Neural Correlates of Semantic Plausibility in Sentence Comprehension in High and Low Working Memory Groups

Louise Stanczak
Gloria Waters
David Caplan
Neal J. Pearlmutter

Neal J. Pearlmutter
Northeastern University
Psychology Department, 125 NI
Boston, MA 02115
pearlmutter@neu.edu
617- 373-3040 (T)
617- 373-8714 (F)
Abstract

The neural correlates of semantic plausibility during sentence comprehension were investigated in high and low working memory groups using event-related fMRI. Sentences rated as plausible, implausible, or anomalous were presented followed by a comprehension question. Comprehension was more difficult for sentences with lower plausibility in behavioral measures, and increased BOLD signal was found as plausibility decreased in bilateral frontal and temporoparietal regions. These areas may reflect computation of plausibility and/or semantic selection for the most appropriate meaning. When activation was compared between high and low working memory groups, no differences were found as function of plausibility during sentence comprehension or when answering questions about sentences.
The semantic plausibility of a sentence affects comprehension. Participants in sentence comprehension experiments spend more time reading sentences with meanings rated as less plausible, particularly at words that are implausible based on expectancies from the prior context. Examples 1-3 demonstrate sentences of decreasing plausibility. Eye-tracking measures found that “carrots” was read with the greatest ease in plausible sentences and the greatest difficulty in anomalous sentences (Rayner, Warren, Juhasz, & Liversedge, 2004).

(1) John used a knife to chop the large carrots for dinner. (plausible)
(2) John used an axe to chop the large carrots for dinner. (implausible)
(3) John used a pump to inflate the large carrots for dinner. (anomalous)

Neuroimaging studies have reported greater BOLD signal when processing implausible sentences as compared to plausible sentences (Cardillo et al., 2004; Friederici et al., 2003; Kiehl, Laurens, & Liddle, 2002; Kuperberg et al., 2000, 2003; Ni et al., 2000). However, there is little consistency in the areas of activation reported, making it difficult to determine the brain areas that support plausibility during sentence comprehension. Studies have reported left inferior frontal activation (Cardillo et al., 2004), left fronto-temporal activation (Kuperberg et al., 2003), bilateral fronto-temporal activation (Kiehl, Laurens, & Liddle, 2002; Ni et al., 2000), or bilateral temporal and right premotor activation (Friederici et al., 2003). Differences in stimulus materials (e.g., number of words, manipulating a noun versus a verb) or methodology (e.g., presentation modality, task, statistical analyses) may account for the range of brain areas reported. Cardillo et al. (2004) and Kiehl, Laurens, and Liddle (2002) presented whole, written sentences manipulating the terminal noun for plausibility, but they did not control for
semantic content across the conditions compared. Kuperberg et al. (2003) also presented sentences visually but used rapid serial visual presentation (RSVP) and critically manipulated a terminal verb. In the aural modality, Ni et al. (2000) presented shorter three-word sentences, and Friederici et al. (2003) presented sentences in German. The experimental tasks, which may have influenced the attentional focus of the participants, were plausibility judgment (Friederici et al., 2003; Kiel, Laurens, & Liddle 2002; Kuperberg et al., 2003), lexical decision (Cardillo et al., 2004), and animacy detection (Ni et al. 2000). None of these studies investigated plausibility using a comprehension task, which is arguably a better paradigm for measuring the brain regions critical in processing plausibility during sentence comprehension.

The areas of activation reported in the studies above may reflect an increase in neural activity for the more difficult, implausible sentences and/or a decrease in neural activity for the easier, plausible sentences. Cardillo et al. (2004) is the only study that has investigated the nature of the BOLD signal differences between implausible and plausible sentences by comparing both sentence types to neutral sentences (see examples 4-6).

(4) The boy bounced the ball. (plausible)

(5) The next item is a chair. (neutral)

(6) Pasta is my favorite kind of wall. (implausible)

A region of interest (ROI) analysis was performed on the left inferior prefrontal cortex, the only area of activation significant in this study when comparing implausible to plausible sentences in an omnibus analysis. An increase in the BOLD response was found for implausible sentences relative to neutral sentences, while no difference was found between plausible and neutral sentences. These results suggest that the activation in the
left inferior prefrontal cortex was driven by the additional cognitive demands required to process implausible sentences.

