

Ambiguity in the Brain: What Brain Imaging Reveals About the Processing of Syntactically Ambiguous Sentences

Robert A. Mason, Marcel Adam Just, Timothy A. Keller, and Patricia A. Carpenter
Carnegie Mellon University

Two fMRI studies investigated the time course and amplitude of brain activity in language-related areas during the processing of syntactically ambiguous sentences. In Experiment 1, higher levels of activation were found during the reading of unpreferred syntactic structures as well as more complex structures. In Experiments 2A and 2B higher levels of brain activation were found for ambiguous sentences compared with unambiguous sentences matched for syntactic complexity, even when the ambiguities were resolved in favor of the preferred syntactic construction (despite the absence of this difference in previous reading time results). Although results can be reconciled with either serial or parallel models of sentence parsing, they arguably fit better into the parallel framework. Serial models can admittedly be made consistent but only by including a parallel component. The fMRI data indicate the involvement of a parallel component in syntactic parsing that might be either a selection mechanism or a construction of multiple parses.

Functional neuroimaging provides a unique opportunity to gain insight into the processing of linguistic ambiguity, because it indicates how much brain activity is associated with the comprehension of different types of ambiguous and comparable unambiguous sentences. Purely behavioral studies have easily demonstrated that being led down a linguistic “garden path” (being led to interpret an ambiguity in favor of a more likely but ultimately incorrect interpretation) results in longer processing times and larger error probabilities, but these studies cannot provide a measure of the amount of computation that is being performed per unit time. fMRI offers a proxy for amount of computation per unit time, namely the amount of brain activity per unit time. With a single-sentence, event-related, experimental design, the present fMRI study provided a measurement of the amount of brain activity every 1,500 ms for different types of ambiguous and unambiguous sentences.

The reason that syntactic ambiguity is inherently interesting is because it presents the cognitive system with a choice, a fork in the

road of parsing. A representation of any sentence is incrementally constructed as each successive word of a sentence is read. When a word in which the structural interpretation is ambiguous is encountered, one of several plausible parsing strategies could be applied. Much research in psycholinguistics has been concerned with empirically determining which one of the plausible strategies is actually used by human comprehenders. What occurs at the choice point is likely to be indicative of more general strategic and architectural properties of the language processing system.

When the syntactically ambiguous word is encountered, one way to deal with it is simply to choose one of the interpretations and discard the other. This single-parse strategy can be considered a serial model. An alternative strategy is to simultaneously construct dual parses corresponding to the two interpretations of the ambiguous word. This has been referred to as a parallel model. Two recent reviews of the parsing literature (Gibson & Pearlmut-ter, 2000; Lewis, 2000) have indicated that there are two classes of viable parsing models that can account for the behavioral data collected thus far. In probabilistic serial models (Traxler, Pickering, & Clifton, 1998), the determination of which single parse to follow is made on the basis of some type of race-horse selection of which parse is more likely. In ranked parallel models (Earley, 1970; Gibson, 1998; Gibson & Pearlmut-ter, 1998; Jurafsky, 1996; Pearlmut-ter & Mendelsohn, 1999; Spivey & Tanenhaus, 1998; Stevenson, 1994), there are mechanisms for ranking the likelihood of the alternative parses and for following both of them as long as resources are available. In the extreme case of no additional resources being available, the ranked parallel model will reduce to a serial model. It is also worth noting that the probabilistic serial model could legitimately be classified as a hybrid model because the consideration of parses is done in parallel. Both of these classes of models have some type of a reanalysis component for error recovery in cases in which only the incorrect parse remains active.

Both Lewis (2000) and Gibson and Pearlmut-ter (2000) have proposed that existing evidence makes it no longer appropriate to simply ask if parsing is a serial or a parallel process. Gibson and Pearlmut-ter proposed that the critical question becomes “whether

Robert A. Mason, Marcel Adam Just, Timothy A. Keller, and Patricia A. Carpenter, Department of Psychology, Center for Cognitive Brain Imaging, Carnegie Mellon University.

Portions of this work were reported at the 40th Annual Meeting of the Psychonomic Society, Los Angeles, CA, November, 1999. This research was supported by the National Institute of Mental Health Grant MH-29617; National Institute of Mental Health, Senior Scientist Awards, MH-00661 and MH-00662; and the National Institute for Neurological Disorders and Stroke Grant PO1-NS35949.

We appreciate the assistance of current and former members of the Center for help in conducting the experiments, particularly Jennifer Roth, Kurt Schimmel, and Holly Zajac for their additional assistance in defining regions of interest. We thank Victor A. Stenger for assistance with the spiral pulse sequence. We also thank Erik Reichle, F. Gregory Ashby, Keith Rayner, Fernanda Ferreira, and four anonymous reviewers for helpful comments on earlier versions of this article.

Correspondence concerning this article should be addressed to Robert A. Mason, Department of Psychology, Center for Cognitive Brain Imaging, Carnegie Mellon University, Pittsburgh, Pennsylvania 15213. E-mail: rmason@andrew.cmu.edu

or not there are some circumstances in which multiple constructions are maintained" (p. 231). This question has not been easily decidable. Brain imaging offers an exciting new technique that helps illuminate some conditions under which multiple constructions are considered and/or constructed. The processes for considering multiple parses or the construction of multiple parses should be accompanied by an increase in cognitive workload measurable with fMRI. The brain imaging data we report below provide new information about the processing of a specific type of ambiguity (i.e., the main verb/reduced relative [MV/RR] ambiguity), which helps to constrain possible classes of parsing models. Specifically we report that there is additional processing as shown in cortical activation during the reading of ambiguous-preferred sentences that has not been found in reading times. We propose that parsing models must have a mechanism that allows them to account for an increase in resource consumption during the processing of temporarily ambiguous syntactic constructions. Thus, parallel models within which multiple parses are maintained until disambiguation will be consistent with our data. Similarly, serial models with some type of parallel, resource-consuming mechanism for making a probabilistic assessment of which parse to follow will be consistent (Frazier & Clifton, 1996; Traxler et al., 1998). In addition to shedding light on this particular parsing issue, the brain imaging data also advance our knowledge about the cortical areas involved in syntactic parsing.

Before describing the possible parsing strategies in more detail, we describe a syntactic ambiguity that has been the object of much previous research, largely because it provides a good venue to study this issue. Below is an example of a temporarily syntactically ambiguous sentence, the MV/RR; both versions are presented here:

1. MV: "The experienced soldiers warned about the dangers before the midnight raid."
2. RR: "The experienced soldiers warned about the dangers conducted the midnight raid."

In the ambiguous sentences, the point of ambiguity is at the verb *warned*. This word can be interpreted as the main verb of the sentence, as in Sentence 1, or it can be interpreted as a past participle to begin the formation of a reduced relative clause, reduced from "soldiers who were warned. . ." as in Sentence 2. These ambiguous sentences can be contrasted with unambiguous counterparts that maintain the same sentence structures such as the following:

3. MV control: "The experienced soldiers spoke about the dangers before the midnight raid."
4. RR control: "The experienced soldiers who were told about the dangers conducted the midnight raid."

There has been considerable debate concerning the determination of which parse to follow in either a probabilistic serial model or a ranked parallel model. With respect to this example, the first interpretation is said to be the preferred interpretation of the ambiguity, relative to the second interpretation. In our study, the RR sentences were always the less likely parses than the MV structures. Therefore, from now on we refer to the RR sentences as

unpreferred and MV sentences as *preferred* when it is not necessary to specify the sentence type (as in a contrast with prepositional phrase [PP]-attachment sentences).

When an ambiguity is ultimately resolved in favor of the unpreferred interpretation, the sentence is referred to as a garden-path sentence (e.g., Bever, 1970). The name arises from the view that readers (or listeners) seem to follow an erroneous parse of the sentence, either on the basis of frequency of occurrence or the use of a specific rule-based preference. By the time decisive disambiguating information becomes available, the reader has already traveled down a path towards the incorrect interpretation and has been garden pathed.

Researchers have previously measured reading times and error rates to determine the parsing model that best describes the functioning of the human sentence comprehension system. Many of the predictions concern the comparison between ambiguous and comparable unambiguous sentences. In early work using the MV/RR ambiguity and several other types of syntactic ambiguities, the serial model was generally supported by the empirical findings. Very often, no differences in behavioral measures of performance were found between ambiguous and unambiguous sentences, so long as the ambiguous sentences were resolved in favor of the preferred interpretation and did not have a strong biasing context (e.g., Frazier & Rayner, 1982). Furthermore, when ambiguous sentences were resolved in favor of the unpreferred interpretation, this resulted in longer self-paced reading times (e.g., Taraban & McClelland, 1988), longer reading times as measured by oculomotor activity (e.g., Frazier & Rayner 1982), and lower grammaticality judgments than their unambiguous counterparts (e.g., Frazier, 1978). Thus, serial models can account for the behavioral data because they assume that there is no workload associated with selecting the preferred parse. The hybrid probabilistic serial model and the ranked parallel model both assume that the workload and reading time may not be a sensitive enough measurement.

The difficulty with the conclusion that there is no additional processing load for ambiguous sentences resolved in favor of the preferred interpretation is that measuring processing load during comprehension is difficult. Note that the conclusion that there is no additional processing during the reading of ambiguous main verb sentences compared with unambiguous main verb sentences is a null prediction. The hypothesized increased processing intensity could be manifested in two possible ways: It may be reflected in longer reading times with the same brain activation intensity per unit time or it may be seen not in the reading time at all but only in higher brain activation intensity. Thus, we may discover that there is an increase in cortical processing for these ambiguous but preferred sentences in the absence of a reading time difference.

The consideration of parsing strategies and amount of cognitive resources used in sentence processing can be related to the science of cognitive brain imaging. A key linking assumption is that, within some dynamic range, an increase in the amount of cognitive processing will be reflected in an increase in the amount of brain activation. For example, as the structure of a sentence is made more complex (holding the lexical content constant), the comprehension processes result in more and more cortical activity in terms of both the volume of activation and its amplitude (Just, Carpenter, Keller, Eddy, & Thulborn, 1996). It is possible to relate the various hypotheses about sentence parsing strategies to differential predictions about cortical activity by considering the amount

of computational load predicted by each model for the various types of sentences.

Consider first the class of ranked parallel models. During the processing of ambiguous sentences, there are times in which two or more possible parses are maintained in parallel. This maintenance of multiple parses should be more resource consuming and should be manifested as additional cortical activity. Furthermore, because the ranked parallel models allow for the pruning of lower ranked possible parses, they also predict that garden-path reprocessing will occur. To summarize, the ranked parallel models predict additional cortical activity during the processing of any ambiguous sentence during the time in which there are sufficient resources to maintain multiple parses. In addition, in the case of insufficient resources, increased cortical activity is expected if the disambiguating information favors a parse that has not been maintained; this prediction is simply that a garden-path effect will produce additional brain activation.

Probabilistic serial models also have little difficulty accounting for a garden-path effect in cortical activity. Similar to the pruned parses in the ranked parallel models, an unpreferred parse is not in working memory when the disambiguating information is encountered. These models are also consistent with the expectation of additional cortical activity that is due to the need to reparse the sentence. However, unlike ranked parallel models, probabilistic serial models predict no difference in the brain activation associated with the processing of ambiguous versus unambiguous sentences provided that the ambiguous sentences are resolved in favor of the most probable resolution. It is possible to generate a prediction of greater activation for ambiguous sentences from a serial model but only with an additional assumption of a resource-consuming parallel processing component. In hybrid probabilistic serial models, there must be some type of parallel mechanism for making a probabilistic assessment of which parse to follow, such as a thematic processor with a race-based mechanism (Frazier & Clifton, 1996; Traxler et al., 1998). If it is assumed that this choice process also consumes resources during the selection of a single parse, these hybrid models would also be consistent with additional cortical activity in main verb sentences even in the absence of a reading time effect. Experiment 2 tests the critical distinction between ranked parallel and most probabilistic serial models, not including this hybrid variation.

