Neural Indicators of Inference Processes in Text Comprehension: An Event-related Functional Magnetic Resonance Imaging Study

Uwe Friese[1], Roland Rutschmann[2], Markus Raabe[2], and Franz Schmalhofer[1]

Abstract

■ During language comprehension, readers or listeners routinely infer information that has not been stated literally in a given text or utterance in order to construct a coherent mental representation (situation model). We used a verification task in a behavioral study and in an event-related functional magnetic resonance imaging (fMRI) experiment to investigate the inference construction process. After having read sentences that mention the outcome of an event explicitly, implicitly, or not at all, participants verified the plausibility of short statements with respect to the context of the just read sentence. The results of the behavioral study established the verification task as a valid method for studying inferences. In the fMRI study, the dorsomedial prefrontal cortex was the most prominent area that was involved in the processing of inference statements. Regions in the left and right temporal lobes were associated with comparison processes that are based on the propositional representations of context sentence and test statements. The results are discussed with respect to levels of representations and the memory systems that underlie the verification process in the different sentence conditions. ■

INTRODUCTION

While reading or listening, language comprehenders are not just passively receiving a message. In addition, they are continuously engaged in constructive processes to form a coherent representation of the described situation. In other words, they build a situation model that represents the state of affairs described by the text (see Van Dijk & Kintsch, 1983). Establishing a situation model of a text includes the construction of inferences, that is, representations of information that are not stated explicitly in a text, but that can be inferred on the basis of prior knowledge. This article is concerned with the question of where and how such inferences are formed in the human brain.

There are few neuroimaging studies that refer directly to inference processes in the brain. So far, the predominant paradigm has been to manipulate the coherence between short sentences because different theoretical accounts on inferencing agree that inferences are generated automatically if they are needed to establish local coherence between subsequent sentences (Graesser, Singer, & Trabasso, 1994; McKoon & Ratcliff, 1992). In a functional magnetic resonance imaging (fMRI) study on causal inferences, Mason and Just (2004) varied the causal relatedness of two-sentence passages. They supposed that inference generation and integration processes should be enhanced in moderately related sentences. Their results showed that the bilateral dorsolateral prefrontal cortices (dlPFC) might be involved in the generation of inferences, and that right hemisphere language areas seem to play a role in the subsequent integration of inferences into the reader's text representation. Using a very similar paradigm, Kuperberg, Lakshmanan, Caplan, and Holcomb (2006) identified a bilaterally distributed network of areas that seemed to mediate causal inferences, including lateral and medial portions of the prefrontal cortex (PFC), the middle (MTG) and superior temporal gyri (STG), as well as the inferior parietal lobes. Ferstl and von Cramon (2001) also studied the coherence of sentence-pairs using fMRI. Their participants had to decide whether two sentences were pragmatically related. Contrary to studies that report right hemisphere activations, the authors found two areas—one in the left dorsomedial prefrontal cortex (dmPFC) and one smaller region in the posterior cingulate gyrus and precuneal area—that showed higher activations in coherent trials than in incoherent trials. The authors related the activation in the medial portion of the superior frontal gyrus to the initiation of inference processes. Furthermore, Virtue, Haberman, Clancy, Parrish, and Jung Beeman (2006) conducted an fMRI study in which participants read stories that consisted of

[1]University of Osnabrueck, Germany, [2]University of Regensburg, Germany

multiple sentences without any concomitant secondary task. They measured neural activity at two critical points in the text, namely, when a verb implied an inference (verb point) and when an inference was needed to establish coherence between successive sentences (coherence break). Inference generation was associated with activity in the right superior temporal gyrus at the verb point, whereas at the coherence break, activity was predominantly found in the left superior temporal gyrus. Thus, research so far has suggested that there are a number of regions that seem to contribute to the processing of inferences, but the specific functions of these regions again remain somewhat unclear.

The goal of the present study was to further specify the functional neuroanatomy of inference generation processes. We chose the text comprehension model of Van Dijk and Kintsch (1983) as an appropriate theoretical framework because, firstly, the theory specifies how the coherence of sentences can be operationalized based on different levels of representations. Secondly, the theory specifies some important memory structures and processes that are tapped during language comprehension.

Van Dijk and Kintsch (1983) distinguish between three basic levels of text representations: a surface level, a propositional level, and a situational level. At the surface level, physical characteristics and the verbatim wording of the text are represented. The propositional level contains information about the linguistic structure of the text and the meaning of its elements. Finally, at the situational level, a mental model of the situation (situation model) described in the text exists (i.e., a representation of what the text refers to).