Plausibility also provides information that affects syntactic parsing during sentence comprehension. Plausibility cues are used to resolve syntactically ambiguous sentences (Ferreira & Clifton, 1986; Garnesy, Pearlmutter, Myers, & Lotocky, 1997; Holmes, Stowe, & Cupples, 1989; MacDonald, Pearlmutter, & Seidenberg, 1994; Pickering & Traxler, 1998; Trueswell, Tanenhaus & Garnsey, 1994) or aid in the assignment of thematic roles for unambiguous, syntactically complex sentences (Traxler, Morris, & Seely, 2002).

Additionally, the use of plausibility information to resolve syntactically ambiguous sentences or syntactically complex, unambiguous sentences has been found to interact with working memory capacity. High span participants were sensitive to the plausibility of alternative interpretations during syntactic ambiguity resolution, while low span participants were not affected by these cues (Pearlmutter & MacDonald, 1995). Importantly, these effects were not driven by individual differences in linguistic knowledge related to plausibility constraints, as no differences were found between high and low span groups using an offline plausibility-rating task, but were only found when plausibility cues were used to disambiguate during online sentence comprehension. Using syntactically unambiguous sentences, King and Just (1991) found that low span participants benefited more than high span participants from plausibility information to aid in thematic role assignment when answering questions regarding syntactically complex clauses compared to simple clauses. Additionally, Traxler, Williams, Blozis, & Morris (2005) found a three-way interaction between working memory capacity,
syntactic complexity, and the use of plausibility to constrain thematic role assignment (but see Caplan & Waters, 2002 for an alternative view).

One limitation from the psychological literature is that no studies have investigated whether the differences in the use of plausibility information during sentence comprehension between high and low span participants are independent of sentence complexity, i.e., whether high and low span participants differ during the online processing of sentences as a function of plausibility, irrespective of syntactic complexity or ambiguity resolution. A limitation of the neuroimaging literature is that no studies have investigated possible differences in neurovascular responses to the use of plausibility information during sentence comprehension between high and low span participants. The present study addresses the questions raised by investigating the neural correlates of high and low working memory groups as a function of plausibility ratings in syntactically simple sentences, using a sentence comprehension task.

Methods

Participants

Sixteen right-handed, native English-speaking, Boston University students between age 19-29 were recruited (13 females, 3 males). Participants were paid and informed written consent was obtained. Participants were healthy volunteers with no history of neurological or psychiatric illnesses. Based on a one-hour behavioral session consisting of working memory and speed of sentence processing measures (see below), eight high and eight low working memory span participants were chosen, controlling for speed of sentence processing (see Waters et al., 2003), from a pool of 120 participants.
**Working Memory Evaluation.** Three working memory tests were administered: Alphabet Span, Subtract 2 Span, and a modified version of the Daneman and Carpenter (1980) Reading Span task called Sentence Span (see Waters & Caplan, 1996). In Alphabet Span, participants are required to repeat a series of monosyllabic, unrelated words presented aurally after rearranging them in alphabetical order. For Subtract 2 Span, participants are required to repeat a random sequence of numbers presented aurally, ranging from 2-9, after subtracting 2 from each number. In Sentence Span, acceptability judgments to sentences presented visually in the cleft subject form (e.g., “It was the child that bit into the fruit.”) are made while recalling the final word of each sentence. For each task, span was defined as the highest span size that the words/numbers were recalled accurately in three out of five trials. An additional .5 was added to the span score if two out of five trials were correct in the following span size. A composite working memory score was calculated by adding the three working memory scores, as Waters and Caplan (2004) found that the composite had better test-retest reliability than any one span measure alone. Using the composite score, low span participants were in the lowest span quartile (13.5 or lower) while high span participants were in the highest quartile (18.5 or higher).

**Speed of Sentence Processing:** As individual differences in speed of sentence processing efficiency have produced differential neural activation patterns (Waters et al., 2003), high and low span groups were controlled for speed of sentence processing. Acceptability judgments were made on 100 cleft-object (CO) and 100 object-subject (OS) sentences presented visually. Using the RTs and accuracy from CO, OS, and cleft-
subject (from Sentence Span) sentences, high and low span participants were within one standard deviation of the mean.

**Stimulus Materials**

All sentences had 10 words and consisted of two clauses. 240 sentences were constructed as 80 triplets (see examples 7-9). The direct object following the first verb (fourth word) was plausible, implausible, or anomalous. Sentences within the same triplet were followed by the same yes/no comprehension question, which did not query plausibility. Participants saw 80 sentences from each plausibility condition, i.e., one sentence from each of the 80 triplets. Three separate lists were counterbalanced across participants, with each list further divided into 8 sublists for presentation during 8 functional MRI runs. Participants saw 30 sentences during each run, 10 from each plausibility condition.

