.

MacDonald, M. C., Just, M., A., & Carpenter, P. A. (1992). Working memory constraints on the processing of syntactic ambiguity. *Cognition, 24*, 56-98.

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.

Shipley, W. C. (1940). A self-administered scale for measuring intellectual impairment and deterioration. *Journal of Psychology, 9*, 371-377.

Spivey, M. J., & Tanenhaus, M. (1994). Referential context and syntactic ambiguity resolution. In C. Clifton, L. Frazier, & K. Rayner (Eds.), *Perspectives on Sentence Processing*. Hillsdale, NJ: Erlbaum.

Waters, G. S., & Caplan, D. (1996). The measurement of verbal working memory capacity and its relation to reading comprehension. *Quarterly Journal of Experimental Psychology, 49*, 51-79.

Wechsler, D. (1981). The Wechsler Adult Intelligence Scale – Revised. New York: Psychological Corporation.

West, R. F., Stanovich, K. E., & Mitchell, H. R. (1993). Reading in the real world and its correlates. Reading Research Quarterly, 28(1), 35-50.