Previous neuroimaging research indicates some of the brain locations associated with sentence parsing. Syntactic processing is associated with activity in several areas; the most prominent among them are two language-processing areas—left inferior frontal gyrus (Broca's area) and left superior/middle temporal gyrus (Wernicke's area). In PET studies using a task subtraction method, researchers have focused on Broca's area by further dividing it into pars triangularis and pars opercularis in the search for a syntactic focal point (Caplan, Alpert, & Waters, 1998, 1999; Stromswold, Caplan, Alpert, & Rauch, 1996). In an fMRI investigation of syntactic processing, Just et al. (1996) found that not only did Broca's and Wernicke's areas show greater activation during the reading of more difficult syntactic constructions but their right homologues also showed an increase in activation. The data from these studies do not consistently point to a single location in the cerebral cortex being the site of syntactic processing but instead indicate that a network of areas participates in syntactic processing. Although these studies validate the idea of a language net-

work, they also demonstrate that syntactic processing is driven largely by activity in the inferior frontal gyrus and the posterior, superior, and middle temporal gyri. Thus, our investigation focuses on Broca's and Wernicke's areas, two prominent members of the network. This focus excludes data acquisition in other cortical areas that is superior and inferior to the band of selected areas, with the benefit of a higher sampling rate within the focused band.

An examination of which areas activate during the presentation of syntactically ambiguous sentences should help to refine our understanding of the language network. A strong localist hypothesis (i.e., that specific cognitive processing can be described as occurring in a single limited brain area) might lead us to expect increases that are due to ambiguity in only a single brain area. However, many functional neuroimaging studies suggest that it is more likely that we would see an effect of syntactic difficulty in both major parts of the language network, Broca's and Wernicke's areas (e.g., Just et al., 1996). Of interest is the relative magnitude of any ambiguity effect in these two areas. The magnitude could be similar in the two areas, or it could be different, and in the extreme, it could be null in one of the two areas. Furthermore, the effect of processing an ambiguity and any effect that is due to reanalysis of an incorrectly generated parse may involve two areas differentially.

Experiments 1, 2a, and 2b measured the fMRI response to individual sentences, using an event-related paradigm (e.g., Buckner et al., 1996; Carpenter, Just, Keller, Eddy, & Thulborn, 1999; Dale & Buckner, 1997). This enabled us to examine the time course of the fMRI response to individual sentences. The paradigm makes it possible to measure the brain activation associated with the comprehension of different types of ambiguities and allows the comparison of the processing of ambiguities and nonambiguous control sentences. In addition, this design permits the randomization of the presentation order of different types of items, an important issue in studies of ambiguity. Experiment 1 compared the brain activation during the reading of two types of ambiguous sentences. Experiments 2a and 2b compared ambiguous sentences with unambiguous sentences.

Experiment 1

Although we have focused the discussion so far on the difference between the processing of ambiguous and unambiguous sentences, it is important to first establish that brain activation is sensitive to the extra processing involved during the reading of garden-path ambiguous sentences. Behavioral research has consistently shown that in the absence of prior biasing context, garden-path ambiguous sentences result in longer reading times than the more preferred parse of an ambiguous sentence. For this reason, the main purpose of Experiment 1 was to compare the intensity and time course of the brain activation associated with the processing of ambiguous sentences that were resolved with either the preferred or unpreferred interpretation, always presenting ambiguous sentences. A second purpose was to compare the comprehension of two types of ambiguity: sentences that were ambiguous with respect to prepositional phrase attachment versus reduced relative clause/main verb construction. The time course of the brain activation in each of the four conditions was measured using fMRI.

Table 1
Sample Sentences for Experiments 1 and 2

Sentence type	Sentence
Experiment 1	
PP attachment sentences	
Preferred PP to VP (control)	Laura cleaned the kitchen floor with Clorox bleach before going to bed last night.
Unpreferred PP to NP (experimental)	Laura cleaned the kitchen floor with scuff marks before going to bed last night.
Reduced-relative clause sentences	
Preferred MV (control)	The experienced soldiers warned about the dangers before the midnight raid.
Unpreferred RR (experimental)	The experienced soldiers warned about the dangers conducted the midnight raid.
Experiment 2	
Unambiguous sentences	
Preferred MV (control)	The experienced soldiers spoke about the dangers before the midnight raid.
Unpreferred RR (experimental)	The experienced soldiers who were told about the dangers conducted the midnight raid.
Ambiguous sentences	
Preferred MV (control)	The experienced soldiers warned about the dangers before the midnight raid.
Unpreferred RR (experimental)	The experienced soldiers warned about the dangers conducted the midnight raid.

Note. PP = prepositional phrase; VP = verb phrase; NP = noun phrase; MV = main verb; RR = reduced relative.

There are many other types of syntactically ambiguous constructions than the MV/RR ambiguity that we have used as an example above. A second type that has been researched in the psycholinguistic literature involves prepositional phrase attachments, as in the following:

5. PP attached to verb phrase (VP): "The landlord painted all the walls with enamel though it didn't help the appearance of the place."
6. PP attached to noun phrase (NP): "The landlord painted all the walls with cracks though it didn't help the appearance of the place."

These two sentences are identical up to the ambiguous prepositional phrase "with [NP]." At that point the sentence is ambiguous; the PP can be attached to the verb or to the immediately preceding NP. The preferred interpretation is to attach the PP to the verb, as is the case in Sentence 5, whereas the unpreferred interpretation is to attach the PP to the preceding NP.¹ Under assumptions defined previously concerning cortical activity, there should be more cortical activity during the comprehension of Sentence 6 than Sentence 5.

In the case of the MV/RR ambiguity, both the probabilistic serial model and the ranked parallel model predict more brain activation for the unpreferred interpretation, namely the RR as described previously. Although the two types of ambiguity have not been compared in a single study, across studies the RR constructions, as seen in Sentence 2, typically have resulted in longer processing times than the PP constructions as seen in Sentence 6

(RR in MacDonald, Just, & Carpenter, 1992; PP in Rayner, Carlson, & Frazier 1983). The RR sentences should therefore result in more cortical activity than the PP sentences.

Method

Participants

In Experiment 1 the participants were 10 right-handed paid volunteer college students (3 women). Each participant gave signed informed consent that had been approved by the University of Pittsburgh and the Carnegie Mellon Institutional Review Boards. Participants were familiarized with the scanner, the fMRI procedure, and the sentence comprehension task before the study started.

Materials

Many of the stimulus items were identical or slight modifications of sentences that have been used in various syntactic ambiguity behavioral studies (MacDonald et al., 1992; Rayner, Carlson, & Frazier, 1983). A sample set of sentences appears in Table 1. Participants read a total of 40

¹ The unpreferred structure for the PP attachment sentences is said to violate the principle of minimal attachment (Frazier, 1978). The minimal attachment principle posits that a new phrase is attached to an unfolding syntactic representation in the simplest manner possible. In the unpreferred structure, this principle is violated because attaching the PP to the NP would require the generation of a complex NP containing both the "the kitchen floor" NP and the "with scuff marks" PP. The unpreferred structure (the RR construction) in the MV/RR ambiguity also violates the principle of minimal attachment.

sentences, 10 sentences in each of four conditions in the study. The same quasi-random presentation order was used for all participants, using a Latin square design.

Four 30-s fixation epochs, consisting of an "X" at the center of the screen, were presented at the beginning, end, and at approximate trisections of the sentence set, to provide a baseline measure of activation. All the remaining intersentence intervals were filled with a 12-s rest period, also consisting of a centered X, to allow the hemodynamic response to approach baseline between sentences.

Presentation

Each single trial began with the entire sentence being presented for 10 s. Eighty-five percent of all reading times in a pilot behavioral experiment fell into this range. For the RR-unpreferred condition, 82% of the reading times fell into this range (*M* reading times in the four conditions ranged from 6.3 s to 7.3 s).

A yes-no comprehension question immediately followed the sentence. The comprehension questions were designed to be sure that the participant was reading the sentences. Care was taken so that the questions did not always refer to thematic roles. The purpose of this was so that readers would not anticipate a question referring to alternative readings of the ambiguous sentences and thus cause them to read in a more strategic and less natural fashion. Participants were told to respond as quickly as possible within a 4-s limit. Few failures to respond within the time limit occurred (approximately 4% of the trials in all experiments; response failures did not vary significantly across conditions). No items were excluded from the analysis because of incorrect responses. After the participant answered the question or 4 s had elapsed, an X appeared on the screen for the rest period. The sentence presentation, probe presentation, response, and the 12-s rest that followed constituted between 23 and 26 s, depending on the response time.

Scanning Procedures

A seven-slice oblique axial prescription (approximately 10° angle relative to a straight axial) was set that covered the middle to superior portions of the temporal lobe (i.e., superior temporal gyrus [STG]; including Wernicke's area) and the inferior frontal gyrus (IFG; including Broca's area). Figure 1 shows the location of the slices for 1 of the participants. The onset of each sentence was synchronized with the beginning of the acquisition of the superiormost slice (Slice 0).

Cerebral activation was measured using blood oxygenation level-dependent (BOLD) contrast (Kwong et al., 1992; Ogawa, Lee, Kay, & Tank, 1990). Imaging was done on a 3.0T scanner at the MR Research Center at the University of Pittsburgh Medical Center. The acquisition

parameters for the gradient-echo EPI with seven oblique axial slices were as follows: TR = 1.5 s, TE = 25 ms, flip angle = 90°, 128 × 64 acquisition matrix, 5-mm thickness, 1-mm gap, RF head coil. The structural images with which the functional images were coregistered were 124-slice, axial, T₁-weighted 3-D SPGR volume scans that were acquired in the same session for each participant with TR = 25 ms, TE = 4 ms, flip angle = 40°, FOV = 18 cm, and a 256 × 256 matrix size.

Data Analysis

The functional activation was assessed in two main regions of interest (ROIs) that were defined in each hemisphere using an anatomical parcellation method, one that relies on limiting sulci and anatomically landmarked coronal planes to segment cortical regions (Caviness, Meyer, Markris, & Kennedy, 1996; Rademacher, Galaburda, Kennedy, Filipek, & Caviness, 1992). As shown in Figure 2, the STG ROIs included the posterior, superior (T1a and T1p or BA22), and middle temporal gyrus regions (T2a, T2p, and TO2 or BA22 and 37). The IFG (inferior frontal gyrus) ROIs included orbital, pars triangularis, and pars opercularis portions of the IFG region (FOC, F3t and F30or BA44, 45 and 47). The ROIs in the functional images were defined for each participant with respect to coregistered structural images. The main focus of the data analysis was on these two ROIs in the left hemisphere.

The interrater reliability of this ROI-defining procedure between two trained staff members was evaluated for four ROIs in 2 participants in another study in this laboratory. The reliability measure was obtained by dividing the size of the set of voxels that overlapped between the two raters by the mean of their two set sizes. The resulting eight reliability measures were in the 78% to 91% range, with a mean of 84%, which is as high as the reliability reported by the developers of the parcellation scheme.

The image preprocessing corrected for in-plane head motion and signal drift by using procedures and software developed by Eddy, Fitzgerald, Genovese, Mockus, and Noll (1996). Data sets with large amounts of in-plane or out-of-plane motion were discarded without further analysis.

The voxels of interest within the four ROIs were identified by computing separate voxel-wise *t* statistics (using a threshold of *t* > 5.0) that compared the activation for the baseline fixation condition with the combination of all experimental conditions. The mean total number of voxels in all ROIs was 1,520. A *t* threshold greater than 5.0 was selected to give a Bonferroni-corrected alpha level of *p* < .025 after taking into account the average number of voxels and approximately 70 degrees of freedom for each of the voxel-wise *t* tests within a participant.

