Research Report

# Neural activity of inferences during story comprehension

*Sandra Virtue[a],\*, Jason Haberman[b], Zoe Clancy[c], Todd Parrish[c], Mark Jung Beeman[c]*

[a]*Department of Psychology, 2219 North Kenmore Ave., DePaul University, Chicago, IL 60614, USA*
[b]*University of California, Davis, CA 95616, USA*
[c]*Northwestern University, Chicago, IL 60611, USA*

## ARTICLE INFO

## ABSTRACT

In this event-related functional magnetic resonance imaging (fMRI) study, participants listened to and comprehended short stories implying or explicitly stating inference events. The aim of this study was to examine the neural mechanisms that underlie inference generation, a process essential to successful comprehension. We observed distinct patterns of increased fMRI signal for implied over explicit events at two critical points during the stories: (1) within the right superior temporal gyrus when a verb in the text implied the inference; and (2) within the left superior temporal gyrus at the coherence break or when participants need to generate an inference to understand the story. To find the most compelling evidence of neural activity during inference generation, we examined fMRI signal at these two critical points separately for people with high working memory capacity (i.e., those individuals who are most likely to draw inferences during text comprehension). Interestingly, high working memory individuals showed greater fMRI signal for implied than explicit events in the left inferior frontal gyrus at the coherence break compared to low working memory individuals. This present study provides evidence that areas within the superior temporal gyrus and inferior frontal gyrus are heavily recruited when individuals generate inferences, even during ongoing comprehension that demands many cognitive processes. In addition, the data suggest that the right hemisphere superior temporal gyrus is particularly involved during early inferential processing, whereas the left hemisphere superior temporal gyrus is particularly involved during later inferential processing in story comprehension.

## 1. Introduction

When readers comprehend text, they carry out a variety of cognitive processes. One process essential for successful text comprehension is the generation of causal inferences. Specifically, readers must make connections between causes and their consequences when events are inferred in a text. Consider the passage: "John was going to a wedding tonight. He had been sitting around the house in his jeans, so he went to his bedroom to find some clothes. Soon he came out wearing his tuxedo." After reading the last sentence in the passage, readers need to generate the inference that John changed his clothes to understand why he was wearing jeans but is now wearing a tuxedo. If readers do not generate this inference, they reach a gap in understanding (i.e., a coherence break) and would likely have difficulty comprehending further information in the text.

Drawing inferences is critical to understanding discourse, and a large literature discusses the characteristics of the text (Graesser et al., 1994; Kintsch, 1988; Kintsch and van Dijk, 1978; McKoon and Ratcliff, 1992; van den Broek, 1990) and the characteristics of the individual (Estevez and Calvo, 2000; Just and Carpenter, 1992; Linderholm, 2002; Linderholm and van den Broek, 2002; Singer et al., 1992) that affect what inferences are drawn and when. However, relatively little attention has been given to the neural activity that occurs during inference generation and how this activity supports specific cognitive processes that likely enable people to draw inferences. In the current study, we investigate the neural activity that occurs when people draw inferences and we propose possible ideas for specific cognitive processes that may be essential during inference generation and throughout text comprehension in general.

The focus of the current study is on text comprehension and on connective inferences that facilitate comprehension (i.e., causal inferences). Other types of inferences occur outside the domain of text comprehension, such as inferences made during logical reasoning (Goel and Dolan, 2004; Noveck et al., 2004). While both types of inferences may share some component processes, they are also likely to differ. For instance, inferences during text comprehension occur rapidly and easily—indeed they can be difficult to inhibit—whereas logical inferences may occur deliberately during effortful problem solving.

There are several types of inferences that individuals can draw during discourse comprehension.[1] In this study, we investigate a particular type of causal inference or a connection between an antecedent and a consequence, known to be important to the successful comprehension of text. The text comprehension literature distinguishes between two major types of causal inferences: predictive (i.e., forward) inferences and bridging (i.e., backward) inferences. When readers generate predictive inferences, they generate expectations about what events will happen next in a text. When readers generate bridging inferences, they link current text information with previous text information or information from their background knowledge to fill in a causal gap. By using the divided visual field paradigm—when subjects respond to inference-related words presented to one visual field to index semantic processing in the contralateral hemisphere—researchers have shown that each hemisphere differently processes information that supports predictive and bridging inferences. Specifically, the right hemisphere is more sensitive than the left hemisphere to concepts related to predictive inferences (Beeman et al., 2000), to anaphoric inferences (i.e., a specific type of bridging inference) with multiple antecedents (Virtue and van den Broek, 2005), and to bridging and predictive inferences preceded by weakly constrained text (Virtue et al., in press). In

contrast, the left hemisphere is more sensitive than the right hemisphere to concepts related to bridging inferences (Beeman et al., 2000), to anaphoric inferences with one antecedent (Virtue and van den Broek, 2005), and to predictive and bridging inferences preceded by strongly constrained text (Virtue et al., in press). Thus, the two hemispheres seem to process semantic information differently (Chiarello et al., 1990) as comprehenders generate inferences.

Recent functional magnetic resonance imaging (fMRI) studies have extended these findings by exploring the neural substrates related to inference generation (Mason and Just, 2004). We propose that two specific areas will be heavily recruited during inference generation: the superior temporal gyrus (STG) and the inferior frontal gyrus (IFG). Areas within the STG—extending into the superior temporal sulcus (STS) and often into the middle temporal gyrus (MTG)—may be important, among other things, during the semantic integration of inferential information (Jung-Beeman, 2005). For example, activity in the STG increases as individuals read sentence pairs that are causally linked (Mason and Just, 2004), read texts without titles (St. George et al., 1999), detect inconsistencies in story information (Ferstl et al., 2005), solve insight problems (Jung-Beeman et al., 2004), and carry out syllogistic reasoning (Goel and Dolan, 2001). To complete these tasks, participants need to connect information that is not explicitly stated. For example, when participants are given texts that are difficult to comprehend without a title, they must link several pieces of information in a text with information from their background knowledge. The ability to connect this information that is not explicitly stated may occur in a similar manner as when readers connect information during inference generation. Areas within the IFG may be important, among other things, during the semantic selection of inferential information. For example, activity in the IFG increases as participants choose between several alternative answers (Barch et al., 2000; Thompson-Schill et al., 1997). In addition, neural activity in IFG increases when individuals generate a single use or action associated with a presented noun (Posner et al., 1988). Usually, the left IFG is predominantly active during the generation of nouns, however, the right IFG is also active when people generate an unusual use or action for a noun (Seger et al., 2000). To complete these tasks, participants need to select semantic information among competing alternatives. A similar process of selecting the appropriate information could also occur when comprehenders need to select an inference during inference generation. Thus, areas within the STG and IFG are likely to play important, and possibly different, roles during the processing of inferences.