(7) Vanessa threw the **javelin** but did not win the competition. (plausible)

(8) Vanessa threw the **feather** but did not win the competition. (implausible)

(9) Vanessa threw the **situation** but did not win the competition. (anomalous)

*Comprehension Question:* Was the throw good enough to win?

Plausibility was determined by norms using the first clause in each sentence, rated on a plausibility scale from 1-7 (7=most plausible) by 144 college-aged participants. The mean plausibility ratings were as follows: plausible = 6.4, implausible = 3.9, anomalous = 1.7. Paired t-tests found significant differences between all three plausibility conditions (plausible vs. implausible: t=2063, p < .001, plausible vs. anomalous: t=9969, p < .001, implausible vs. anomalous: t=1527, p < .001). There was also a significant effect of word length for the critical noun between each plausibility condition as sentences with greater...
plausibility had significantly shorter critical nouns (plausible = 5.78 ch/word, implausible = 6.06 ch/word, anomalous = 6.67 ch/word). Finally, there was a smaller effect of word frequency such that the plausible sentences had critical words that were significantly more frequent than the implausible or anomalous sentences (log frequency of plausible = 1.37, implausible = 1.23, anomalous = 1.26).

**Procedure**

Sentences were presented using RSVP of a single word in the middle of the screen with one trial totaling 10 s. Each trial began with a 500 ms fixation, and then a 500 ms blank screen. Each word was presented for 300 ms followed by a 150 ms blank screen. A 1000 ms blank screen followed the terminal word, and then the comprehension question appeared for 3650 ms.

Participants held an MRI compatible button box in their left hand to record response RTs and accuracy. Participants were instructed to read each sentence carefully and answer the comprehension question as quickly and accurately as possible, using their index finger to indicate “yes” or their middle finger to indicate “no”. While in the scanner, participants were familiarized with the task in a short practice session.

An event-related design presented sentences in a pseudo-randomized order with pseudo-randomized inter-trial intervals ranging from 2-12 s, for optimal efficiency in deconvolution and estimation of the hemodynamic response (Burock et al., 1998; Dale, 1999; Dale & Buckner, 1997). Sentences were presented from a Dell latitude D600 computer using the software package StimPres 1.1 to the back of the scanner with a Sharp LCD projector and were viewed by the participant as a reflection in a mirror attached to the head coil.
Image Acquisition

Gradient-echo planar MR images were acquired on a 3T Allegra at Massachusetts General Hospital. A head coil was used for radiofrequency transmission and reception, with foam pads inserted inside to minimize head movement. Two T1-weighted images were collected with 128 sagittal slices in each volume (TR=7.25, TE=3, flip angle=7°, slice thickness=1.33 mm, matrix size=192 x 256; FOV=256 mm, in-plane resolution=1.33x1 mm). Eight T2*-weighted images were collected for each of the 8 sublists (TR=2; TE=30 ms, slice thickness=3 mm, interslice gap=0.9 mm, FOV=200 mm, in-plane resolution=3.1 mm), consisting of 30 slices with the BOLD contrast along the AC-PC plane. Each functional scan consisted of 220 volume acquisitions for a total of 6600 images.

Image Analysis

The two T1-weighted images were averaged for a model of each participant’s cortical surface, through an automated procedure involving segmentation of the cortical white matter, tessellation of the gray/white border, and inflation of the folded surface tessellation patterns (Dale, Fischl, & Sereno, 1999; Dale & Sereno, 1993). Automatic correction of topological defects in the resulting manifold was performed (Fischl, Liu, & Dale, 2001). Each participant’s reconstructed brain was spherically normalized then morphed onto an averaged spherical surface representation to optimally aligned the main sulcal/gyral features and match morphologically homologous cortical locations across participants, while minimizing metric distortion (Fischl, Sereno, & Dale, 1999; Fishl, Sereno, Tootel, & Dale, 1999). This formed a spherical-based coordinate system onto
which the selective averages and variances of each participant’s functional data were resampled and spherically smoothed.