For our experiments we translated and modified sentence material that was used in numerous previous behavioral studies on inference processing (McKoon & Ratcliff, 1986; e.g., Yang, Perfetti, & Schmalhofer, 2007; Schmalhofer, McDaniel, & Keefe, 2002; McDaniel, Schmalhofer, & Keefe, 2001; Keefe & McDaniel, 1993; Potts, Keenan, & Golding, 1988). These materials consist of sentences such as, *While the flight attendant served the passenger a glass of red wine turbulence caused the wine to spill* (see also Table 1). Referring to this context sentence, the statement *The wine was spilled* can be characterized as an explicit repetition. By modifying the ending of the context sentence, different relationships between the representation of the test statement *The wine was spilled* and the representation of the context sentence can be established. If the context sentence of the example above ends with a paraphrase of the verb "spilled": *turbulence caused the wine to splash*, the test statement's representation differs from the explicit version mainly in the surface representation. In this case, the representations at the situational level and the propositional level overlap to a large degree. The "inference" ending *turbulence occurred which was very severe* provides enough information for the reader to infer that *The wine was spilled* is a plausible consequence. With regard to the context sentence, McDaniel et al. (2001) have shown that the inference (that the

Table 1. Illustration of the Text Material and the Experimental Conditions

Condition	*Headline*	*Sentence*	*Verification Probe*	*Expected Response*
Explicit	On the airplane	While the flight attendant served the passenger a glass of red wine turbulence caused the wine to *spill*.	"wine spilled"	yes
Paraphrase	On the airplane	While the flight attendant served the passenger a glass of red wine turbulence caused the wine to *splash*.	"wine spilled"	yes
Inference	On the airplane	While the flight attendant served the passenger a glass of red wine turbulence *occurred which was very severe*.	"wine spilled"	yes
Unrelated	On the airplane	While the flight attendant served the passenger a glass of red wine *the plane was at cruising altitude*.	"wine spilled"	no
Filler	The injured pet	The veterinarian was not able to cure Daniel's sick cat and finally she suggested to put it down.	"cat climbed"	no
Pseudo	Pseudowords	Gij vorniß hul pon eunif urh pirka sohldu rusq oinik wasin ehk eosep og peeze schmurlisan ilf uis.	"ilf uis"/"gij vorniß"	yes/no

The original text material was presented in German. In each trial, participants received a headline, a sentence or pseudoword sequence, and a verification probe. In the explicit, paraphrase, inference, and unrelated conditions, variations of the same theme (airplane flight) were created, while the identical verification probe ("wine spilled") was used. Filler trials were included to balance the expected ratio of yes and no responses. In pseudoword trials, the test probe consisted of either the first two or the last two pseudowords of the sequence.

wine was spilled) is only represented at the situational level. Finally, if the context sentence is ''unrelated'' in that it does not provide an explanation for the test statement (*While the flight attendant served the passenger a glass of red wine, the plane was at cruising altitude*), it can be assumed that the statement is not part of the representation of the context sentence at all.

Based on previous work (McDaniel et al., 2001; Kintsch, Welsch, Schmalhofer, & Zimny, 1990; Van Dijk & Kintsch, 1983), it can be argued that the representations of explicit test statements match those of the context sentences on surface, propositional, and situational level. Paraphrases share corresponding representations only at the propositional and situational levels. In contrast, for inference statements only, the situation-level representations match. Lastly, the representational levels of unrelated statements cannot be mapped to the representation of the context sentence at all. Consequently, these differences in the representations typically lead to differential behavioral response patterns in sentence recognition tasks. Usually, explicit sentences are recognized with the shortest response latencies, followed by paraphrase sentences, inferences, and unrelated sentences (e.g., Schmalhofer & Glavanov, 1986). Correspondingly, the proportion of ''yes'' responses decreases in the same sequence—it is highest for explicit sentences and lowest for unrelated sentences. The (standardized) differences between these endorsement rates have been used to estimate the strength of the memory trace of the test sentence at the three representational levels in the following manner (Kintsch et al., 1990). The difference between explicit and paraphrase conditions is supposed to reflect the strength of the surface-level representation. An estimate of the propositional representation is given by the difference between paraphrases and inferences, and finally, the strength of the situational representation can be approximated by the difference between inferences and unrelated sentences.

In our study, we used an analogous rationale to define contrasts that characterize the neural correlates of processing at the different levels of representation. By contrasting explicit, paraphrase, inference, and unrelated conditions with each other, we can draw specific conclusions about which brain regions are involved in the processing at the surface, propositional, and situational levels of text representations. In particular, we can isolate where inferences are constructed from the situation model by comparing paraphrase trials—where both situation model and propositional representations exist—with inference trials, which presumably are built from the situation model only.

With respect to the comparison processes that work at the different levels of representation, we assume the following. First, we suppose that both the context sentence and the verification statement can be accessed in working memory (Graesser, Millis, & Zwaan, 1997). In the explicit condition, the surface-level match of verification statements and the corresponding words in the context sentences can therefore be detected within the working memory system. Secondly, for comparisons at the propositional level, semantic memory has to be tapped. Thirdly, determining the relations between situational representations additionally requires more constructive processes that are closely linked to the episodic memory system (e.g., schema-driven processes, see Lea, Mulligan, & Walton, 2005; Kintsch et al, 1990).