There is some evidence that areas within the STG and IFG are involved in drawing inferences. For example, fMRI signal increases in the right STG as textual constraints demand greater inference generation (Mason and Just, 2004) and signal in the right IFG increases as the text encourages readers to connect information across sentences (Robertson et al., 2000). However, several essential questions remain unanswered about how exactly the STG and IFG may be involved during inference generation. First, is the neural activity found within the STG and IFG in previous studies apparent throughout text comprehension or are these brain areas highly involved at

---

[1] One type of inference often cited in the theory of mind (TOM) literature links characters' mental states to explain and predict behavior during the comprehension of stories. For example, studies have shown that specific neural activity occurs when individuals infer the intention of others' mental states (Fletcher et al., 1995; Rilling et al., 2004). Although these inferences about characters' states of mind can be important, the current study focuses on a different set of inferences about physical events that do not necessitate consideration of the characters' state of mind.

different times during the generation of inferences? Based on behavioral findings, we know that inferential processing can vary at different points throughout a text. For example, when people comprehend events that are implied in a text, they show greater priming for inference-related words than unrelated words in the right hemisphere. However, when people comprehend text when a coherence break occurs, they show greater priming for inference-related words than unrelated words in the left hemisphere (Beeman et al., 2000). Although the hemispheric processing of text varies throughout comprehension, previous neuroimaging research has measured neural activity over entire blocks of text (Robertson et al., 2000) or during pairs of sentences (Ferstl and von Cramon, 2001; Mason and Just, 2004). Thus, it is difficult to know if these results generalize to neural activity found at different times throughout the entire process of inference generation.

With recent advances in neuroimaging techniques, information about the time course of neural activity during inference generation can be obtained. Specifically, by using an event-related fMRI design, we will investigate the neural activity that occurs at two critical time points: when the verb in a text implies an inference in a text (i.e., verb point) and when comprehenders must generate an inference to establish coherence (i.e., a coherence break). Because we are able to examine neural activity at critical time points throughout text comprehension, we also will be able to present more naturalistic texts (i.e., stories that consist of multiple sentences) in this study.

Second, will increases in neural activity within the STG and IFG during inference generation be apparent even in the presence of other ongoing text comprehension processes? While comprehending a passage, people have many concepts available to them that are related to many different inferences. Thus, it is quite possible that fMRI signal related to specific inferences could be drowned out by a cacophony of other processes that occur during text comprehension. For example, the specific inferences that are generated and when they are generated depend on factors related to the individual comprehender (e.g., goals for comprehension (Linderholm and van den Broek, 2002), working memory capacity (Estevez and Calvo, 2000; Linderholm, 2002), background knowledge (Graesser et al., 1994; Singer et al., 1992), and the text (e.g., the number of causal connections, the more strongly the text points to a specific event, and the presence of coherence breaks or contradictions (van den Broek, 1990; Virtue et al., in press)). Behavioral (Estevez and Calvo, 2000; Linderholm, 2002) and electrophysiological (St. George et al., 1997) evidence show that high working memory comprehenders are more likely than low working memory comprehenders to generate inferences. Thus, to find the most compelling evidence of inference generation during text comprehension, we will examine neural activity within the STG and the IFG in comprehenders who are most likely to generate an inference, specifically, participants who have high working memory capacities.

In the current study, we will investigate the neural activity as comprehenders process inferences at two critical time points during story comprehension: when the verb in a text implies an inference and at the coherence break. To minimize the effects of other factors in the text at these two time points, the stories we contrasted vary in only a few words which were necessary to either imply or explicitly state the inference. By using an event-related fMRI design, we will measure neural activity at these two critical time points as participants comprehend naturalistic stories that contain multiple sentences. In addition, we will examine fMRI signal in high working memory capacity individuals, who are most likely to show evidence of inference generation.

## 2. Results

### 2.1. Participants and comprehension task performance

Data from seventeen Northwestern University undergraduates (12 females and 5 males, all right-handed native English speakers) were included in the analyses. To encourage comprehension, we asked participants true/false questions about the stories after each imaging run. Two additional participants who answered fewer than 63% of the comprehension questions correctly were not included in the analyses. The remaining 17 participants answered on average 87% of the comprehension questions correctly (SD = 7.59%, range 75–100%).

### 2.2. MR results

At each of the two critical time points, we contrasted fMRI signal from 40 implied events (e.g., to find some clothes) to fMRI signal from 40 explicitly stated events (e.g., to change his clothes) that occurred while participants comprehended 20 short stories. At the verb point, fMRI signal increased more following implied events than following explicit events in the right superior temporal gyrus (Table 1; Fig. 1). This cluster in the right STG has two separate foci (one at 44, –20, 10; the second at 58, –16, 6). Later in each passage at the coherence break, fMRI signal increased more following implied events than following explicit events in one cluster above threshold in the left STG (centered at –54, –46, 6 with a volume of 3750 mm$^3$) and a second cluster just below the threshold (centered at –48, 3, –19 with a volume of 460 mm$^3$) (Table 1; Fig. 2). Other brain areas are highly active throughout story comprehension; some of these could contribute to drawing inferences, without showing increased fMRI signal above this high baseline. However, because no other clusters of voxels reliably differed across implied and explicit events at either time point, we have no evidence of inference-specific activity in other brain areas.