Functional images were corrected for motion using the AFNI algorithm (Cox, 1996) and time-points with excessive movement were manually deleted, affecting less than 2% of the data. Two functional runs from one participant were also excluded due to excessive movement at the end of the session. Images were spatially smoothed using a Gaussian filter full width half maximum (FWHM) of 6 mm. Plausible, implausible, and anomalous trials were selectively averaged across the eight runs for each participant. A finite impulse response (FIR) model estimated the hemodynamic response average at each TR to yield 14 mean and variance values from –6 s pre-stimulus onset to 20 s post-stimulus onset. The data was spherically smoothed onto the surface tessellation with a FWHM of 8 mm.

Random effects analyses using a t-statistic across voxel-wise maps were conducted on the high span group, low span group, and whole group, using a significance level of $p < 10^{-3}$ (uncorrected for multiple comparisons). This was completed for each of the three comparisons: implausible - plausible, anomalous - plausible, anomalous - implausible. Two time-intervals were averaged to reflect the critical word and the comprehension question. Since the critical word appeared about 3 s into the stimulus presentation for 300 ms, and BOLD signal changes are expected approximately 4-6 s post-stimulus onset (Turner, 1997), activation was averaged from 8-14 s into the stimulus presentation to reflect the critical word. The comprehension question appeared about 6.5 s into the stimulus presentation for 3650 ms, so activation was averaged from 14-20 s into the stimulus presentation to reflect the comprehension question. The activation from the
low span group was compared to the high span group for each of the three comparisons at both time-intervals to investigate differences between the working memory groups.

To correct for multiple comparisons, a clustering program (Doherty et al., 2004) was run on the high-low span and the whole group statistical parametric maps at both time-intervals for all three comparisons to identify clusters which individually exceeded a cluster-threshold of $p < .05$, a voxel-threshold of $p < .01$, and a cluster size of $200 \text{ mm}^2$. Monte Carlo simulations calculated the likelihood of extracting one or more clusters of a given size and threshold under the null hypothesis. This was accomplished by generating a volume of Gaussian distributed random numbers for each participant that was processed in the same manner as the collected experimental data.

**Results**

**Behavioral Data**

*Accuracy:* Table 1A shows the mean percent correct for both working memory groups in each plausibility condition. A 2 (group) x 3 (plausibility) ANOVA was performed with group as a between subjects factor and plausibility as a within subjects factor. There were no main effects of working memory span, $F(1, 14) = .02$, NS, or plausibility, $F(2, 28) = 1.17$, NS, and no interaction, $F(2, 28) = 2.67$, NS.

*Reaction Time:* Table 1B shows the mean RTs by group and plausibility for correct responses within 3 standard deviation of the mean. A 2 x 3 ANOVA showed a significant main effect of plausibility, $F(2, 28) = 10.76$, $p < .001$. A Tukey’s test revealed that RTs for anomalous sentences were significantly longer than for implausible sentences ($p = .01$) and plausible sentences ($p < .001$), but plausible and implausible
sentences did not significantly differ. No main effect of working memory group, $F(1,14) = 1.34$, NS, or interaction between working memory group and plausibility, $F(2, 28) = .96$, NS, was present.

**Imaging Data**

*Effects at the Critical Word:* No differences between high and low span groups were found between any of the three plausibility comparisons; therefore, no activation maps are shown. When collapsing across high and low span groups, implausible sentences had greater BOLD signal than plausible sentences in the left inferior frontal sulcus and middle frontal gyrus (see ROI 1 in the top of Table 2 and Figure 1). In this comparison, plausible sentences showed greater BOLD response than implausible sentences in the right postcentral sulcus (see ROI 2 in Table 2 and Figure 1). Anomalous sentences demonstrated greater BOLD signal than plausible sentences in the left inferior frontal sulcus, precentral sulcus, and intraparietal sulcus (see ROIs 3-4 in the middle of Table 2 and Figure 1) and the right middle frontal gyrus, inferior frontal sulcus, pars triangularis, and angular gyrus (see ROIs 5-6 in the middle of Table 2 and Figure 1). Anomalous sentences had greater BOLD response than implausible sentences in the left precentral sulcus and right postcentral gyrus (see ROIs 7-8 in the bottom of Table 2 and Figure 1). No areas of activation were greater for plausible or implausible sentences than anomalous sentences.