Time Series Analysis

The time series data for each voxel consisted of the raw signal intensity in 16 consecutive images (i1-i16), acquired 1,500 ms apart. A mean time series for each activated voxel of each participant (*M* activated voxels for left IFG = 12 and left STG = 16, using the *t* > 5.0 threshold) was formed by collapsing across the 10 sentence tokens per condition in the experiment. These 16 intervals were then segmented into three separate interval regions: i1-i4; i5-i10; and i11-i16. The first interval region (IR1) consisted of data that were collected during the first 6 s of each trial, during which the hemodynamic response was rising but had not reached asymptotic levels. This interval region is typically discarded in block epoch designs, and it was expected that few if any differences that were due to the experimental manipulation would be revealed in this interval. The second interval region (IR2) reflected the time in which hemodynamic response was near asymptotic activity levels, reflecting the encoding and comprehension of the sentences. The end of this region corresponded to 6 s after the offset of the sentence and onset of the question; this 6 s is equivalent to the delay of the hemodynamic response's rise to asymptote. Within the third interval region (IR3), the hemodynamic response reflected the late processing of the question and was decreasing in response to the fixation

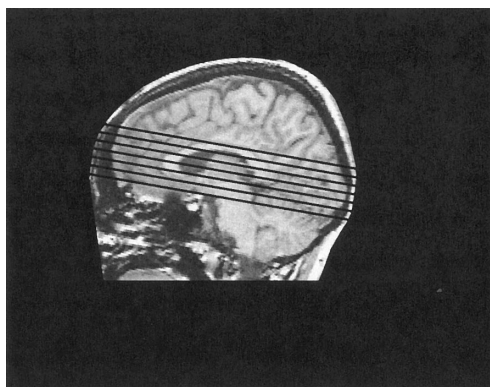


Figure 1. The slice prescription for a typical participant.

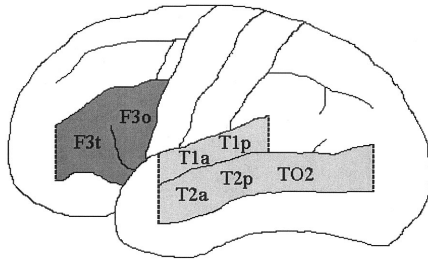


Figure 2. The anatomical areas included in the inferior frontal gyrus (IFG) and the superior temporal gyrus (STG) regions of interest (ROIs). IFG is shown in dark gray and includes F3o and F3t. The STG ROI is shown in light gray and includes T1a, T1p, T2a, and T2p.

point that signaled the end of the trial. The choice of adding a constant 6 s from the onset of the sentence for the beginning of IR2 and from the onset of the question for the beginning of IR3 is taken from an estimate of the rise of the hemodynamic response function to the response delay (e.g., Bandettini, Jesmanowicz, Wong, & Hyde, 1993). Inferential statistics were performed on the time-course curves as a whole and also on the three interval regions.

Results

Time Series

The time-series results show that the brain activation intensities were higher for unpreferred sentences. The curves in Figures 3 and 4 show no differences across conditions for the first interval region, while the hemodynamic response was rising. However, after 6 s (approximately the fourth image), the curves began to diverge. The preferred versions' signal intensities quickly leveled off, whereas the unpreferred conditions continued their increase in intensity. It is also clear that the time-course curves are bimodal. The second mode is likely due to an increase in activity as a result

of reading and answering the question. After the second mode, the most difficult sentences (i.e., unpreferred reduced relatives) decayed to baseline from a higher intensity than the other sentence types and remained higher at each subsequent time slice.

Functional Imaging Analyses of Variance

The mean raw signal intensities were analyzed in four separate 2 (left IFG vs. left STG) $\times 2$ (preferred vs. unpreferred) $\times 2$ (MV/RR vs. PP) $\times N$ (intervals) analyses of variance (ANOVAs) that differed only in the number of intervals used in each analysis (where $N = 16$ for the combined analysis, $n = 4$ for IR1, and $ns = 6$ for IR2 and IR3). Effects were tested against participant variability by collapsing across active voxels for each. In all analyses reported, an alpha level of .05 was the criterion for statistical significance. Mean percentage changes from fixation baseline for all analyses are reported in Table 2.

Combined intervals analysis. As predicted, the unpreferred sentences resulted in higher signal intensity than the preferred sentences. This garden-path effect was significant, $F(1, 8) = 30.94$, $MSE = 172.198$. Two other effects were only marginally significant in the participants. First, higher signal intensities were associated with the processing of MV/RR than with PP sentences, $F(1, 8) = 3.53$, $MSE = 145.307$, $p < .10$. Second, the higher signal intensity for unpreferred sentences over preferred sentences was greater in the MV/RR constructions than in the PP constructions, $F(1, 8) = 3.34$, $MSE = 297.404$, $p = .105$.

IR1. As expected, there were no significant or marginally significant differences for this region, the first 6 s of sentence processing, in the ANOVA on the basis of participant variability.

IR2. The mean signal intensity for the unpreferred condition was greater than the preferred condition, $F(1, 8) = 38.28$, $MSE = 144.796$. The greater signal intensity associated with the unpreferred condition was larger for the MV/RR sentences than for the prepositional phrase sentences, $F(1, 8) = 3.52$, $MSE = 89.61$, $p <$

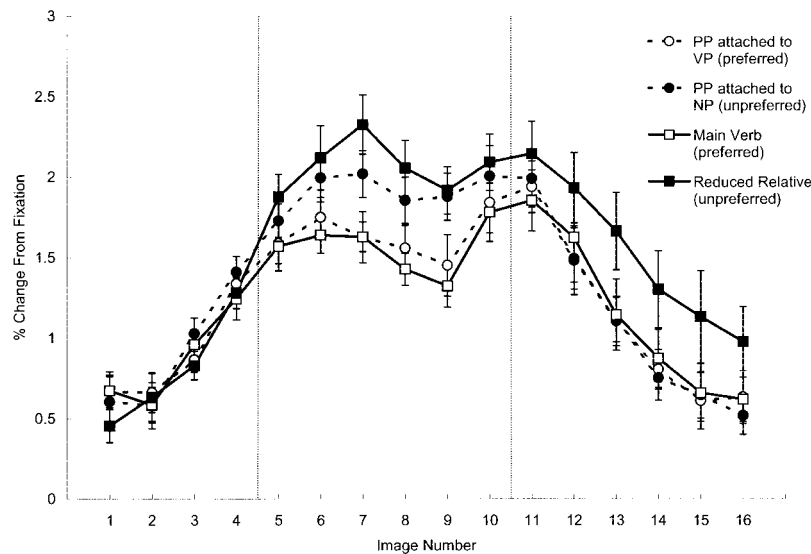


Figure 3. The average time course curves of the activated voxels for participants in the left inferior frontal gyrus (IFG) region of interest in Experiment 1 as measured in percentage change in signal intensity compared with the fixation condition

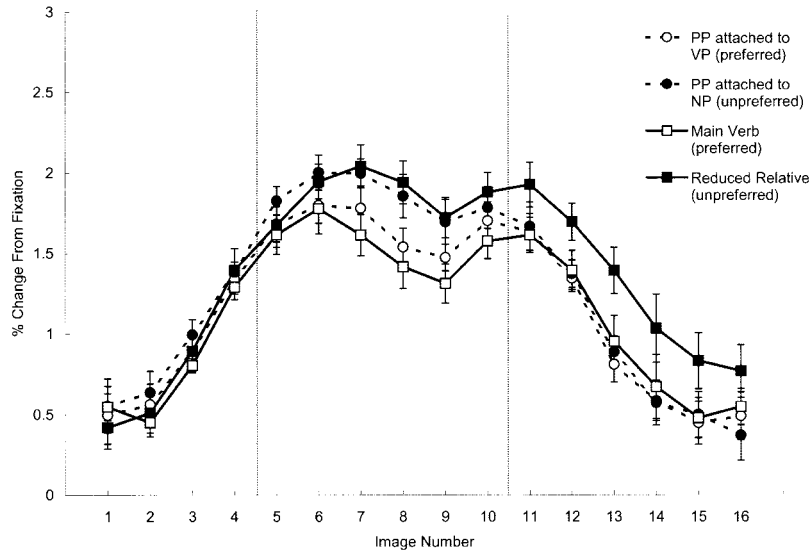


Figure 4. The average time course curves of the activated voxels for participants in the left superior temporal gyrus (STG) region of interest in Experiment 1 as measured in percentage change in signal intensity compared with the fixation condition.

.10. Furthermore, the additional processing required by the unpreferred sentences was larger in the left IFG than in the left STG; the ROI × Preference interaction was significant, $F(1, 8) = 8.17$, $MSE = 19.058$.

IR3. As in IR2, the unpreferred sentences were accompanied by higher signal intensities than the preferred sentences, $F(1, 8) = 11.36$, $MSE = 165.270$. In addition, MV/RR sentences had higher signal intensities than PP sentences, $F(1, 8) = 8.40$, $MSE = 284.981$. Furthermore, the difference between the unpreferred and

preferred conditions was larger for the MV/RR sentences than for the PP sentences, $F(1, 8) = 4.56$, $MSE = 324.113$, $p < .065$.

Right Hemisphere

Few participants showed activation that was detectable in this single-item paradigm in the right IFG and right STG ROIs. Only 4 of 10 participants showed any activation in right IFG and only 6 of 10 in right STG. In addition, those cases in which there were

Table 2

The Mean Percent Change in Signal Intensities for Left Broca and Left Temporal as a Function of Sentence Type (RR vs. PP) and Preference (Unpreferred vs. Preferred) for Experiment 1

Sentence type	Left Broca			Left temporal		
	RR	PP	Difference	RR	PP	Difference
Total (i1–i16)						
Unpreferred	1.54	1.35	0.19	1.38	1.26	0.12
Preferred	1.22	1.24	-0.02	1.13	1.16	-0.03
Difference	0.32	0.11		0.25	0.10	
Interval Region 1 (i1–i4)						
Unpreferred	0.80	0.90	-0.11	0.80	0.89	-0.09
Preferred	0.86	0.88	-0.02	0.77	0.82	-0.05
Difference	-0.07	0.02		0.03	0.07	
Interval Region 2 (i5–i10)						
Unpreferred	2.06	1.91	0.15	1.87	1.86	0.01
Preferred	1.56	1.63	-0.07	1.55	1.66	-0.11
Difference	0.50	0.28		0.32	0.20	
Interval Region 3 (i11–i16)						
Unpreferred	1.52	1.08	0.44	1.28	0.89	0.38
Preferred	1.13	1.09	0.03	0.94	0.88	0.06
Difference	0.40	-0.01		0.33	0.01	

Note. Intervals corresponding to specific tables are noted in parentheses. RR = reduced relative; PP = prepositional phrase; i = interval.

any activated voxels in the right hemisphere rarely amounted to more than three voxels of activation (one participant had nine activated voxels in right IFG and five voxels in right STG). Because of the sparse amount of data, analyses for these regions are not further reported.

Behavioral Performance

Two behavioral measures were collected during the experiment: response times to the comprehension questions and error rates on the comprehension questions. For prepositional attachment sentences, comprehension question response times were 1,954 ms for preferred sentences and 2,239 ms for unpreferred sentences, whereas for RR sentences, they were 2,136 ms for preferred and 2,568 ms for unpreferred. The comprehension question response times were longer for unpreferred sentences than preferred sentences, $F(1, 8) = 43.95$, $MSE = 26,337.795$, and longer for the MV/RR sentences compared with prepositional attachment sentences, $F(1, 8) = 7.78$, $MSE = 75,643.086$. Consistent with the signal intensity data, the longer response times for the unpreferred sentences were greater for MV/RR sentences than the prepositional attachment sentences; this interaction of ambiguity and preference was marginally significant, $F(1, 9) = 6.74$, $MSE = 9,705.704$, $p = .0616$. The average error rates for the four conditions were 6.2% for PP preferred, 3.7% for PP unpreferred, 7.4% for MV preferred, and 42% for RR unpreferred. The high error rate for the RR unpreferred resulted in a significant Sentence Type \times Preference interaction, $F(1, 8) = 30.77$, $MSE = 0.813$, as well as main effects of sentence type, $F(1, 8) = 25.13$, $MSE = 1.132$, and preference, $F(1, 8) = 19.45$, $MSE = 0.965$.