Previous behavioral studies indicate that high working memory individuals are more likely to generate inferences than low working memory individuals (St. George et al., 1997). Thus, we expected to see the best evidence of inference-related processing in the superior temporal gyrus and in the inferior frontal gyrus in individuals with high working memory capacity. Indeed, in the current study, individuals' working memory scores correlated with the strength of fMRI signal increases in the left STG for implied relative to explicit events at the coherence break (r = 0.38, p < 0.05; one tailed). Therefore, we separately analyzed the fMRI data of the high working memory group to identify clusters showing stronger

**Table 1 – All areas larger than 400 mm³ at *p* < .005 showing stronger fMRI signal for implied events compared to explicitly stated events following the verb point and the coherence break**

| Critical event | Gyrus/ Structure | Brodmann area | Volume | Center Coordinates | | | Mean percent signal | Maximum percent signal | Mean *t* | Maximum *t* | Maximum *t* coordinates | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | *x* | *y* | *z* | | | | | *x* | *y* | *z* |
| **Verb point** | | | | | | | | | | | | | |
| Implied > Explicit | Right superior temporal gyrus | 41, 22 | 1391 | 49 | −19 | 9 | 0.16 | 0.10 | 3.7 | 4.8 | 46 | −18 | 10 |
| **Coherence break** | | | | | | | | | | | | | |
| Implied > Explicit | Left middle and superior temporal gyrus | 21, 22 | 3750 | −54 | −46 | 6 | 0.21 | 0.33 | 3.8 | 5.3 | −56 | −53 | 5 |
| | Left middle temporal gyrus | 21 | 460 | −48 | 3 | −19 | 0.16 | 0.20 | 3.6 | 4.2 | −46 | 4 | −24 |

All clusters are shown in descending order of average percent signal change. Location of cluster centers and peak *t* values are shown in Talairach coordinates.
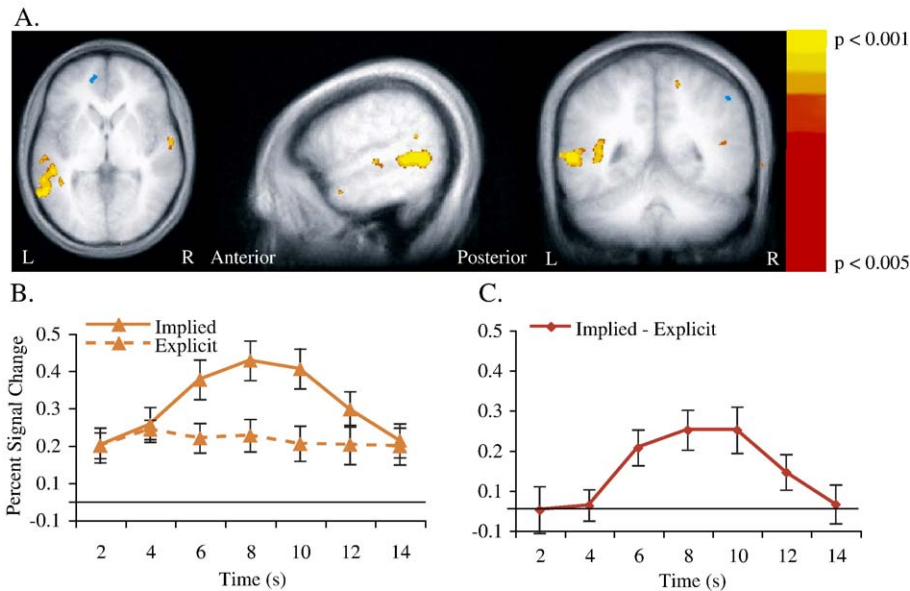
fMRI signal for implied than for explicit events in the texts. We also directly contrasted fMRI signal change for implied relative to explicit events in the high working memory group to that of the low working memory group. All clusters for these comparisons at both the verb point and the coherence break over 400 mm³ at *p* < 0.005 are reported in Table 2 (and for the sake of completion, the implied versus explicit comparison within the low working memory group is also reported).

At the verb point, the high working memory group showed stronger signal for implied than explicit events in the posterior cingulate, but not in any of the predicted areas (i.e., the STG and the IFG). Similarly, the contrast between high and low working memory groups at the verb point did not reveal evidence of greater fMRI signal for implied than explicit events in these areas. However, at the coherence break, the three largest clusters showing stronger signal for implied than for explicit events in the high working memory subjects were precisely the areas predicted to be important for inference generation: the right STG, the left STG, and the left IFG. Moreover, in directly contrasting the two groups at the coherence break, the largest areas showing stronger signal change for implied over explicit events for the high working memory group than for the low working memory group were in left IFG/anterior insula and the left STG (Table 2; Fig. 3). For
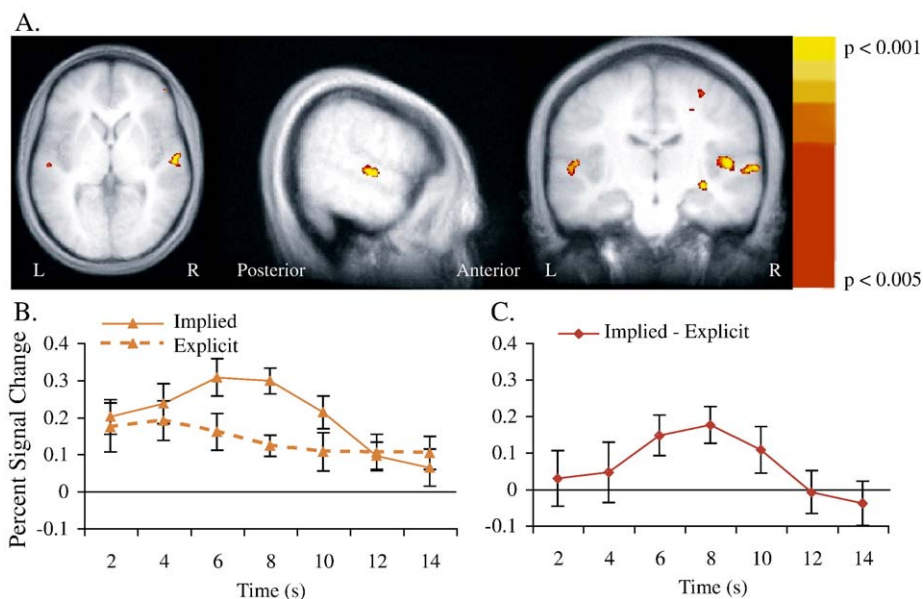


**Fig. 1 – fMRI signal in the right superior temporal gyrus following the verb point. (A) Voxels showing stronger fMRI signal for implied events compared to explicitly stated events (*t* = 3.25, *p* < .005) following the verb point overlaid on the averaged normalized structural image of all subjects. The cluster is in right superior temporal gyrus with a cluster volume of 1391 mm³; with the peak *t* value of 4.8 occurring at 46, –18, 10. (B) Group average signal change in percent signal for the seven observed time points following implied events (orange solid line) and explicitly stated events (orange dashed line) across the entire right superior temporal gyrus. Error bars represent the standard error of the mean signal change at each time point. (C) Implied event signal minus explicit event signal across the active region in right superior temporal gyrus. Error bars show the standard error of the mean of the difference at each time point.**

Fig. 2 – fMRI signal in the left superior temporal gyrus following the coherence break. (A) Voxels showing stronger fMRI signal for implied events compared to explicitly stated events ($t$ = 3.25, $p$ < .005) following the coherence break overlaid on the averaged normalized structural image of all subjects. The cluster is in left posterior superior temporal gyrus with a cluster volume of 3750 mm$^3$; with the peak $t$ value of 5.3 occurring at –56, –53, 5. (B) Group average signal change in percent signal for the seven observed time points following the coherence break for implied events (orange solid line) and explicit events (orange dashed line) across the entire left superior temporal gyrus. Error bars show the standard error of the mean of the difference at each time point. (C) Implied event signal minus explicit event signal following the coherence break across the active region in left superior temporal gyrus. Error bars show the standard error of the mean of the difference at each time point.

the high and low working memory comparison at the coherence break, there was also a reliable cluster in left ventral anterior cingulate; the next biggest cluster was in the right STG (centered at 51, 21, 5), but it was smaller than our size threshold with a volume of 344 mm$^3$.