Table 2 about here

Figure 1 about here

*Effects at the Comprehension Question:* No differences between high and low span groups were found when comparing the three plausibility conditions; therefore, no
activation maps are shown. When collapsing across the groups, questions following implausible sentences showed greater BOLD signal than questions following plausible sentences in the left frontal pole and right middle frontal gyrus (see ROIs 1-2 in the top of Table 3 and Figure 2). Questions following anomalous sentences demonstrated greater BOLD signal than questions following plausible sentences in the left superior frontal gyrus, inferior frontal, precentral sulcus, angular gyrus, intraparietal sulcus, and posterior cingulate gyrus (see ROIs 3-6 in Table 3 and Figure 2) and in the right middle frontal gyrus, angular gyrus/sulcus, and anterior cingulate sulcus (see ROIs 7-9 in Table 3 and Figure 2). No differences in BOLD signal were found between the questions following anomalous sentences and implausible sentences.

Table 3 about here

Figure 2 about here

*ROI Analysis for Critical Words:* The analyses above indicate that the time-interval averaged for the critical word was more sensitive to plausibility effects than the time-interval averaged for the comprehension question. To investigate the nature of plausibility effects, the average hemodynamic response at each TR was calculated for each of the 8 clusters identified at the critical word (see Figure 1 for hemodynamic curves). The average hemodynamic response during the time-interval previously described as reflecting the critical word (8-14 s into the stimulus presentation) for each ROI was entered into a 2 (group) x 3 (plausibility) ANOVA. Working memory effects were further investigated using this statistically more powerful analysis.

All 8 ROIs had a significant main effect of plausibility, consistent with the cluster analysis. The main effect of span was non-significant in all ROIs, except one in the right
inferior, middle frontal region, $F(1, 14) = 5.76, p=.03$, (see ROI 5 in Table 2 and Figure 1). For this ROI, high span participants showed a negative percent signal change while low span participants were very close to the pre-stimulus baseline. No interactions between plausibility and working memory span were found.

Because no significant interaction between working memory and plausibility was found, the average hemodynamic response curves collapsed across working memory groups are plotted for each ROI in Figure 1. The left hemisphere demonstrated an increase in BOLD signal relative to the pre-stimulus baseline, which correlated with plausibility such that anomalous sentences produced the largest increase in activity. The right hemisphere showed a decrease in BOLD signal relative to the pre-stimulus baseline, with plausible sentences showing the largest decrease in activity. The one exception to this pattern was the cluster in the right post-central sulcus (see ROI 2 in Table 2 and Figure 1), which is also the only ROI that demonstrated greater activation for plausible sentences than implausible sentences.

Discussion

**Plausibility Activations**

Effects of plausibility were found in both behavioral and neurological data. In the neurological data, activation in a bilateral frontal and temporoparietal network increased as sentences were less plausible. Increased activation in the left frontal cortex as plausibility decreased was found in all comparisons, a finding consistent with many other studies (Cardillo et al., 2004; Hagoort et al., 2004; Kiehl, Laurens, & Liddle 2002; Kuperberg et al., 2003; Ni et al., 2000; for an exception see Friederici et al., 2003). The right hemisphere homolog of the left frontal cortex and bilateral temporal-parietal regions
also increased in BOLD response as plausibility decreased. Differences in activation between plausibility comparisons in the current study and other studies may reflect differences in the degree of implausibility and/or type of violation (e.g., pragmatic versus selection restriction violation).

Activation maps for the critical word showed differences between all three levels of plausibility (see Table 2 and Figure 1). For the comprehension question, which did not probe plausibility, differences were also found between levels of plausibility, but not across the implausible and anomalous sentences comparison (see Table 3 and Figure 2). A similar activation pattern at both the critical word and the comprehension question was found, with additional activation during the comprehension question bilaterally in the cingulate for anomalous sentences. The similarity in the location of activation at both time-intervals and the uni-modal distribution depicted in the hemodynamic response curves of these regions suggest that activation at the comprehension question is an extension of the activation at the critical word.

Two cognitive processes are proposed to play a role in processing plausibility during sentence comprehension. The first process is the computation of semantic plausibility, in which the meaning of a sentence is compared to the likelihood of an event occurring in the real-world based on previous experience and information stored in semantic memory (Hagoort et al., 2004; Neely, 1991). Computations for highly plausible scenarios may be easier than computations for meanings with lower plausibility, as an extended search in semantic memory may be necessary when the meaning of a sentence is implausible or anomalous. The second process is the consideration of new meanings that are unavailable to semantic memory, i.e., new semantic alternatives not previously
learned. This process may be engaged more frequently and more extensively as plausibility decreases. Finally, these two processes must interact in order for the plausibility of various alternative meanings to be compared so that the most appropriate interpretation can be selected.