Discussion

Consistent with behavioral data, the fMRI results showed that additional brain activity occurs during the reading of unpreferred syntactic constructions. This additional processing was manifested in the higher signal intensity associated with the unpreferred sentences compared with the preferred sentences in the overall analysis as well as the IR2 and IR3 independent analyses. Furthermore, the effects were found in two brain regions known to participate in sentence comprehension. This first demonstration of a garden-path effect in imaging data was an indication of the power of the single-trial fMRI method and a validation of its use in fMRI experiments of language processing.

The suggestion of more brain activation for the MV/RR construction than the prepositional phrase construction may be predominantly due to the complex recovery associated with the unpreferred version of the reduced relatives. The preference effect was larger for the MV/RR sentences than PP sentences in both the overall analysis and in IR2. Furthermore, the trend toward a main effect of sentence type in the overall analysis was driven by the significantly higher levels of activation for the reduced relatives that did not appear until the final interval region (IR3).

Experiments 2a and 2b

The results of Experiment 1 allow us to return to the critical question of whether an ambiguity itself, regardless of how it is resolved, produces higher levels of activation than an unambigu-

ous sentence. As was seen in Experiment 1, there is additional cortical activity during the reading of ambiguous-unpreferred sentences compared with the reading of ambiguous-preferred sentences. This is consistent with a ranked parallel model. The construction–maintenance of multiple parses should show a measurable increase in intensity of processing. The ranking–pruning of the correct parse could have resulted in an increase in intensity of processing that was due to recovering the correct parse. The probabilistic serial model also predicts additional brain activation in this case as well. As in the ranked parallel model, the increase in processing could have been a consequence of forcing the parser to reanalyze the sentence on discovery of the incorrect structure. Thus, both models are consistent with the increased brain activity when an ambiguous sentence was resolved in favor of the unpreferred interpretation.

What occurs during the processing of the preferred sentences that is slightly different? The resource-based ranked parallel model predicts that there should be more brain activation during the processing of ambiguous sentences than unambiguous sentences irrespective of which interpretation is ultimately confirmed. Therefore, we would expect an ambiguity effect to be present for preferred sentences as well. This is in contrast to the prediction of the simple probabilistic serial model (without the assumption that the race selection mechanism consumes a significant amount of resources) that predicts no ambiguity effect so long as the ultimate interpretation is the preferred one.

To address the issue of the effect of ambiguity, readers in Experiment 2 were presented with both ambiguous and unambiguous sentences. The unambiguous control sentences were matched to the preferred and the unpreferred syntactic structures. Although the unambiguous controls for the reduced relatives construction were a full relative clause, they are referred to as the unambiguous unpreferred sentences for simplicity. To limit the number of items, only the MV/RR sentences from Experiment 1 were used. These items came primarily from MacDonald et al. (1992); however, an additional 16 items were generated using the MV/RR items from Experiment 1 as a template. This enabled us to increase the number of sentences within a condition from 6 to 10. Samples of the sentences presented in Experiment 2 appear in Table 1. Experiment 2a only included the experimental items. This resulted in half of the sentences including relative clauses, and half of those were garden-path sentences. In Experiment 2b, filler sentences were added. The inclusion of filler sentences was an attempt to prevent readers from focusing on a limited type of sentence structure.

Method

Participants

There were two groups of participants in Experiment 2. In Experiment 2a, the participants were 6 right-handed paid volunteer college students (3 women). In Experiment 2b the participants were 8 right-handed paid volunteer students (3 women). Each participant gave signed informed consent (approved by the University of Pittsburgh and the Carnegie Mellon Institutional Review Boards). Participants were familiarized with the scanner, the fMRI procedure, and the sentence comprehension task before the study started.

Materials and Procedure

As in Experiment 1, participants in Experiment 2a read a total of 40 sentences, 10 sentences in each of four conditions in the study. The same

random presentation order was used for all participants. Sentences were presented using a Latin square design.

Four 30-s fixation epochs, consisting of an X at the center of the screen, provided a baseline activation measure. They were presented at the beginning, end, and at approximate trisections of the study. In addition, the remaining intersentence intervals were filled with a 12-s rest period, also consisting of a centered X, to allow the hemodynamic response to approach baseline between test epochs. Presentation, scanning procedures, and data analysis were identical to Experiment 1.

There were several significant differences in the method for Experiment 2b. The same set of 40 experimental sentences were used; however, they were divided in half and presented in two consecutive functional acquisitions within the same scanning session. In addition, 20 filler items were added to the materials. These filler items did not contain temporary syntactic ambiguities of the type that we are studying and were split evenly across the two acquisitions. This resulted in two functional acquisitions during which the participant saw 30 trials, 20 experimental (5 in each of the four conditions), and 10 fillers for a total of 60 trials across the two acquisitions. Each acquisition was 15 min and 6 s in length. A break of approximately 2–5 min occurred between the two acquisitions during which the participant was not removed from the scanner and was instructed to hold his or her head completely still. The division of the experiment into two acquisitions was deemed necessary to limit the duration of a continuous functional acquisition.

Scanning Procedures

The scanning procedures for Experiment 2a were the same as in Experiment 1. For Experiment 2b, several aspects of the scanning procedure were different, including the scanner. Imaging was done on a 3.0T scanner at the MR Research Center at the University of Pittsburgh Medical Center using a spiral pulse sequence in which slices were not interleaved. Improvements in the scanner enabled us to use a 16 slice oblique axial prescription (approximately 10° angle) while using the same TR. The 16 slices were selected to ensure coverage of the middle to superior portions of the temporal lobe (STG, including Wernicke’s area) and the IFG (including Broca’s area). The onset of each sentence was synchronized with the beginning of the acquisition of the most superior slice (Slice 0).

The acquisition parameters for the spiral scan pulse sequence with 16 oblique axial slices were as follows: TR = 1.5 s, TE = 18 ms, flip angle = 90°, 64 × 64 acquisition matrix, 5-mm thickness, 1-mm gap, RF head coil. The structural images with which the functional images were coregistered were 124-slice axial T₁-weighted 3-D SPGR volume scans that were acquired in the same session for each participant, with TR = 25 ms, TE = 4 ms, flip angle = 40°, FOV = 24 cm, and a 256 × 192 matrix size.

Results

Time Series

Comprehending ambiguous sentences produced higher levels of brain activation than comprehending unambiguous sentences, as shown in Figures 5, 6, 7, and 8. As in Experiment 1, there was no difference across conditions for the first interval region, but after 6 s the curves began to diverge. The signal intensity for the unpreferred-ambiguous sentences increased above the activity for the other three curves, especially in the left IFG. The preferred-ambiguous sentences did not increase in intensity as much as the preferred-ambiguous sentences. However, the critical finding was that the percentage change in signal intensity from fixation for the preferred-ambiguous sentences was greater than that of the preferred-unambiguous sentences for almost every image in IR2 (the exception was two out of the six IR2 images in left IFG in Experiment 2a).

As in Experiment 1, inferential statistics were performed on the time-course curves as a whole and also on the separate interval regions, demarcated by the vertical lines in the time-course graphs.

Functional Imaging ANOVAs

The mean raw signal intensities were analyzed in four separate 2 (left IFG vs. left STG) × 2 (preferred MV vs. unpreferred RR) × 2 (unambiguous vs. ambiguous) × N (intervals) ANOVAs (where N = 16 for the combined analysis, n = 4 for IR1, and ns = 6 for

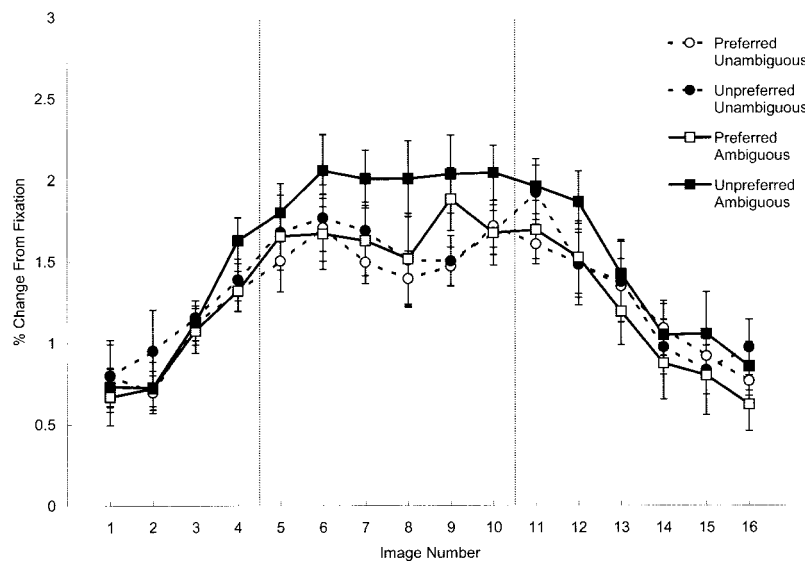


Figure 5. The average time course curves of the activated voxels for participants in the left inferior frontal gyrus (IFG) region of interest in Experiment 2a as measured in percentage change in signal intensity compared with the fixation condition.

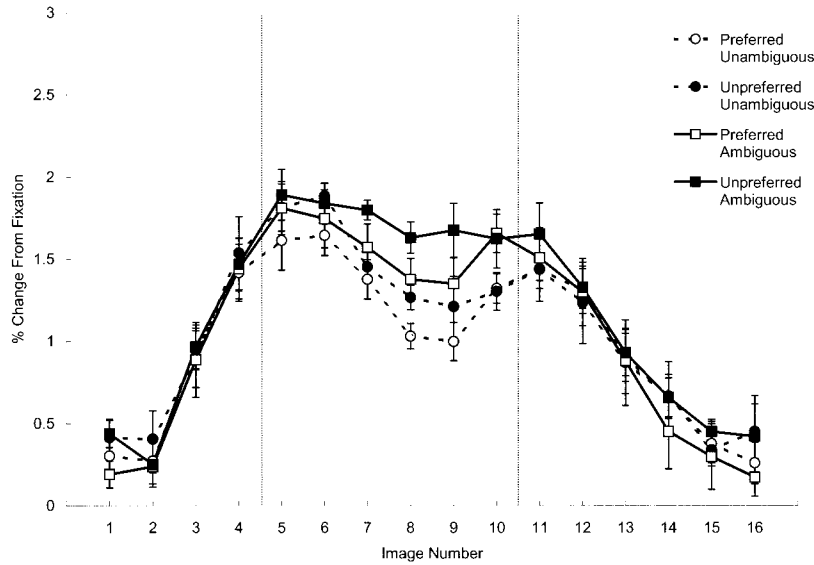


Figure 6. The average time course curves of the activated voxels for participants in the left superior temporal gyrus (STG) region of interest in Experiment 2a as measured in percentage change in signal intensity compared with the fixation condition.

IR2 and IR3). As in Experiment 1, effects were tested against participant variability by collapsing across active voxels for each participant for Experiment 2a (F_a). The mean raw signal intensities from Experiment 2b were analyzed as percentage change from a fixation baseline and tested against participant variability (F_b). For both analyses, an alpha level of .05 was the criterion for statistical significance. Mean percentage change from fixation baseline for all analyses are reported in Tables 3 and 4.