## 3.    Discussion

We observed greater fMRI signal, indicating increased neural activity in several brain areas, when people comprehend stories that imply causal events than when people comprehend stories that explicitly state causal events. This increase in fMRI signal was observed with naturalistic texts, using fMRI as a covert measure. That is, participants merely comprehended stories without performing any concomitant secondary task during comprehension. These processing differences were observed over and above the demanding processing required for ongoing comprehension of the texts, despite the fact that the texts differed by only a few words and were identical at the coherence break. In addition, we found evidence that this neural activity varied at specific times during inference generation and for comprehenders with different levels of working memory capacity.

These results lead to interesting speculations about what cognitive processes may be involved during the generation of inferences. The increases in fMRI signal for implied events compared to explicit events suggest that unique cognitive processes may be utilized when comprehenders

generate inferences. As predicted, the current study found greater signal for implied than explicitly stated events in the STG.[2] This is consistent with a proposal that inference generation requires semantic activation, integration, selection, and/or incorporation of the inferred concept, with distinct brain areas involved in each of these processes (Beeman, 1993; Beeman et al., 2000; Jung-Beeman, 2005; Jung-Beeman et al., 2004). For example, comprehending passages that imply events may make semantic integration more difficult than comprehending passages that explicitly state an event. In fact, increased semantic integration has been associated with increased activity in the STG (as in our study, sometimes extending into STS and MTG) (Goel and Dolan, 2001; Jung-Beeman et al., 2004; Mason and Just, 2004; St. George et al., 1999).

This model of inferencing (Beeman, 1993; Beeman et al., 2000; Jung-Beeman, 2005; Jung-Beeman et al., 2004) further suggests that comprehenders must select the appropriate semantic concept to resolve a break in coherence (e.g., to select the concept changed). Increased semantic selection has been associated with increased activity in the inferior frontal gyrus (Barch et al., 2000; Frith et al., 1991; Seger

---

[2] Although parts of both superior temporal gyrus clusters extend close to auditory cortex, both implied and explicit events in the stories contained equivalent levels of auditory input. Moreover, while our stories were presented auditorily, similar regions have been identified in several recent studies with visually presented words in which the signal was interpreted as reflecting increased lexical–semantic integration (Jung-Beeman et al., 2004; Kircher et al., 2001).

**Table 2 – All areas larger than 400 mm³ at $p < 0.005$ showing fMRI signal for implied events compared to explicitly stated events in high working memory subjects, a comparison of high and low working memory subjects, and in the low working memory subjects following the verb point and the coherence break**

| Critical event | Gyrus/Structure | Brodmann area | Volume | Center coordinates | | | Mean percent signal | Maximum percent signal | Mean t | Maximum t | Maximum t coordinates | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | x | y | z | | | | | x | y | z |
| *High working memory subjects* | | | | | | | | | | | | | |
| Verb point | | | | | | | | | | | | | |
| Implied > Explicit | Right posterior cingulate | 23, 31 | 641 | 5 | −63 | 16 | 0.23 | 0.33 | 5.0 | 8.6 | 3 | −61 | 18 |
| Coherence break | | | | | | | | | | | | | |
| Implied > Explicit | Right superior temporal gyrus | 22 | 953 | 54 | −17 | 3 | 0.21 | 0.30 | 4.9 | 7.0 | 48 | −18 | 3 |
| Implied > Explicit | Left superior temporal gyrus | 22 | 781 | −55 | −27 | 3 | 0.32 | 0.40 | 4.6 | 6.8 | −56 | −21 | 5 |
| Implied > Explicit | Left inferior frontal gyrus | 22 | 406 | −36 | 27 | −4 | 0.23 | 0.29 | 5.4 | 9.9 | −36 | 29 | −4 |
| | | | | | | | | | | | | | |
| *High–Low working memory subjects* | | | | | | | | | | | | | |
| Verb point | | | | | | | | | | | | | |
| HWM > LWM | Left inferior parietal lobule | 40 | 750 | −59 | −27 | 34 | 0.37 | 0.52 | 3.8 | 4.8 | −59 | −28 | 35 |
| LWM > HWM | Right cingulate gyrus | 32 | 531 | 1 | 31 | 26 | 0.48 | 0.63 | 4.0 | −5.0 | 1 | 34 | 23 |
| Coherence break | | | | | | | | | | | | | |
| HWM > LWM | Left inferior frontal gyrus/Left insula | 47, 13 | 1031 | −40 | 20 | 0 | 0.42 | 0.57 | 3.8 | 5.0 | −46 | 14 | −4 |
| HWM > LWM | Left superior temporal gyrus | 21 | 812 | −58 | −29 | 0 | 0.41 | 0.55 | 3.7 | 4.6 | −59 | −31 | −2 |
| HWM > LWM | Left ventral anterior cingulate | 25 | 594 | −5 | 14 | −9 | 0.34 | 0.44 | 3.9 | 5.5 | −4 | 14 | −7 |
| | | | | | | | | | | | | | |
| *Low working memory subjects* | | | | | | | | | | | | | |
| Coherence break | | | | | | | | | | | | | |
| Implied > Explicit | Left superior temporal gyrus | 22 | 797 | −55 | −49 | 4 | 0.24 | 0.34 | 4.5 | 6.2 | −54 | −46 | 3 |
| Implied > Explicit | Left middle occipital gyrus | 37 | 656 | −42 | −69 | 4 | 0.17 | 0.23 | 4.6 | 6.5 | −36 | −74 | 0 |
| Explicit > Implied | Left insula | 13 | 562 | −36 | 16 | 4 | 0.20 | 0.26 | 4.7 | −6.9 | −36 | 16 | 5 |
| Implied > Explicit | Left fusiform | 20 | 531 | −38 | −41 | −12 | 0.15 | 0.20 | 4.9 | 9.2 | −39 | −44 | −10 |

All cluster sizes are shown in descending order of average percent signal change. Location of cluster centers and peak t values are shown in Talairach coordinates.

et al., 2000; Thompson-Schill et al., 1997). Consistent with this idea, the current study found greater IFG activity for implied relative to explicit events at the coherence break in high working memory participants than in low working memory participants. Thus, it is likely that in this study the high working memory comprehenders were selecting the inference when they reached a coherence break in the text more often than were the low working memory comprehenders.