The areas activated in the current study are consistent with previous research on the retrieval of semantic knowledge and the selection of semantic alternatives. The temporoparietal junction has been proposed to be a part of a distributed store of semantic knowledge (Price, 2000). The left prefrontal cortex is hypothesized to retrieve semantic alternatives for individual words (Wagner et al., 2001), integrate sentence meaning with real-world knowledge (Hagoort et al., 2004), and select an appropriate target amongst competing alternatives (Thompson-Schill et al., 1997; Thompson-Schill et al., 1999). Therefore, greater bilateral frontal-temporoparietal activation as a function of plausibility suggests that comprehending sentences with decreased plausibility places additional demands on the retrieval of alternative meanings, the calculation of plausibility, and/or the selection of the most appropriate meaning.

The right postcentral sulcus showed greater activation for plausible than for anomalous sentences (see ROI 1 in Table 2 and Figure 1). This finding has not been reported in other studies and requires further investigation.

**Hemispheric Differences in Hemodynamic Response Curves**

The ROI analysis showed that, aside from the right post-central sulcus, regions in both hemispheres demonstrated greater BOLD signal for less plausible sentences, but this pattern is positive relative to the pre-stimulus baseline in the left hemisphere and negative relative to the pre-stimulus baseline in the right hemisphere. This pattern is inconsistent
with the interpretation that the left hemisphere is activated while the right hemisphere is inhibited as a function of plausibility. If this was the case, the anomalous sentences, which showed the greatest activation in the left hemisphere, would be the most inhibited in the right hemisphere, but this mirrored pattern of activation is not observed (see Figure 1).

A more viable possibility is that activation in both hemispheres is driven by the same cognitive processes, e.g., those mentioned previously, but the pre-stimulus baseline activity in the two hemispheres is unequal. If the right hemisphere is engaged in another task prior to stimulus onset and is more active than the left hemisphere, then a similar change in BOLD signal within each hemisphere related to plausibility could present as a positive hemodynamic curve in the left hemisphere and a negative hemodynamic curve in the right hemisphere. This would indicate that the left hemisphere is more active with respect to the pre-stimulus baseline while processing sentences as a function of plausibility, while the right hemisphere disengages from the more demanding pre-stimulus baseline task to process sentences as a function of plausibility. Therefore, the “negative” response in the right hemisphere may reflect task-induced deactivation, which is proposed to result from the interruption of other processing by an experimental task (Binder et al., 1999; Chen et al., in press). This possibility, and the nature of the right hemisphere pre-stimulus baseline task, requires further investigation.

A second alternative for the observed pattern of hemodynamic response is that the underlying cognitive processes that drive the left and right hemispheres differ, producing laterality effects in the hemodynamic response. In support of this alternative, laterality differences in the selection of single-word meanings have been posited such that the right
hemisphere maintains multiple meanings regardless of context while the left hemisphere selects the appropriate meaning or suppresses the inappropriate meaning (Corney & Evans, 2000; Faust & Chiarello, 1998; Faust & Lavidor, 2003). Two arguments against this alternative are that homologous regions of activation would be less likely to emerge if the two hemispheres were engaged in entirely distinct cognitive processes, and it not clear why maintaining multiple meanings of less plausible sentences in memory would lead to a reduction in BOLD signal relative to a pre-stimulus baseline.

Working Memory and Plausibility in Sentence Comprehension

Plausibility did not interact with working memory capacity in either the behavioral or the neuroimaging measures. This finding can be interpreted with respect to previous studies. Using plausibility to resolve syntactic ambiguity (e.g., Pearlmutter & MacDonald, 1995) or aid in processing syntactically complex structures (e.g., King & Just, 1991; Traxler et al., 2005) may differ between high and low span participants. Evidence from Pearlmutter and MacDonald suggests that high span participants are better at integrating plausibility cues during syntactic ambiguity resolution, while evidence from King and Just suggests that low span participants are better at integrating plausibility cues to aid in comprehension of unambiguous, syntactically complex sentences. As these studies differ along additional lines (e.g., online versus offline measures of sentence comprehension), further research is necessary to understand the nature of these interactions. In contrast, the results from the present study and Pearlmutter and MacDonald found that plausibility effects in unambiguous, structurally simple sentences did not interact with working memory capacity. In conclusion, high and low span participants may not differ in the computation and/or use of plausibility information.
per se, but they may differ in the ability to integrate plausibility cues with other complex, syntactic processes during sentence comprehension.
References


Author Note

Louise Stanczak, Department of Speech, Language and Hearing Science, Boston University; Gloria Waters, Department of Speech, Language, and Hearing Science, Boston University, David Caplan, Neuropsychology Laboratory, Massachusetts General Hospital; Neal Pearlmutter, Department of Psychology, Northeastern University.