Combined analysis. Higher signal intensities were associated with the comprehension of ambiguous sentences than of unambig-

uous sentences, $F_a(1, 5) = 19.15$, $MSE = 36.935$, and $F_b(1, 7) = 24.00$, $MSE = 1.688$. In addition, the unpreferred sentences resulted in higher signal intensity than the preferred sentences. This effect was significant, $F_a(1, 5) = 23.82$, $MSE = 78.874$, and $F_b(1, 7) = 15.31$, $MSE = 0.912$. Thus, the main effects of both variables, ambiguity and sentence type, were significant and the two variables did not interact in the analysis of the entire time course. In addition, in Experiment 2b, the ambiguity effect was significant for both preferred sentences, $F_b(1, 7) = 11.70$, $MSE = 1.622$, and for unpreferred sentences, $F_b(1, 7) = 10.56$, $MSE = 2.045$.

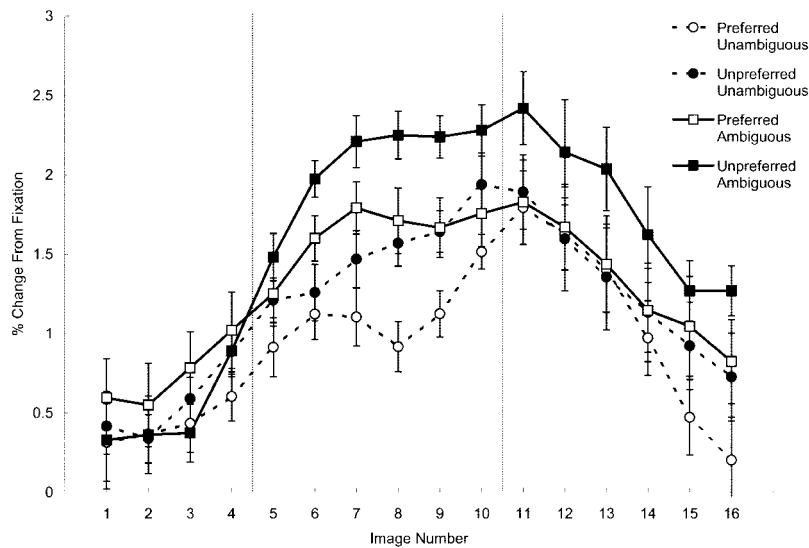


Figure 7. The average time course curves of the activated voxels for participants in the left inferior frontal gyrus (IFG) region of interest in Experiment 2b as measured in percentage change in signal intensity compared with the fixation condition.

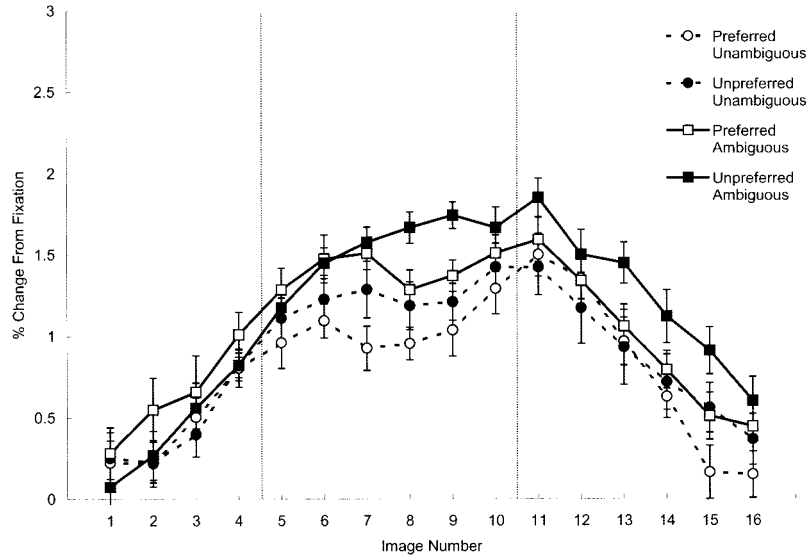


Figure 8. The average time course curves of the activated voxels for participants in the left superior temporal gyrus (STG) region of interest in Experiment 2b as measured in percentage change in signal intensity compared with the fixation condition.

IR1. As expected there were no significant differences for this region, the first 6 s of sentence processing, for either Experiments 2a or 2b.

IR2. Both of the experimental variables, ambiguity and sentence type, had a reliable effect in IR2. There were higher signal intensities associated with ambiguous sentences than unambiguous sentences, $F_a(1, 5) = 10.57$, $MSE = 180.182$, and $F_b(1, 7) = 19.56$, $MSE = 1.835$; likewise, there was greater activation following unpreferred constructions as opposed to preferred construc-

tions, $F_a(1, 5) = 14.86$, $MSE = 81.975$, and $F_b(1, 7) = 20.57$, $MSE = 1.414$.

To examine whether the disadvantage associated with ambiguous sentences occurred for both the preferred as well as the unpreferred sentences, separate ANOVAs were performed. There was a significant disadvantage for the ambiguous sentences compared with the unambiguous as seen in the test performed on the unpreferred conditions, $F_a(1, 5) = 15.21$, $MSE = 89.860$, and $F_b(1, 7) = 10.61$, $MSE = 1.356$. The ambiguity effect was present

Table 3
The Mean Percent Change in Signal Intensities for Left Broca and Left Temporal as a Function of Ambiguity (Ambiguous vs. Unambiguous) and Sentence Type (Unpreferred-RR vs. Preferred-MV) for Experiment 2a

Sentence type	Left Broca			Left temporal		
	Ambig.	Unambig.	Difference	Ambig.	Unambig.	Difference
Total (i1-i16)						
Unpreferred	1.52	1.36	0.17	1.19	1.07	0.11
Preferred	1.28	1.28	0.01	1.06	0.99	0.06
Difference	0.24	0.08		0.13	0.08	
Interval Region 1 (i1-i4)						
Unpreferred	1.05	1.07	-0.02	0.78	0.81	-0.03
Preferred	0.95	0.98	-0.04	0.69	0.73	-0.05
Difference	0.11	0.09		0.09	0.08	
Interval Region 2 (i5-i10)						
Unpreferred	1.99	1.64	0.36	1.74	1.49	0.25
Preferred	1.67	1.55	0.12	1.59	1.33	0.25
Difference	0.32	0.09		0.16	0.16	
Interval Region 3 (i11-i16)						
Unpreferred	1.37	1.26	0.11	0.91	0.83	0.07
Preferred	1.12	1.20	-0.08	0.77	0.82	-0.06
Difference	0.25	0.06		0.14	0.01	

Note. Intervals corresponding to specific tables are noted in parentheses. RR = reduced relative; MV = main verb; i = interval; Ambig. = ambiguous; Unambig. = unambiguous.

Table 4
The Mean Percent Change in Signal Intensities for Left Broca and Left Temporal as a Function of Ambiguity (Ambiguous vs. Unambiguous) and Sentence Type (Unpreferred-RR vs. Preferred-MV) for Experiment 2b

Sentence type	Left Broca			Left temporal		
	Ambig.	Unambig.	Difference	Ambig.	Unambig.	Difference
Total (i1–i16)						
Unpreferred	1.38	1.04	0.34	1.08	0.84	0.24
Preferred	1.13	0.81	0.32	0.98	0.75	0.23
Difference	0.24	0.22		0.10	0.09	
Interval Region 1 (i1–i4)						
Unpreferred	0.43	0.49	–0.06	0.40	0.40	0.01
Preferred	0.65	0.38	0.27	0.58	0.41	0.18
Difference	–0.22	0.11		–0.18	–0.01	
Interval Region 2 (i5–i10)						
Unpreferred	1.81	1.33	0.49	1.45	1.16	0.29
Preferred	1.43	0.98	0.45	1.32	0.98	0.34
Difference	0.39	0.35		0.13	0.18	
Interval Region 3 (i11–i16)						
Unpreferred	1.57	1.11	0.46	1.16	0.81	0.35
Preferred	1.16	0.94	0.22	0.90	0.74	0.15
Difference	0.41	0.17		0.26	0.06	

Note. Intervals corresponding to specific tables are noted in parentheses. RR = reduced relative; MV = main verb; i = interval; Ambig. = ambiguous; Unambig. = unambiguous.

in the preferred sentences as well, $F_a(1, 5) = 6.37$, $MSE = 96.033$, $p = .053$ and $F_b(1, 7) = 16.14$, $MSE = 0.924$.

There was an indication that left IFG played a larger role in the processing of the ambiguous sentences than did left STG; however, the evidence for this was slightly different in Experiments 2a and 2b. In Experiment 2b, the ambiguity effect was larger in left IFG than left STG; this ROI \times Ambiguity effect was significant, $F_b(3, 21) = 3.43$, $MSE = 0.413$. The sentence-type effect (i.e., unpreferred RR showed a greater percentage change from fixation than did preferred MV sentences) was also larger in left IFG than left STG but was only marginally significant, $F_b(3, 21) = 2.41$, $MSE = 0.503$, $p = .095$. In Experiment 2a, the ambiguity and preference effects varied across the two ROIs as indicated by a significant ROI \times Ambiguity \times Preference interaction, $F_a(1, 5) = 8.18$, $MSE = 18.370$. The large ambiguity effect in unpreferred sentences compared with the smaller ambiguity effect in preferred sentences held true for the voxels in left IFG, $F_a(1, 5) = 17.11$, $MSE = 12.776$, but the two variables did not interact in left STG ($p > .30$).

IR3. Although there were trends toward an ambiguity effect and a sentence-type effect in IR3, none of the differences tested were significant in Experiment 2a. In the Experiment 2b analysis, both the ambiguity and sentence type effects were significant: ambiguity, $F_b(1, 7) = 16.83$, $MSE = 1.007$, and preference, $F_b(1, 7) = 7.15$, $MSE = 3.375$. As was the case for the second region, the ambiguous sentences resulted in higher signal intensities than the unambiguous sentences for both the preferred sentences, $F_b(1, 7) = 6.28$, $MSE = 0.517$, and the unpreferred sentences, $F_b(1, 7) = 4.88$, $MSE = 3.232$.

Right Hemisphere

Only 1 participant showed activation in right STG. Even though 5 of the 6 participants had three or fewer activated voxels in right

IFG, the sparse amount of data precluded analyzing and reporting the data for these regions. In Experiment 2b, there was a larger number of participants with more than three active voxels on the right hemisphere (4 of 8 in IFG and 4 of 8 in STG).

Individual Voxel Specificities

Because there were two orthogonally manipulated independent variables, it was possible to examine the response characteristics of individual voxels in each of the four experimental conditions. Two alternative hypotheses were examined. The workload hypothesis posits that the same set of additional voxels are activated in a more demanding condition regardless of the nature of the comprehension demand (independent variable). The computation-specific hypothesis posits that which set of additional voxels is activated in a more demanding condition depends on the nature of the comprehension demand.

The activating voxels were sorted into mutually exclusive sets depending on the condition or conditions for which they were activated, defined as the t statistic, comparing their activation level to the fixation condition being above the $t > 5$ threshold.² For example, if a voxel was activated in the preferred condition regardless of ambiguity, that voxel would be in the preferred subset. With four conditions, there were 15 possible subsets ($2^4 - 1 = 15$) into which a voxel's behavior could have been classified. The mean size of each subset of voxels was expressed as a percentage of all activating voxels, for left IFG and left STG. A t test was

² This voxel subset activation pattern is consistent with the activation of voxel subsets from Experiment 1; in that study, it was also true that although there were activated voxels for specific components, the activation of individual voxels depended to a large degree on the amount rather than the type of computational demand.