It is also possible that the IFG activation found for high working memory comprehenders reflects a different, but equally important, cognitive process other than semantic selection. For example, researchers have proposed that IFG activity is a reflection of difficulty or effortful processing (Mason and Just, 2004). However, an a priori assumption might be that low working memory participants would have more difficulty than high working memory participants comprehending the implied events—yet they show less

signal change in the IFG as well as in the left STG. Alternatively, this IFG activity could reflect effort in cognitive mapping (Robertson et al., 2000) or additional semantic retrieval processes (Wagner et al., 2001) that are carried out to help comprehenders organize information to create a coherent representation of a story. Whatever the specific process is that is carried out in the IFG, it seems to be important for completing causal inferences and occurs more strongly in high than in low working memory comprehenders. Thus, event-related fMRI may be a useful tool to identify specific cognitive processes, and their underlying neural mechanisms, involved when people draw causal inferences.

With respect to hemisphere differences, previous studies demonstrated that, when an event is inferred in a text, inference concepts are more accessible in the right hemisphere than in the left hemisphere (Beeman et al., 2000). Therefore, it was not surprising that early in the inference
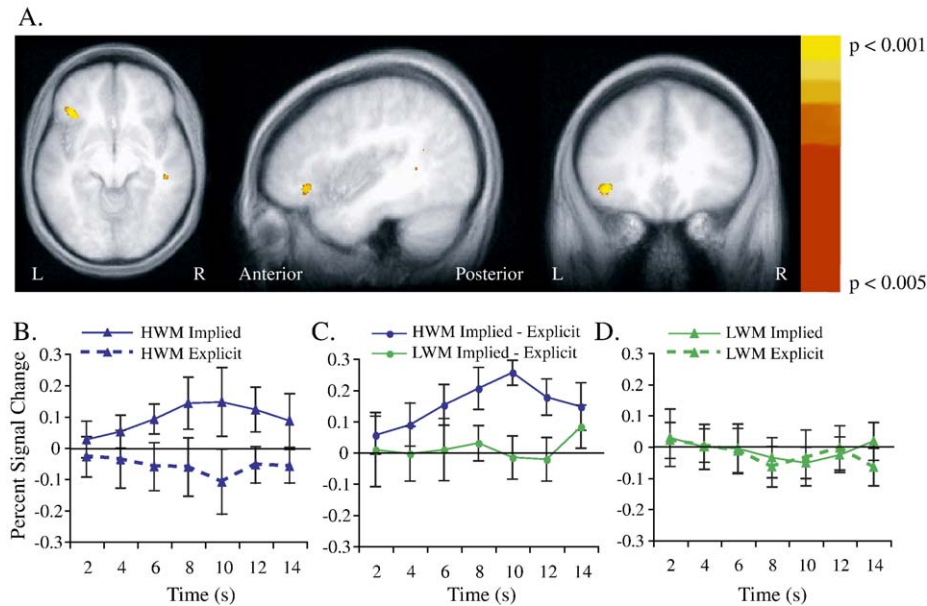
Fig. 3 – fMRI signal in left inferior frontal gyrus following the coherence break for high working memory (HWM) participants compared to low working memory (LWM) participants. (A) Voxels showing stronger fMRI signal for implied events compared to explicitly stated events for high working memory subjects compared to low working memory subjects following the coherence break overlaid on the averaged normalized structural image of 8 high working memory subjects and 9 low working memory subjects. Three clusters are over 500 mm³ at the standard threshold of $t = 3.28$, $p < .005$: one cluster is in the left inferior frontal gyrus/insula with a volume of 1031 mm³, the second cluster is in the left superior temporal gyrus with a volume of 812 mm³, and the third cluster is in the left anterior cingulate with a volume of 594 mm³. (B) Average signal change in percent signal for the seven observed time points following implied events (blue solid line) and explicitly stated events (blue dashed line) across the entire left inferior frontal gyrus/insula for high working memory subjects. Error bars show the standard error of the mean of the difference at each time point. (C) Implied event signal minus explicit event signal following the coherence break for high working memory subjects (blue solid line) compared to low working memory subjects (green solid line) across the active region in the left inferior frontal gyrus. Error bars show the standard error of the mean of the difference at each time point. (D) Average signal change in percent signal for the seven observed time points following implied events (green solid line) and explicitly stated events (green dashed line) across the entire left inferior frontal gyrus/insula for low working memory subjects. Error bars show the standard error of the mean of the difference at each time point.

generation process the right STG showed the strongest signal for implied events compared to explicit events. Signal in the right STG has been observed for a variety of linguistic tasks emphasizing semantic integration (Ferstl et al., 2005; Goel and Dolan, 2001; Jung-Beeman et al., 2004; Kircher et al., 2001; Mason and Just, 2004; St. George et al., 1999). The involvement of right hemisphere language areas, generally (Meyer et al., 2000), and right anterior temporal areas, specifically (Ferstl et al., 2005), has often been attributed to increased difficulty without specific formulation of what processes are involved. However, in the current study at the coherence break, implied events elicited increased activity in the right STG of the high working memory group, who would be expected to have less difficulty drawing inferences. At the verb point, activation for implied events was moderately correlated with increased performance on comprehension questions, though this was not statistically reliable ($r = 0.37$, $p < 0.1$; one tailed).

Although the different versions of the stories in the current study did not differ in content at the coherence break, they did differ in processing demands and fMRI signal in the STG. Consistent with prior behavioral work

(Beeman et al., 2000), this effect was predominantly in the left hemisphere. Thus, the verb that implied or explicitly stated inference events in previous sentences influenced how comprehenders processed later semantic information when generating inferences later at the coherence break. These results suggest that the right STG may initially be more sensitive than the left STG to causal connections in a text (Beeman, 1993; Beeman et al., 2000) and ultimately engage additional processes earlier than the left STG. The current results cannot conclusively determine whether the signal in the STG reflects increased effort to semantically integrate the story, successful integration, or some other cognitive process. In fact, it is possible that both semantic integration and effort to integrate information are occurring, especially in comprehenders who may have more cognitive resources available to them (e.g., high working memory individuals).