This project was supported by NIH R01 grant DC05237 to NP. The authors thank Michael Amato, Seth Amsden, Vered Argaman, Evan Chen, Joanna Currier, Grazyna Furman, Alexandra Geyer, Yuka Kominato, Daphne Robakis, Elizabeth Tatro, and Naa-Solo Tettey for help with stimulus construction, data collection, and/or data processing.

Correspondence concerning this article should be addressed to Neal J. Pearlmutter, Department of Psychology, Northeastern University, 125 NI, Boston, MA 02115. email: pearlmutter@nev.edu.
Table 1.

A. Mean Percent Correct (SD) for the Comprehension Task

<table>
<thead>
<tr>
<th></th>
<th>Plausible</th>
<th>Implausible</th>
<th>Anomalous</th>
</tr>
</thead>
<tbody>
<tr>
<td>High Span</td>
<td>90.1 (30)</td>
<td>90.3 (30)</td>
<td>91.1 (29)</td>
</tr>
<tr>
<td>Low Span</td>
<td>92.9 (26)</td>
<td>90.9 (29)</td>
<td>88.2 (32)</td>
</tr>
<tr>
<td>Average</td>
<td>91.5 (28)</td>
<td>90.6 (29)</td>
<td>89.7 (30)</td>
</tr>
</tbody>
</table>

B. Mean RT in ms (SD) for the Comprehension Task

<table>
<thead>
<tr>
<th></th>
<th>Plausible</th>
<th>Implausible</th>
<th>Anomalous</th>
</tr>
</thead>
<tbody>
<tr>
<td>High Span</td>
<td>1911 (516)</td>
<td>1958 (521)</td>
<td>2021 (526)</td>
</tr>
<tr>
<td>Low Span</td>
<td>2095 (532)</td>
<td>2104 (538)</td>
<td>2156 (543)</td>
</tr>
<tr>
<td>Average</td>
<td>2003 (528)</td>
<td>2031 (536)</td>
<td>2088 (531)</td>
</tr>
</tbody>
</table>

RTs included in Table 1B were from accurate responses within 3 standard deviation of the mean.
Table 2.

Significant ROIs at the Critical Word

<table>
<thead>
<tr>
<th>Contrast</th>
<th>ROI #</th>
<th>Region</th>
<th>Talaraich (x, y, z)</th>
<th>Size (mm²)</th>
<th>Cluster t-statistic</th>
</tr>
</thead>
<tbody>
<tr>
<td>Implausible – Plausible</td>
<td>1</td>
<td>L. Inferior Frontal Sulcus, Middle Frontal Gyrus</td>
<td>(-60, 20, 9)</td>
<td>784</td>
<td>4.9</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>R. Postcentral Sulcus</td>
<td>(22, -28, 66)</td>
<td>469</td>
<td>-3.0</td>
</tr>
<tr>
<td>Anomalous – Plausible</td>
<td>3</td>
<td>L. Inferior Frontal Sulcus, Precentral Sulcus</td>
<td>(-51, 10, 41)</td>
<td>1574</td>
<td>4.1</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>L. Intraparietal Sulcus</td>
<td>(-41, -50, 40)</td>
<td>598</td>
<td>3.0</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>R. Middle Frontal Gyrus, Inferior Frontal Sulcus, Pars Triangularis</td>
<td>(24, 35, 24)</td>
<td>564</td>
<td>3.3</td>
</tr>
<tr>
<td></td>
<td>6</td>
<td>R. Angular Gyrus</td>
<td>(32, -60, 49)</td>
<td>227</td>
<td>2.5</td>
</tr>
<tr>
<td>Anomalous – Implausible</td>
<td>7</td>
<td>L. Precentral Sulcus</td>
<td>(-33, 13, 29)</td>
<td>395</td>
<td>2.8</td>
</tr>
<tr>
<td></td>
<td>8</td>
<td>R. Postcentral Gyrus</td>
<td>(36, 29, 37)</td>
<td>381</td>
<td>3.4</td>
</tr>
</tbody>
</table>

The ROI #s above correspond to Figure 1 and represent significant clusters for the average activation in all participants at the critical word (8-14 s into the stimulus presentation). X, Y, and Z represent Talaraich and Tournoux (Talairach & Tournoux, 1988) coordinates.
Table 3.