Table 5
Mean Percent of Voxels Belonging to Each of the 15 Voxel Sets in Left STG and Left IFG, Labeled by the Set's Behavior in Each of the Four Conditions

Set no.	Condition				Region	
	Unambiguous preferred	Unambiguous unpreferred	Ambiguous preferred	Ambiguous unpreferred	Left STG	Left IFG
1	1	1	1	1	30.25*	22.17*
2	0	0	0	1	24.42*	29.25*
3	0	0	1	1	5.54	8.56*
4	0	1	0	1	2.63	5.78
5	1	1	0	0	0.00	0.46
6	1	0	1	0	0.00	0.37
7	1	0	0	0	3.13	3.14
8	0	1	0	0	1.85	1.98
9	0	0	1	0	4.60	2.92
10	1	0	0	1	2.39	4.98*
11	0	1	1	0	2.44	2.08
12	1	1	1	0	0.75	1.39
13	1	1	0	1	3.92	5.21*
14	1	0	1	1	5.09	5.12*
15	0	1	1	1	12.98*	6.58

Note. A 1 indicates that voxels in the set were above threshold in that condition. A 0 indicates that voxels in the set were below threshold in that condition. Sets 1 and 2 combined to account for over half of the activated voxels in the two regions of interest. Sets 3–6 indicate the proportion responding to either a specific sentence type or whether the sentence was ambiguous or unambiguous. Sets 7–9 include voxels that activated in only one of the four conditions with the exception of the ambiguous-unpreferred condition. Sets 10–15 collectively show the remaining subsets. STG = superior temporal gyrus; IFG = inferior frontal gyrus.

* $p < .05$ level, uncorrected for multiple comparisons.

performed on each subset of voxels to determine if the percentage of activated voxels was significantly different from zero.

The response characteristics of individual voxels suggest that their activation in this task depends primarily on the amount of computational demand rather than on the precise quality of that demand, consistent with the workload hypothesis. Two subsets accounted for the majority (approximately 50% for both left IFG and left STG) of the activated voxels. The largest subset consisted of those voxels that activated only in the most difficult condition, that is, when the sentence was an unpreferred-ambiguous sentence (24.42% for left STG and 29.25% for left IFG), shown as Set 2 in Table 5.³ This subset of voxels can therefore be characterized as activating only when the processing demand is extremely high. The second largest subset (Set 1) consisted of voxels that were activated in all four of the experimental conditions (30.25% for left STG and 22.17% for left IFG), showing an absence of specificity to the experimental variables. This suggests that most of the activating voxels in this/these regions were sensitive to the level of demand (the workload) rather than to each type of sentence (preferred MV or unpreferred RR) or to whether a sentence was ambiguous. The remaining sets of voxels that showed a meaningful amount of activation all included the most difficult condition.⁴

Behavioral Performance

There were no significant differences in the average question-answering response times among the various conditions in either Experiments 2a or 2b (the four conditional means ranged from 2,240 ms to 2,427 ms across the two experiments). The error rate was 5% in Experiment 2a for questions following preferred-

unambiguous sentences as well as questions following preferred-ambiguous sentences; the error rates for these two conditions in Experiment 2b were 1.3% and 1.9%, respectively. Error rates for comprehension questions following the unpreferred-unambiguous sentences were 13.3% for Experiment 2a and 2.5% for Experiment 2b; when the question followed the unpreferred-ambiguous sentences, these rates were 26.7% for Experiment 2a and 11.3% for Experiment 2b. As was the case in Experiment 1, the error rates for questions following the unpreferred-ambiguous sentences were greater than following the other types of sentences, which resulted in a significant Ambiguity \times Sentence Type interaction, $F_a(1, 5) = 16.00$, $MSE = 0.167$, and $F_b(1, 7) = 6.19$, $MSE = 0.426$.

³ Some aspects of data analysis here were different than in the analyses of the time course data. First, IR1, which revealed no difference in time course among conditions, was eliminated from consideration, limiting the analysis to IR2 and IR3 (Images 5 through 16). Second, the voxels of interest were those that showed a reliable difference in activation from the fixation condition in any one or more of the four experimental conditions.

⁴ The ordering of the two most active subsets differed slightly between left STG and left IFG. In left STG, the voxels that were active in all experimental conditions formed the largest subset, whereas the opposite was true in left IFG; the voxels that were active only during the processing of ambiguous, RR sentences formed the largest subset. This pattern may indicate that left IFG was more sensitive to the manipulation of syntactic difficulty than was left STG. This pattern is consistent with the analysis of the time course from Experiment 2a, which indicated that left IFG showed a larger ambiguity by preference interaction than did left STG. Furthermore, the set of voxels that responded if a sentence was ambiguous (Set 3) was significantly different from zero in left IFG but not in left STG.

Unlike the ambiguity effect in the signal intensity data, the ambiguity effect in the error rates was only present following unpreferred sentences, $F_a(1, 5) = 4.71$, $MSE = 1.133$, $p = .082$, and $F_b(1, 7) = 18.05$, $MSE = 0.339$, and not following preferred sentences ($F_s < 1$).⁵

Discussion

What was perhaps the most surprising result in this experiment was the finding of an ambiguity effect for preferred (MV) constructions. On-line behavioral results consistently showed little if any ambiguity effect in processing nonminimal attachment sentences such as the MV ambiguous sentences when they were resolved in the preferred direction (e.g., Binder, Duffy, & Rayner, 2001; Ferreira & Clifton, 1986; Frazier & Rayner, 1982). These behavioral results led to the development of single-parse theories that predict the absence of an ambiguity effect. For example, a key component of the classic garden-path theory was that the reader initially parses the sentence according to systematic syntactic preferences. Under this assumption there should be no additional processing associated with ambiguous versions of a preferred structure compared with unambiguous versions; in both cases the parsing follows the same (and ultimately correct) interpretation, and no additional structures or reanalysis are necessary. Nevertheless, the new results here indicated that there was a higher level of brain activity when the sentence was ambiguous, even if it was resolved in favor of the preferred syntactic structure.

Although the processing of ambiguous sentences need not consume more time than unambiguous sentences, it requires additional processing. It is clear from these results that there is a cost associated with processing an ambiguous sentence. If we assume that the additional activation is due to multiple parses being considered and/or constructed, then these results are consistent with models that allow multiple constructions to occur. Any of the ranked parallel parsing models, for example, can easily incorporate these findings (e.g., Gibson, 1998; Jurafsky, 1996; Pearlmuter & Mendelsohn, 1999).

As mentioned in the introduction, a hybrid version of the probabilistic serial model can be made consistent with these data, provided that an additional assumption is made. A mechanism by which the parser makes a probabilistic assessment of which parse to follow, such as a thematic processor with a race-based mechanism (Frazier & Clifton, 1996; Traxler et al., 1998), could consume resources during the selection of the most likely parse. Under such a model, the same prediction would be made as under a ranked parallel model; there should be additional cortical activity as a result of this probabilistic comparison but without the accompanying reading time. Although the results do not allow us to distinguish between this model and a ranked parallel model, the critical assumption underlying the hybrid probabilistic serial model is that a parallel process occurs so that a selection can be made of which parse to construct.

Our new results, showing that there is extra work associated with processing a syntactic ambiguity that is resolved in the preferred interpretation, rule out some models and leave some remaining contenders. The finding rules out contemporary models that do not make provision for either a parallel mechanism that selects a parse at the point of ambiguity or a parallel construction and maintenance of multiple parses. We have described two mod-

els that, with different degrees of awkwardness, can both account for the findings. Perhaps, if the temporal resolution of fMRI could be improved, an experiment in which the distance between the ambiguity and the disambiguation region is lengthened could enable a distinction to be made between these two types of models. The ranked parallel model predicts that the additional cortical activation should be maintained until the disambiguation, whereas the probabilistic serial model predicts that the additional activation should occur only near the point of ambiguity.

In addition to a race-based probabilistic serial model account, there is a hypothesis that would allow any probabilistic serial model to produce an ambiguity effect. It is likely that when an ambiguous verb is encountered, the incorrect interpretation is occasionally selected, requiring extra processing (and hence extra cortical activation) that is due to reanalysis. Because within the fMRI results we cannot distinguish between initial parsing and reanalysis, our ambiguity effect for the preferred (MV) sentences could have occurred because a reanalysis stage was necessary for a subset of the items. However, if this account were correct, one would expect the ambiguity effect to occur not only in the fMRI results but also in reading time results. That there was a discrepancy between the reading time results and brain imaging results for MV sentences leads to the conclusion that the parallel parsing hypothesis is more parsimonious. An effect of extra processing that is due to reanalysis almost incontrovertibly implies that there be extra processing time, whereas maintaining multiple parses could result in additional cognitive processing without an increase in processing time.

It is also possible that there are systematic individual differences to be found in the time at which the additional activation associated with the processing of ambiguous sentences occurs. MacDonald et al. (1992) found that high-span readers had longer reading times than low-span readers throughout the ambiguous region, whereas the low-span readers showed increased reading times only on encountering the disambiguating region in unpreferred constructions. A similar pattern might also emerge in future brain imaging studies. A direction for future research would be to select participants on the basis of their scores on the reading span task and look for individual differences in brain activation data that are analogous to MacDonald et al.'s reading time data.

In the future of ambiguity research, other types of syntactic ambiguities should be investigated, because each ambiguity type has its own characteristics. The ambiguous MV sentences used here are missing an argument that exists in the unambiguous versions and therefore may not be as good a sentence as the unambiguous versions.⁶ Although it is true that the verbs will miss a usually present argument in our ambiguous MV versions and this could incur additional processing as seen in cortical activity, it is unclear if the cognitive processing that is required would be

⁵ Although it is possible that the effect of reading the question may have contributed to the rise of the hemodynamic response in the last three images of IR2, processing the question was unlikely to have caused a difference across conditions in this region. This conclusion is supported by the fact that there were no reading time or error rate differences between ambiguous and unambiguous MV conditions.

⁶ We thank Fernanda Ferreira for pointing out this possibility as well as several ways to address the issue.

sufficient to result in the measurable increase in cortical activation that was found in Experiment 2. It is clear that further research could determine the viability of this alternative hypothesis. The conclusions from this article could be made more general by examining brain activation in the context of other types of ambiguities, such as the PP attached to VP sentences from Experiment 1. For example, by substituting “using” for “with” in “The landlord painted all the walls with enamel though it didn’t help the appearance of the place,” the sentence is made unambiguous, allowing for an ambiguity contrast. If an ambiguity effect is found with several types of ambiguities within the context of more likely interpretations, we would have stronger support for a parallel model.

Although the surprising ambiguity effect in the preferred sentences was perhaps the most interesting result, another goal in the second experiment was to simply investigate the overall ambiguity effect. As was seen in both the overall analysis and the analysis of the second interval region, ambiguous sentences were accompanied by higher signal intensities than unambiguous sentences. Several interactions in the second interval region are suggestive of the roles of left IFG and left STG during the processing of ambiguous sentences. The reanalysis–recovery from the unpreferred versions tended to produce higher activation levels in Broca’s area than in Wernicke’s area. This conclusion is based on the result that the extra resources required when the sentence was unpreferred and ambiguous were reflected in larger differences in signal intensity in left IFG than in left STG. This conclusion is consistent with previous results that suggest left IFG plays a central role in the processing of syntactic constructions (e.g., Caplan et al., 1998).