A few pieces of evidence suggest that the activity in the anterior STG of high working memory participants might reflect successful semantic integration. In the current study, two patterns of STG activity were apparent at the coherence break: high working memory participants showed activity in

mid-anterior STG and low working memory participants showed activity in posterior STG. Speculatively, this may reflect that high working memory participants have successfully generated the inference and in fact may be incorporating them into memory (e.g., integrating the specific information that John changed from jeans into a tuxedo). In contrast, low working memory participants may be still activating information to try to resolve information related to the inference at the coherence break. These group differences did not surpass strict criteria, so this interpretation is tentative, but it fits nicely with other results (Jung-Beeman et al., 2004; Kircher et al., 2001). Finally, prior behavioral studies (Beeman, 1993) suggest that specific inference-related concepts are available in the right hemisphere prior to the coherence break—at a time analogous to when the verb implied the inference in the current experiment when we observe greater neural activity for implied than explicit events in the right STG. In addition, inference-related concepts are more accessible in the left hemisphere at the coherence break—at the time when increased fMRI signal for implied events compared to explicit events occurred predominantly in the left STG. When considering all the evidence, we propose that neural activity in the right STG when the inference event was inferred likely reflects successful semantic integration. However, future studies are still needed to further delineate semantic integration effort from successful semantic integration, let alone incorporation of inferences into story representations.

Based on these results, two important issues must be considered. First, we propose that the brain areas involved in this study are not the only areas involved during inference generation. Rather, areas within the superior temporal gyrus and inferior frontal gyrus are likely part of a larger collection of areas utilized during the comprehension of text. These brain areas are particularly taxed when people draw inferences and are likely necessary for the generation of a coherent situation model of a text (Long and Baynes, 2002). Second, we propose that a network of complex processes are involved during the generation of inferences. The current study provides evidence of at least two possible cognitive processes (e.g., semantic integration and selection) that may be involved at two critical time points (e.g., the verb point and the coherence break). However, it is possible that other cognitive processes and neural substrates are required for the generation of other types of inferences (Ferstl and von Cramon, 2001), or even for the type studied here, but may have not been observed in our particular contrast. For example, it would be interesting to examine the neural activity of different types of inferences utilizing a lower baseline of neural activity. In addition, future investigations could examine the neural substrates involved during both causal inferences during text comprehension and logical inference generation during reasoning or problem solving (Goel and Dolan, 2004; Noveck, et al., 2004). The current study was aimed at identifying the neural substrates of causal inferences that are essential to the successful comprehension of text. However, it is likely that, during both causal and logical inference generation, similar cognitive and neural processes may occur. Thus, it would be fruitful to investigate the relation between these similarly named but independently studied processes.

## 4.    Conclusion

In conclusion, we observed distinct patterns of neural activity as people draw inferences while comprehending short stories. Importantly, we observed this with naturalistic stories using very subtle alterations of the text, with no secondary task, and in the presence of other ongoing processes during story comprehension. These results provide evidence that, as people generate inferences during comprehension, areas within the superior temporal gyrus and inferior frontal gyrus play important roles in creating a coherent representation of text. Specifically, neural activity within the right superior temporal gyrus occurred earlier during comprehension of inference events, whereas activity within the left superior temporal gyrus occurred at the coherence break. These fMRI results nicely map onto prior work showing that the right hemisphere is sensitive to appropriate inference concepts earlier than is the left hemisphere. In addition, this study found evidence of increased inferior frontal gyrus activity at the coherence break for individuals with high working memory capacity. Thus, this study provides support for the critical role of the superior temporal gyrus and the inferior frontal gyrus during the generation of inferences and, more generally, when comprehending text.

## 5.    Experimental procedures

### 5.1.    Participants

Seventeen participants (5 men and 12 women, age range 18–28) were native English speakers, right-handed (M = 0.83, range 0.4 to 1.0), as assessed by a five-item preference questionnaire (with a range of +1.0 to –1.0) (Bryden, 1982), and without significant history of neurological disorders. All had normal or corrected-to-normal visual acuity. The Northwestern University Institutional Review Board approved all experimental procedures, and written informed consent was obtained from each participant.

### 5.2.    Materials and procedures

During fMRI scanning, participants listened over headphones to 20 stories describing 80 critical inference events. The stories were played by an MRA, Inc. audio system. There was no overt task during story comprehension. Specifically, participants were instructed to listen carefully to each of the stories and to be prepared to answer true/false questions about them. The true/false questions did not refer to the implied or explicit events and were presented on the screen following each scanning run. Each story described four episodes in which there was a central inference event in each episode. In each of the four episodes, we created two versions: one in which the verb implied an inference in a text and one in which it was explicitly stated. To minimize any effects due to other factors in the text, the two versions

**Table 3 – Sample story when the inference event was implied or explicitly stated**

Nancy called John to say she'd pick him up early for her best friend's wedding. John had been sitting around the house in his jeans, so he went to his bedroom to find[1] some clothes/to change his clothes. Soon he came out wearing his tuxedo, [2] which had belonged to John's father, but was still fashionable and looked like new.
John was pleased that the tuxedo fit so well and was in such good condition. But then John noticed that it was all creased and wrinkled, so he got to work[1]/so he started ironing. After a few minutes, his shirt was completely smooth[2], so he got dressed again, and put on his shoes. John knew that Nancy would arrive any minute and wanted to have everything ready so they could leave as soon as she got there. John hurriedly pulled the blender out of the bag, got out the scissors and tape, and set them all down[1]/started wrapping. When John showed Nancy the box, she said it looked pretty in the flowered paper[2], so they carefully put the flowered box in the car, and then drove to the wedding. The wedding took place out in the country in an old church and the ceremony was beautiful. After the ceremony, John and Nancy stood outside, their hands full of rice, and waited[1] to see the newlyweds/threw it at the newlyweds. Soon the rice was in their hair and clothes[2] and John took pictures, while Nancy wished her friend a happy honeymoon.