Significant ROIs at the Comprehension Question

<table>
<thead>
<tr>
<th>Contrast</th>
<th>ROI #</th>
<th>Region</th>
<th>Talaraih (x, y, z)</th>
<th>Size (mm$^2$)</th>
<th>Cluster t-statistic</th>
</tr>
</thead>
<tbody>
<tr>
<td>Implausible – Plausible</td>
<td>1</td>
<td>L. Frontal Pole</td>
<td>(-31, 52, 0)</td>
<td>506</td>
<td>3.8</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>R. Middle Frontal Gyrus</td>
<td>(30, 24, 38)</td>
<td>305</td>
<td>2.6</td>
</tr>
<tr>
<td>Anomalous – Plausible</td>
<td>3</td>
<td>L. Superior Frontal Gyrus</td>
<td>(-45, 9, 21)</td>
<td>831</td>
<td>3.6</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>L. Inferior Frontal, Precentral Sulcus</td>
<td>(-21, 28, 38)</td>
<td>236</td>
<td>3.1</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>L. Angular Gyrus, L. Intraparietal Sulcus</td>
<td>(-38, -49, 33)</td>
<td>355</td>
<td>3.6</td>
</tr>
<tr>
<td></td>
<td>6</td>
<td>L. Posterior Cingulate</td>
<td>(-12, -45, 30)</td>
<td>265</td>
<td>3.0</td>
</tr>
<tr>
<td></td>
<td>7</td>
<td>R. Middle Frontal Gyrus</td>
<td>(33, 31, 29)</td>
<td>379</td>
<td>3.5</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(28, 17, 50)</td>
<td>331</td>
<td>2.9</td>
</tr>
<tr>
<td></td>
<td>8</td>
<td>R. Angular Gyrus, Sulcus</td>
<td>(42, -54, 41)</td>
<td>354</td>
<td>3.0</td>
</tr>
<tr>
<td></td>
<td>9</td>
<td>R. Anterior Cingulate</td>
<td>(1, 31, 26)</td>
<td>395</td>
<td>3.8</td>
</tr>
</tbody>
</table>

The ROI #s above correspond to Figure 2 and represent significant clusters for the average activation in all participants at the comprehension question (14-20 s into the stimulus presentation). X, Y, and Z represent Talairach and Tournoux (Talairach & Tournoux, 1988) coordinates. No areas of activation were found the Anomalous – Implausible contrast.
Figure Captions

Figure 1. Significant ROIs for average neural activity in all subjects (high and low span) at the critical word (8-14 s into stimulus presentation) are depicted on the lateral surface of the left and right hemispheres for 3 comparisons. *Activation for Implausible – Plausible:* 1. L. inferior frontal sulcus, middle frontal gyrus, 2. R. postcentral sulcus. *Activation for Anomalous – Plausible:* 3. L. inferior frontal sulcus, precentral sulcus, 4. L. intraparietal sulcus, 5. R. middle frontal gyrus, inferior frontal sulcus, pars triangularis, 6. R. angular gyrus. *Activation for Anomalous – Implausible:* 7. L. precentral sulcus, 8. R. postcentral gyrus. Hemodynamic response curves for these ROIs are shown.

Figure 2. Significant ROIs for average neural activity in all subjects (high and low span) at the comprehension question (14-20 s into stimulus presentation) are depicted on the lateral and medial surface of the left and right hemispheres. *Activation for Implausible – Plausible:* 1. L. frontal pole, 2. R. middle frontal gyrus. *Activation for Anomalous – Plausible:* 3. L. superior frontal gyrus, 4. L. inferior frontal, precentral sulcus, 5. L. angular gyrus, intraparietal sulcus, 6. R. middle frontal gyrus, 7. R. angular gyrus/sulcus, 8. L. posterior cingulate gyrus, 9. R. anterior cingulate sulcus. No significant differences were found for Anomalous – Implausible.
Figure 1.

L. Lateral View

p = 10^{-1.85}

Implausible - Plausible

Anomalous - Plausible

Anomalous - Implausible

R. Lateral View

Voxel-level thresholds

R. Postcentral Gyrus

p = 10^{-1.85}

Anomalous
Figure 2.

L. Lateral View R. Lateral View L. Medial View R. Medial View

Implausible - Plausible

1 2

3 4 5

6 7 8 9

Anomalous - Plausible

Voxel-level thresholds

p = 10^{-1.85}