Although the time course of the intensity of activation in activated voxels tells a story about how the brain processes syntactic ambiguities, a look into the distribution of activated voxels suggests that the amount of brain recruitment is largely independent of the syntactic variables per se. Instead, the recruitment of voxels is indicative of workload in general. The voxel subset analyses showed that the largest subsets of voxels exceeded threshold in either all experimental conditions or in only the most difficult condition (unpreferred ambiguous). A strong localist hypothesis would suggest that there are sets of voxels (i.e., brain substrates) that activate specifically to various stimulus events, such as encountering an ambiguous sentence. If that were true, there would have been a large subset of voxels that exceeded threshold only when encountering any ambiguous sentence. This was clearly not the case; the subset of voxels that responded to ambiguous sentences was much larger when the sentences were resolved as an unpreferred construction than as a preferred construction. Moreover, it seems very unlikely that there is a set of voxels that activates specifically to sentences that are syntactically ambiguous and resolved in a RR construction. Considering that the unpreferred-ambiguous sentences have been shown to be difficult sentences to process (i.e., result in large reading times and lower grammaticality judgments), a plausible hypothesis is that general cognitive workload is predictive of voxel recruitment. Therefore, although the responses of sets of voxels reflect specific cognitive processing, the amount of cortical tissue used in a task seems to be indicative of a more general metric of cognitive workload.

General Discussion

Together, the results of the two experiments lead to several new conclusions. First, the ambiguous sentences evoked higher levels of brain activation, particularly when they were the unpreferred construction. However, even when the ambiguous sentences were resolved in favor of the preferred construction, they still produced higher levels of brain activation than a matched unambiguous sentence. This result is particularly salient because of the previous difficulty in finding differences between the processing of ambiguous and unambiguous sentences in these specific constructions when using traditional behavioral measures. Binder et al. (2001) failed to find a garden-path effect in eye-movement measures for preferred sentences even when thematic fit information and discourse context were biased toward the unpreferred meanings. The finding of additional cortical activation for ambiguous sentences regardless of sentence type is consistent with both ranked parallel models and hybrid probabilistic serial models, which include the assumption that the selection between alternatives consumes resources. Furthermore, probabilistic serial models may also be consistent with the data, provided that some of the ambiguous preferred sentences were erroneously parsed, thus requiring a reanalysis; however, the difference between reading time data and imaging data is troubling for these models.

Another important conclusion was that the brain activity increased as a function of the nonpreference and complexity of the sentence. The more complex sentences, like RR clauses, were accompanied by higher levels of activation than the PP sentences. Likewise, the unpreferred sentences were accompanied by higher levels of activation than the preferred sentences. For the most difficult sentences, the RRs, the amount of brain activation persisted at a high level for a longer period than the other conditions; that is, these sentences showed significantly higher levels of signal intensity into the third interval region. Last, the late syntactic processing, such as the reanalysis–recovery required by the unpreferred ambiguous sentences, appears to evoke more activation in Broca’s area than in Wernicke’s area.

The results from this brain imaging study offer much more information than just providing another dependent measure that confirms the difficulty of processing garden-path sentences. First, as described above, the results show the extra brain workload imposed by syntactic ambiguity. Second, the results show that the workload imposed by ambiguity and by garden-path resolutions is supported by at least two areas, left IFG and left STG, rather than just a single area being involved. It is quite likely that there are multiple processes involved in the extra workload. Third, the results provide an index of the temporal distribution of the work of left IFG and left STG in syntactic ambiguity processing.

Left IFG and left STG have somewhat different temporal profiles of activation, suggesting that they play slightly different yet interdigitated roles. Both left IFG and left STG are immediately recruited to handle the extra workload required by syntactically ambiguous sentences. This was seen in the increase in activation following the fourth image. Although both areas show an effect, the increase was greater for left IFG. Furthermore, only in left IFG did the increased activation extend over later images, including those images in which the hemodynamic response was probably associated with processing the probe question and even with the decay of activation to baseline. It would appear that left STG is

centrally involved in the processing of syntactic ambiguities but that more of the burden of any reanalysis–recovery from incorrect parses, or reactivation of a discarded parse, falls on left IFG. It is possible that left IFG (Broca's area) is involved in the internal generation of abstract syntactic representations that are reiteratively communicated to left STG (Wernicke's area) for interpretation and elaboration through the activation of semantic representations, an idea suggested by Keller, Carpenter, and Just (2001). This type of collaboration between the two areas could produce the pattern of data seen here. The generation of multiple syntactic constructions for ambiguous sentences would result in greater activation in left IFG compared with the activation associated with unambiguous sentences. Similarly, the elaboration of those multiple syntactic constructions would result in greater activation for ambiguous sentences in left STG. However, the considerable overlap in the semantic representations for the two constructions might result in the ambiguity effect being less in left STG than it would be in left IFG. This idea is consistent with ranked parallel models that propose resources are preserved through shared representations (e.g., Earley, 1970; Pearlmuter & Mendelsohn, 1999) as well as hybrid probabilistic serial models.

The results generally showed that brain activation, as measured through fMRI, is a very useful measure of cognitive processing. Current methods and techniques available in brain imaging research make it possible to examine the time course of brain activation during tasks as complex as reading. With these advances, it is now possible to use brain imaging data to inform cognitive theory and perhaps to make distinctions between theories that may be functionally different but indistinguishable in terms of behavioral reaction-time results. Reciprocally, research in brain imaging and cortical function benefits from being guided by cognitive models that make predictions about the time course and the content of the processing of sentences.

References

- Bandettini, B. A., Jesmanowicz, A., Wong, E. C., & Hyde, J. S. (1993). Processing strategies for functional MRI of the human brain. *Magnetic Resonance Imaging in Medicine*, *30*, 161–173.
- Bever, T. G. (1970). The cognitive basis for linguistic structure. In J.R. Hayes (Ed.), *Cognitive development of language* (pp. 279–363). New York: Wiley.
- Binder, K. S., Duffy, S. A., & Rayner, K. (2001). The effects of thematic fit and discourse context on syntactic ambiguity resolution. *Journal of Memory and Language*, *44*, 297–324.
- Buckner, R. L., Bandettini, P. A., O'Craven, K. M., Savoy, R. L., Peterson, S. E., Raichle, M. E., & Rosen, B. R. (1996). Detection of cortical activation during averaged single trials of a cognitive task using functional magnetic resonance imaging. *Proceedings of the National Academy of Sciences of the United States of America*, *93*, 4878–4883.
- Caplan, D., Alpert, N., & Waters, G. (1998). Effects of syntactic structure and propositional number on patterns of regional cerebral blood flow. *Journal of Cognitive Neuroscience*, *10*, 541–552.
- Caplan, D., Alpert, N., & Waters, G. (1999). PET studies of syntactic processing with auditory sentence presentation. *NeuroImage*, *9*, 343–351.
- Carpenter, P. A., Just, M. A., Keller, T. A., Eddy, W. F., & Thulborn, K. R. (1999). Graded functional activation in the visuospatial system with the amount of task demand. *Journal of Cognitive Neuroscience*, *11*, 9–24.
- Caviness, V. A., Meyer, J., Markris, N., & Kennedy, D. N. (1996). MRI-based topographic parcellation of human neocortex: An anatomically specified method with estimate of reliability. *Journal of Cognitive Neuroscience*, *8*, 566–587.
- Dale, A. M., & Buckner, R. L. (1997). Selected averaging of rapidly presented trials using fMRI. *Human Brain Mapping*, *5*, 1–12.
- Earley, J. (1970). An efficient context-free parsing algorithm. *Communications of the Association for Computing Machinery*, *13*, 94–102.
- Eddy, W. F., Fitzgerald, M., Genovese, C. R., Mockus, A., & Noll, D. C. (1996). Functional imaging analysis—computational olio. In A. Prat (Ed.), *Proceedings in computational statistics* (pp. 39–49). Heidelberg, Germany: Physica-Verlag.
- Ferreira, F., & Clifton, C. (1986). The independence of syntactic processing. *Journal of Memory and Language*, *25*, 348–368.
- Frazier, L. (1978). *On comprehending sentences: Syntactic parsing strategies*. Unpublished doctoral dissertation, University of Connecticut. Bloomington: Indiana University Linguistics Club, 1978.
- Frazier, L., & Clifton, C., Jr. (1996). *Construal*. Cambridge, MA: MIT Press.
- Frazier, L., & Rayner, K. (1982). Making and correcting errors during sentence comprehension: Eye movements in the analysis of structurally ambiguous sentences. *Cognitive Psychology*, *14*, 178–210.
- Gibson, E. (1998). Linguistic complexity: Locality of syntactic dependencies. *Cognition*, *68*, 1–76.
- Gibson, E., & Pearlmuter, N. J. (1998). Constraints on sentence comprehension. *Trends in Cognitive Science*, *2*, 262–268.
- Gibson, E., & Pearlmuter, N. J. (2000). Distinguishing serial and parallel parsing. *Journal of Psycholinguistic Research*, *29*, 231–240.
- Jurafsky, D. (1996). A probabilistic model of lexical and syntactic access and disambiguation. *Cognitive Science*, *2*, 137–194.
- Just, M. A., Carpenter, P. A., Keller, T. A., Eddy, W. F., & Thulborn, K. R. (1996, October 4). Brain activation modulated by sentence comprehension. *Science*, *274*, 114–116.
- Keller, T. A., Carpenter, P. A., & Just, M. A. (2001). The neural bases of sentence comprehension: fMRI examination of syntactic and lexical processing. *Cerebral Cortex*, *11*, 223–237.
- Kwong, K. K., Belliveau, J. W., Chester, D. A., Goldberg, I. E., Weisskoff, R. M., Poncelet, B. P., Kennedy, D. N., Hoppel, B. E., Cohen, M. S., & Turner, R. (1992). Dynamic magnetic resonance imaging of human brain activity during primary sensory stimulation. *Proceedings of the National Academy of Sciences of the United States of America*, *89*, 5675–5679.
- Lewis, R. L. (2000). Falsifying serial and parallel parsing models: Empirical conundrums and an overlooked paradigm. *Journal of Psycholinguistic Research*, *29*, 241–248.
- MacDonald, M. C., Just, M. A., & Carpenter, P. A. (1992). Working memory constraints on the processing of syntactic ambiguity. *Cognitive Psychology*, *24*, 56–98.
- Ogawa, S., Lee, T. M., Kay, A. R., & Tank, D. W. (1990). Brain magnetic resonance imaging with contrast dependent on blood oxygenation. *Proceedings of the National Academy of Sciences of the United States of America*, *87*, 9868–9872.
- Pearlmuter, N. J., & Mendelsohn, A. (1999). *Serial versus parallel sentence comprehension*. Manuscript submitted for publication.
- Rademacher, J., Galaburda, A. M., Kennedy, D. N., Filipek, P. A., & Caviness, V. S., Jr. (1992). Human cerebral cortex: Localization, parcellation, and morphometry with magnetic resonance imaging. *Journal of Cognitive Neuroscience*, *4*, 352–374.
- Rayner, K., Carlson, M., & Frazier, L. (1983). The interaction of syntax and semantics during sentence processing: Eye movements in the analysis of semantically biased sentences. *Journal of Verbal Learning and Verbal Behavior*, *22*, 358–374.
- Spivey, M. J., & Tanenhaus, M. K. (1998). Syntactic ambiguity resolution in discourse: Modeling the effects of referential context and lexical frequency. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *24*, 1521–1543.
- Stevenson, S. (1994). Competition and recency in a hybrid model of

- syntactic disambiguation. *Journal of Psycholinguistic Research*, 23, 295–322.
- Stromswold, K., Caplan, D., Alpert N., & Rauch, S. (1996). Localization of syntactic comprehension by positron emission tomography. *Brain and Language*, 52, 452–473.
- Taraban, R., & McClelland, J. L. (1988). Constituent attachment and thematic role assignment in sentence processing: Influences of content-based expectations. *Journal of Memory and Language*, 27, 597–632.
- Traxler, M. J., Pickering, M. J., & Clifton, C., Jr. (1998). Adjunct attachment is not a form of lexical ambiguity resolution. *Journal of Memory and Language*, 29, 558–592.