Note that the verb point[1] (when the verb implied or explicitly stated the inference) and the coherence break[2] (when comprehenders must generate the inference for successful comprehension) are labeled for each of the four critical inference events in this story.

of the stories differed by only a few words (see Table 3). The versions were intermixed so that for a given story each participant heard two episodes with implied events and two episodes with explicit events. Thus, each participant heard 40 episodes inferring the inference event and 40 episodes explicitly stating the inference event. Half the participants heard one set of stories, the other half heard a second set, so that each episode was heard equally often in each version. Within each episode, we examined two critical time points: (1) the verb point, or the time at which a verb in a text implied or explicitly stated the inference; and (2) the coherence break, or the moment story coherence was impaired unless the inference was generated. Although the two versions of the stories varied at the verb point, the stories were identical at the coherence break. Thus, differences in neural activity at the coherence break should result from comprehender-driven processes reflecting interpretation of the story based on information that occurred earlier in the story. The distance between the verb point and coherence break ranged from 2 to 8 s.

## 5.3. Reading Span task

Outside of the scanner, participants performed a reading span task to assess working memory capacity (Singer and Ritchot, 1996). In this task, participants read aloud a set of sentences that included between two and six sentences. After each set, participants said aloud the last word in each set of sentences. Then, participants were asked to recall two missing words from one of the sentences in the set they had just seen

previously. The total number of final words recalled in the reading span task was used to distinguish low from high working memory capacity individuals. Using a median split of the distribution of final words recalled (out of 44 final words presented), eight participants were classified as a low working memory capacity group (recalled 28–34 total words; $M = 32.11$, $SD = 2.32$) and nine participants were classified as a high working memory capacity group (recalled 35–41 total words; $M = 37.13$, $SD = 2.36$). In addition to our analyses across all subjects, we examined fMRI signal separately for implied and explicit events in the high working memory group. We also conducted analyses that directly compared fMRI signal for the low and high working memory group, and we examined fMRI signal separately for the low working memory group.

## 5.4. MR image acquisition

Imaging was performed at Northwestern University's Center for Advanced MRI using a 3 T Siemens Trio scanner and a standard transmit/receive head coil. Head motion was restricted with plastic calipers built into the coil and a vacuum pillow. Anatomical high-resolution T1-weighted images were acquired in the axial plane parallel to the ACPC plane at the end of every session. Functional images were acquired in the same axial plane as the anatomical images using a gradient echo planar sequence (TR = 2 s allowing 38 slices 3 mm thick, TE = 20 ms, matrix size 64 × 64 in 220 mm field of view). The imaging sequence was optimized for detection of the BOLD effect (Binder et al., 1999) including local shimming and 8 s of scanning prior to data collection to allow the MR signal to reach equilibrium. Participants listened to stories in three runs (run 1 = 9 min, 16 s, run 2 = 9 min, 32 s, run 3 = 8 min, 38 s); across three runs, a total of 823 whole-brain volumes were acquired. The total functional scan session took approximately 30 min with an additional 7 min to collect the anatomical scan.

## 5.5. MR data analysis

Images were co-registered through time using a three-dimensional registration algorithm (Cox, 1996). EPI volumes were spatially smoothed by using a 7.5 mm full-width half-maximum Gaussian kernel to accommodate residual anatomical differences across participants and improve signal-to-noise. Within each run, voxels were eliminated if the signal magnitude changed more than 10% between time points (TR = 2 s) or if the mean signal level was below a threshold defined by the inherent noise in the data acquisition. Finally, all of the runs were transformed (Collins et al., 1994) to conform to a standard stereotaxic atlas (Talairach and Tournoux, 1988) with a final isometric voxel size of 2.5 mm³.

Data were analyzed using general linear model (GLM) analysis (D. Ward, Deconvolution Analysis of fMRI Time Series Data, http://afni.nimh.nih.gov/afni) that extracted average estimated responses to each trial type, correcting for linear drift and removing signal changes correlated with head motion. For each participant, event-related analyses contrasted fMRI signal for three TRs reflecting 4–10 s following the implied events to that of explicit events at the two critical time points. These two time points—the verb point and the

coherence break—were determined by the auditory input; participants made no overt responses during the stories. For each time point, implied minus explicit difference scores from each participant were combined in a second-stage random-effects analysis to identify differences consistent across all participants. A combination of two thresholds was used in combination to determine significance: a $p$ value of less than 0.005 and a cluster size greater than 500 mm$^3$. Specifically, a cluster size threshold was set at 500 mm$^3$ in volume (32 normalized voxels, contiguous or separated by no more than 5 mm; or 14 original-sized voxels) in which each voxel was reliably different across all participants ($t(16) > 3.25$, $p < 0.005$ uncorrected). The $t$ value corresponding to $p < 0.005$ differs in our working memory analyses due to the different subjects in each working memory group; high working memory group ($t(8) > 4.02$), high compared to low working memory group ($t(16) > 3.28$), and low working memory group ($t(7) > 3.82$). The only significant clusters were in the STG, consistent with a priori predictions: right hemisphere STG when the event was implied and left hemisphere STG at the coherence break. For each cluster presented in the tables, we report signal change as a percent of average signal within the cluster.

## Acknowledgments

REFERENCES

Barch, D.M., Braver, T.S., Sabb, F.W, Noll, D.C., 2000. Anterior cingulate and the monitoring of response conflict: evidence from an fMRI study of overt verb generation. J. Cogn. Neurosci. 12, 298–309.

Beeman, M., 1993. Semantic processing in the right hemisphere may contribute to drawing inferences from discourse. Brain Lang. 44, 80–120.

Beeman, M., Bowden, E.M., Gernsbacher, M.A., 2000. Right and left hemisphere cooperation for drawing predictive and coherence inferences during normal story comprehension. Brain Lang. 71, 310–336.

Binder, J.R., Frost, J.A., Hammeke, T.A., Bellogwan, P.S.F., Rao, S.M., Cox, R.W., 1999. Conceptual processing during the conscious resting state: a functional MRI study. J. Cogn. Neurosci. 11, 80–93.

Bryden, M.P., 1982. Laterality: Functional Asymmetry in the Intact Brain. Academic Press, New York.

Chiarello, C., Burgess, C., Richards, L., Pollock, A., 1990. Semantic and associative priming in the cerebral hemispheres: some words do, some words don't … sometimes, some places. Brain Lang. 38, 75–104.

Collins, D.L., Neelin, P., Peters, T.M., Evans, A.C., 1994. Automatic 3D intersubject registration of MR volumetric data in standardized Talairach space. J. Comput. Assist. Tomogr. 18, 192–205.

Cox, R.W., 1996. AFNI: software for analysis and visualization of functional magnetic resonance neuroimages. Comput. Biomed. Res. 29, 162–173.