Appendix

Experiments 1 and 2 Stimuli

PP Attached to VP (Preferred)

The overworked scientist only read news reports on Sundays because he had so little time.
 The executive only called people on the intercom because he was paranoid.
 The kids played all the albums on the stereo before they went to bed.
 We watched the magic acts with amazement and clapped loudly when the act was over.
 The rescue crew reached the victim in minutes but had difficulty pulling him from the water.
 The police protected the witness from danger because he alone could identify the murderer.
 The baby disgusted the woman with his dirty diapers because she never had children herself.
 The landlord painted all the walls with enamel though it didn't help the appearance of the place.
 That kid hit the girl with a bat before he got off the subway and ran up the stairs.
 The salesman annoyed the woman with a sales pitch because she couldn't stand his loud voice.
 The spy saw the cop with binoculars but the cop didn't see him.
 The little girl tried to cut the apple with plastic knives though she wasn't very successful.
 Scott played the compact discs with Jim's CD player to see how bad they sounded.
 Rob bought a new television with his credit card and took advantage of the free delivery.
 Alice nursed the sick cat with warm milk before trying to find a new owner.
 The bitter wife killed her husband with a machete after she caught him cheating on her.
 Sandy named her first son from her list of names as soon as the baby was born.
 Emily hung the informative article on the board after leaving her history class.
 Anne arranged the colorful pamphlets on her floor before she left for the protest.
 The disgusted student threw the book on the ground but picked it up moments later.

PP Attached to NP (Unpreferred)

The overworked scientist only read news reports on tornadoes because he had so little time.
 The executive only called people on the payroll because he was paranoid.
 The kids played all the albums on the shelf before they went to bed.
 We watched the magic acts with rabbits and clapped loudly when the act was over.
 The rescue crew reached the victim in the suit but had difficulty pulling him from the water.
 The police protected the witness from New York because he alone could identify the murderer.
 The baby disgusted the woman with the tailored suit because she never had children herself.
 The landlord painted all the walls with cracks though it didn't help the appearance of the place.
 That kid hit the girl with the long hair before he got off the subway and ran up the stairs.
 The salesman annoyed the woman with the red blouse because she couldn't stand his loud voice.
 The spy saw the cop with a revolver but the cop didn't see him.
 The little girl tried to cut the apple with plastic coating though she wasn't very successful.
 Scott played the compact discs with deep scratches to see how bad they sounded.
 Rob bought a new television with a scratch on it and took advantage of the free delivery.
 Alice nursed the sick cat with a broken leg before trying to find a new owner.
 The bitter wife killed her husband with the limp after she caught him cheating on her.
 Sandy named her first son from her second husband as soon as the boy was born.
 Emily hung the informative article on the war after leaving her history class.
 Anne arranged the colorful pamphlets on the disease before she left for the protest.
 The disgusted student threw the book on the battle but picked it up moments later.

(Appendix continues)

Unambiguous—MV (Preferred)

The experienced soldiers spoke about the dangers before the midnight raid.
 The cotton farmers spoke about bad floods just before harvest time.
 Several angry workers spoke about low wages during the holiday season.
 The frightened kid went through the crowd to the front row.
 A small dog went through the fence into the chicken coop.
 An impatient shopper went through the doors to the sales racks.
 The evil genie ate the golden figs in the ancient temple.
 The kitchen staff ate in the cafeteria after the executives finished.
 The sunburned boys ate the hot dogs at the football stadium.
 The silly boys giggled during the play until the teacher arrived.
 The thoughtless secretaries giggled on the balcony when the parade passed.
 The teenage girls giggled in the hallway while the principal watched.
 The sick child cried early every morning in the hospital room.
 The calico cat cried in the alley after drinking the milk.
 The Indian children cried in the stream after their mothers left.
 The young children sang in the hallway while the adults argued.
 The brown sparrow sang on a branch high above the cat.
 The convicted criminal sang in the cell during the parole hearing.
 A yellow frisbee fell from the roof onto the narrow driveway.
 The large package fell from the plane into the dark jungle.
 Many small stones fell from the cliff during the fierce storm.
 The older kids learned all the dances for the spring recital.
 The young technician learned the computer program from the thick manual.
 The six volunteers learned the complicated procedure without very much trouble.
 The opera star sang in the auditorium while the audience sat quietly.
 The upset infant cried throughout the night in his crib.
 The overweight man ate the entire casserole at the family reunion.
 The injured player went through the locker room to the sauna room.
 The paper money went into the cash register from the annoyed clerk's hand.
 The movie actor sang in the subway station as the train came to a stop.
 The old man ate the special diet at the retirement home's cafeteria.
 The disorderly children giggled after class while the teacher yelled.
 The visiting campers spoke about the bears while they discarded their trash carefully.
 The hard ball fell from the boy's hand onto the baseball field.
 The amateur photographer learned the proper use of his new camera.
 The angry customer spoke about the delay before he was seated.
 The small boy went through the doorway to the playroom.
 The small kitten cried by his mother before getting picked up.
 The head coach ran along the sideline as the game dragged on.
 The horse ran past the barn after while his trainer watched.

Unambiguous—Relative Clause (Unpreferred)

The experienced soldiers who were told about the dangers conducted the midnight raid.
 The cotton farmers who were told about the bad floods had no other crops.
 Several angry workers who were told about low wages decided to file complaints.
 The frightened kid who was shoved through the crowd got separated from Jane.
 A small dog who was shoved through the fence hurt his hind leg.
 An impatient shopper who was shoved through the doors complained to the manager.
 The evil genie who was fed the golden figs went into a trance.
 The kitchen staff who were fed in the cafeteria soon got very sleepy.
 The sunburned boys who were fed the hot dogs got a stomach ache.
 The silly boys who were reprimanded during the play quickly left the auditorium.
 The thoughtless secretaries who were reprimanded on the balcony returned to their desks.
 The teenage girls who were reprimanded in the hallway answered the principal rudely.
 The sick child who was bathed early every morning wanted her rubber duck.
 The calico cat who was bathed in the alley ran into the street.
 The Indian children who were bathed in the stream splashed and shouted loudly.
 The young children who were seen in the hallway were following the adults.
 The brown sparrow who was seen on a branch pecked at an insect.
 The convicted criminal who was seen in the cell plotted a daring escape.
 A yellow frisbee that was thrown from the roof landed in the ditch.

The large package that was thrown from the plane hit several tall trees.
 Many small stones that were thrown from the cliff damaged passing cars below.
 The older kids who were shown all the dances were in the recital.
 The young technician who was shown the computer program caught on right away.
 The six volunteers who were shown the complicated procedure became very good students.
 The opera star who was seen in the auditorium sang while the audience sat quietly.
 The upset infant who was seen throughout the night cried in his crib.
 The overweight man who was fed the entire casserole hosted the family reunion.
 The injured player who was shoved through the locker room took refuge in the sauna room.
 The paper money that was shoved into the cash register became a wrinkled mess.
 The movie actor who was seen in the subway station walked onto the train.
 The old man who was fed the special diet left the hospital yesterday.
 The disorderly children who were reprimanded after class went home crying.
 The visiting campers who were warned about the bears discarded their trash carefully.
 The hard ball that was thrown by the boy broke a window.
 The amateur photographer who was shown the new camera bought a new case.
 The angry customer who was told about the delay punched the teller in the face.
 The small boy who was shoved through the doorway fell onto the floor.
 The small kitten who was bathed by his mother jumped after the ball.
 The head coach who was seen from the sideline shouted at the referee.
 The horse who was pulled past the barn escaped from his trainer.

Ambiguous—MV (Preferred)

The experienced soldiers warned about the dangers before the midnight raid.
 The cotton farmers warned about bad floods just before harvest time.
 Several angry workers warned about low wages during the holiday season.
 The frightened kid pushed through the crowd to the front row.
 A small dog pushed through the fence into the chicken coop.
 An impatient shopper pushed through the doors to the sales racks.
 The evil genie served the golden figs in the ancient temple.
 The kitchen staff served in the cafeteria after the executives finished.
 The sunburned boys served the hot dogs at the football stadium.
 The silly boys called during the play until the teacher arrived.
 The thoughtless secretaries called on the balcony when the parade passed.
 The teenage girls called in the hallway while the principal watched.
 The sick child washed early every morning in the hospital room.
 The calico cat washed in the alley after drinking the milk.
 The Indian children washed in the stream after their mothers left.
 The young children watched in the hallway while the adults argued.
 The brown sparrow watched on a branch high above the cat.
 The convicted criminal watched in the cell during the parole hearing.
 A yellow frisbee dropped from the roof onto the narrow driveway.
 The large package dropped from the plane into the dark jungle.
 Many small stones dropped from the cliff during the fierce storm.
 The older kids taught all the dances for the spring recital.
 The young technician taught the computer program from the thick manual.
 The six volunteers taught the complicated procedure without very much trouble.
 The opera star watched in the auditorium while the audience sat quietly.
 The upset infant watched throughout the night in his crib.
 The overweight man served the entire casserole at the family reunion.
 The injured player pushed through the locker room to the sauna room.
 The paper money dropped into the cash register from the annoyed clerk's hand.
 The movie actor watched in the subway station as the train came to a stop.
 The old man served the special diet at the retirement home's cafeteria.
 The disorderly children called after class while the teacher yelled.
 The visiting campers warned about the bears while they discarded their trash carefully.
 The hard ball dropped from the boy's hand onto the baseball field.
 The amateur photographer taught the proper use of his new camera.
 The angry customer warned about the delay before he was seated.

(Appendix continues)

The small boy pushed through the doorway to the playroom.
 The small kitten washed by his mother before getting picked up.
 The head coach watched from the sideline as the game dragged on.
 The horse raced past the barn while his trainer watched.

Ambiguous—RR (Unpreferred)

The experienced soldiers warned about the dangers conducted the midnight raid.
 The cotton farmers warned about the bad floods had no other crops.
 Several angry workers warned about low wages decided to file complaints.
 The frightened kid pushed through the crowd got separated from Jane.
 A small dog pushed through the fence hurt his hind leg.
 An impatient shopper pushed through the doors complained to the manager.
 The evil genie served the golden figs went into a trance.
 The kitchen staff served in the cafeteria soon got very sleepy.
 The sunburned boys served the hot dogs got a stomach ache.
 The silly boys called during the play quickly left the auditorium.
 The thoughtless secretaries called on the balcony returned to their desks.
 The teenage girls called in the hallway answered the principal rudely.
 The sick child washed early every morning wanted her rubber duck.
 The calico cat washed in the alley ran into the street.
 The Indian children washed in the stream splashed and shouted loudly.
 The young children watched in the hallway were following the adults.
 The brown sparrow watched on a branch pecked at an insect.
 The convicted criminal watched in the cell plotted a daring escape.
 A yellow frisbee dropped from the roof landed in the ditch.
 The large package dropped from the plane hit several tall trees.
 Many small stones dropped from the cliff damaged passing cars below.
 The older kids taught all the dances were in the recital.
 The young technician taught the computer program caught on right away.
 The six volunteers taught the complicated procedure became very good students.
 The opera star watched in the auditorium sang while the audience sat quietly.
 The upset infant watched throughout the night cried in his crib.
 The overweight man served the entire casserole hosted the family reunion.
 The injured player pushed through the locker room took refuge in the sauna room.
 The paper money dropped into the cash register became a wrinkled mess.
 The movie actor watched in the subway station walked onto the train.
 The old man served the special diet left the hospital yesterday.
 The disorderly children called after class went home crying.
 The visiting campers warned about the bears discarded their trash carefully.
 The hard ball dropped by the boy broke a window.
 The amateur photographer taught the new camera bought a new case.
 The angry customer warned about the delay punched the teller in the face.
 The small boy pushed through the doorway fell onto the floor.
 The small kitten washed by his mother jumped after the ball.
 The head coach watched from the sideline shouted at the referee.
 The horse raced past the barn escaped from his trainer.

Received April 25, 2002
 Revision received April 24, 2003
 Accepted April 28, 2003 ■