Estevez, A., Calvo, M.G., 2000. Working memory capacity and time course of predictive inferences. Memory 8, 51–61.

Ferstl, E.C., von Cramon, D.Y., 2001. The role of coherence and cohesion in text comprehension: an event-related fMRI study. Cogn. Brain Res. 11, 325–340.

Ferstl, E.C., Rinck, M., von Cramon, D.Y., 2005. Emotional and temporal aspects of situation model processing during text comprehension: an event-related fMRI study. J. Cogn. Neurosci. 17, 724–739.

Fletcher, P.C., Happe, F., Frith, U., Baker, S.C., Dolan, R.J., Frackowiak, R.S.J., Frith, C.D., 1995. Other minds in the brain: a functional imaging study of "theory of mind" in story comprehension. Cognition 57, 109–128.

Frith, C.D., Friston, K.J., Liddle, P.F., Frackowiak, R.S., 1991. A PET study of word finding. Neuropsychologia 29, 1137–1148.

Goel, V., Dolan, R.J., 2001. The functional anatomy of humor: segregating cognitive and affective components. Nat. Neurosci. 4, 237–238.

Goel, V., Dolan, R.J., 2004. Differential involvement of left prefrontal cortex in inductive and deductive reasoning. Cognition 93, 109–121.

Graesser, A.C., Singer, M., Trabasso, T., 1994. Constructing inferences during narrative text comprehension. Psychol. Rev. 101, 371–395.

Jung-Beeman, M., 2005. Bilateral brain processes for comprehending natural language. Trends Cogn. Sci. 9, 512–518.

Jung-Beeman, M., Bowden, E.M., Haberman, J., Frymiare, J.L., Arambel-Liu, S., Greenblatt, R., et al., 2004. Neural activity when people solve verbal problems with insight. Public Libr. Sci. Biol. 2, 500–510.

Just, M.A., Carpenter, P.A., 1992. A capacity theory of comprehension: individual differences in working memory. Psychol. Rev. 99, 122–149.

Kintsch, W.A., 1988. The role of knowledge in discourse comprehension: a construction–integration model. Psychol. Rev. 95, 163–182.

Kintsch, W.A., van Dijk, T.A., 1978. Toward a model of text comprehension and production. Psychol. Rev. 85, 63–394.

Kircher, T.T., Brammer, J.M., Andreu, N.T., Williams, S.C.R., McGuire, P.K., 2001. Engagement of right temporal cortex during processing of linguistic context. Neuropsychologia 39, 798–809.

Linderholm, T., 2002. Predictive inference generation as a function of working memory capacity and causal text constraints. Discourse Process. 34, 259–280.

Linderholm, T., van den Broek, P., 2002. The effects of reading purpose and working memory capacity on the processing of expository text. J. Educ. Psychol. 94, 778–784.

Long, D.L., Baynes, K., 2002. Discourse representation in the two cerebral hemispheres. J. Cogn. Neurosci. 14, 228–242.

Mason, R.A., Just, M.A., 2004. How the brain processes causal inferences in text: a theoretical account of generation and integration component processes utilizing both cerebral hemispheres. Psychol. Sci. 15, 1–7.

McKoon, G., Ratcliff, R., 1992. Inference during reading. Psychol. Rev. 99, 440–466.

Meyer, M., Friederici, A.D., von Cramon, D.Y., 2000. Neurocognition of auditory sentence comprehension: event related fMRI reveals sensitivity to syntactic violations and task demands. Cogn. Brain Res. 9, 19–33.

Noveck, I.A., Goel, V., Smith, K.W., 2004. The neural basis of conditional reasoning with arbitrary content. Cortex 40, 613–622.

Posner, M.I., Petersen, S.E., Fox, P.T., Raichle, M.C., 1988. Localization of cognitive operations in the human brain. Science 240, 1627–1631.

Rilling, J.K., Sanfey, A.G., Aronson, J.A., Nystrom, L.E., Cohen, J.D.,

2004. The neural correlates of theory of mind within interpersonal interactions. NeuroImage 22, 1694–1703.

Robertson, D.A., Gernsbacher, M.A., Guidotti, S.J., Robertson, R.R.W., Irwin, W., Mock, B.J., et al., 2000. Functional neuroanatomy of the cognitive process of mapping during discourse comprehension. Psychol. Sci. 11, 255–260.

Seger, C.A., Desmond, J.E., Glover, G.H., Gabrieli, J.D.E., 2000. Functional magnetic resonance imaging evidence for right-hemisphere involvement in processing unusual semantic relationships. Neuropsychology 14, 361–369.

Singer, M., Ritchot, K.F.M., 1996. The role of working memory capacity and knowledge access in text inference processing. Mem. Cogn. 24, 733–743.

Singer, M., Andrusiak, P., Reisdorf, P., Black, N.L., 1992. Individual differences in bridging inference processes. Mem. Cogn. 20, 539–548.

St. George, M., Mannes, S., Hoffman, J.E., 1997. Individual differences in inference generation: an ERP analysis. J. Cogn. Neurosci. 9, 776–787.

St. George, M., Kutas, M., Martinez, A., Sereno, M.I., 1999. Semantic integration in reading: engagement of the right hemisphere during discourse processing. Brain 122, 1317–1325.

Talairach, J., Tournoux, P., 1988. Co-planar Stereotaxic Atlas of the Human Brain: 3-dimensional Proportional System: An Approach to Cerebral Imaging. Thieme Medical Publishers, New York.

Thompson-Schill, S.L., D'Esposito, M., Aguirre, G.K., Farah, M.J., 1997. Role of left inferior prefrontal cortex in retrieval of semantic knowledge: a reevaluation. Proc. Natl. Acad. Sci. 94, 14792–14797.

van den Broek, P., 1990. The causal inference maker: towards a process model of inference generation in text comprehension. In: Balota, D.A., Flores d'Arcais, G.B., Rayner, K. (Eds.), Comprehension Processes in Reading. Lawrence Erlbaum Associates, New Jersey, pp. 423–445.

Virtue, S., van den Broek, P., 2005. Hemispheric processing of anaphoric inferences: the activation of multiple antecedents. Brain Lang. 93, 327–337.

Virtue, S., van den Broek, P., Linderholm, T., in press. Hemispheric processing of inferences: the effects of textual constraint and working-memory capacity. Mem. Cogn.

Wagner, A.D., Parâe-Blagoev, E.J., Clark, J., Poldrack, R.A., 2001. Recovering meaning: left prefrontal cortex guides controlled semantic retrieval. Neuron 31, 329–338.