This third assumption may need some further elaboration. The verification judgment in the inference and unrelated condition must be based on the assessment, whether the test statement is a plausible consequence of the situation that is described by the context sentence. For this purpose, previous experiences of the same or similar situations play a crucial role. Technically, in the computational model of Kintsch et al. (1990), this is achieved by measuring the activation that ''flows'' from the multilevel representation of the context sentence to the representation of the test statement. If the test statement fits well into the described context, it will receive activation from the situational level even if it differs on surface and propositional level. Recently, a similar concept for plausibility has been proposed by Connell and Keane (2006) in their general plausibility analysis model (PAM). In PAM, plausibility depends on concept-coherence, which is ''about consistency with previous experience, as measured by the degree of fit between a given scenario and prior knowledge'' (Connell & Keane, 2006). We adopt this notion, and thus, argue that the situation-level evaluation of inference and unrelated statements depends to a high degree on the episodic memory system.

Our participants read sentences and immediately afterward decided whether a short target statement was true with respect to the just read sentence. We used short statements instead of full sentences to be able to precisely determine the time point when the critical information was presented. One possible disadvantage of this approach is that the execution of a verification task does not necessarily reflect the same processes that occur during natural language comprehension. However, it can be argued that intentional verification processes closely resemble those that contribute to normal text comprehension (Singer, 2006). To make sure that our inference sentences effectively led to high affirmation rates in this verification task and to allow comparability to previous research with similar material, we first conducted a behavioral experiment (Experiment 1). We then used event-related fMRI in an independent sample to characterize the brain activity during the processing of the verification statements (Experiment 2).

EXPERIMENT 1

It has been shown in previous research that inferences arise from the situation-level representation and that inferencing is enhanced if readers are instructed to focus on the situation conveyed by the text (McDaniel et al., 2001; Schmalhofer & Glavanov, 1986). These studies used word pronunciation and sentence recognition tasks. In Experiment 1, we tested if we would get comparable results with a statement verification task—a procedure that would be suited better for event-related fMRI acquisition because it does not involve the reading of whole sentences. In addition, it does not require overt verbal responses. We expected to find higher endorsement rates and shorter response latencies for the verification statements in the explicit and paraphrase conditions than for verification statements in the inference condition. In the unrelated condition, the verification statements should be rejected predominantly, which would presumably lead to the slowest response times.

Methods

Participants

There were 40 volunteers in the behavioral study (20 women, mean age = 22.4 years). All were right-handed native speakers of German. They were healthy, had no history of neurological illness, and all had normal or corrected-to-normal vision (contact lenses).

Design and Materials

We used 72 text scenarios. Almost all of them were taken from Yang et al. (2007, pp. 73–88). These were translated into German and adjusted for the purposes of current experiment. We thereby obtained a headline, four sentence variations and one test statement for each scenario. With reference to the test statement, the sentence variations constituted (a) an *explicit* trial, (b) a *paraphrase* trial, (c) an *inference* trial, or (d) an *unrelated* trial (Table 1). The general structure of all sentences was that a situation involving some actions was described. Furthermore, the outcome of this action was mentioned either explicitly or implicitly in the following way. (a) Explicit sentences contained exactly the same words as the test statements. In (b) paraphrases, synonymous phrases were used. The (c) inference sentences did not mention the outcome in question directly but provided sufficient information to infer it. Lastly, the (d) unrelated sentences provided only insufficient and incoherent information regarding the test statement as a possible outcome of the described situation.

Each sentence consisted of 18 words. The first 13 (±1) words were identical in the explicit, paraphrase, and inference conditions. To ensure that each sentence appeared equally often in all conditions across participants, we divided the sentence pool into four lists and counterbalanced lists and conditions by a Latin square. The expected yes and no responses were balanced by including additional 18 filler sentences that were supposed to be correctly answered with "no" like the unrelated sentences. Furthermore, we constructed 18 pronounceable pseudoword sequences by permuting individual letters in sentences that were not contained in any of the other experimental conditions. This pseudoword condition was included as a baseline task for the fMRI study. For the pseudoword sequences, the verification statement consisted of the first two or the last two pseudowords of the sequence. The statement had to be answered with "yes" when the pseudowords in the verification task statement were the same as the last words of the sequence. This baseline task is similar to the verification task in that it includes basic memory and decision processes but lacks language-specific coherence building processes. The presentation of the total 108 trials was pseudorandomized to provide three blocks of 36 trials, six of each of the five experimental conditions (pseudowords, explicit, paraphrase, inference, unrelated) plus six filler trials.

Procedure

All participants were informed about the type of stimuli presented to them and were instructed to give a positive response only if the verification statement was sensible with respect to the preceding sentence or, in case of pseudoword sequences, if the two items of the test probe were identical to the last two items presented in the pseudoword sequence before. They were also asked to read the pseudoword sequences as if it was normal text.

Stimuli were generated with E-prime (Psychology Software Tools, Pittsburgh, PA) and presented on a 22-in. CRT screen. Responses were recorded with a serial response box (Psychology Software Tools) which the participants operated with their right hand (index finger for "yes," middle finger for "no").

We presented the context sentences word-by-word, centered on the screen (rapid serial visual presentation) with 600 msec presentation time for the headline, followed by a blank screen for 600 msec and then successively 300 msec for each of the 18 words, with an interstimulus interval of 300 msec (12 sec in total). After the last word, a central fixation cross appeared for 2 sec, followed by the test statement for a maximum of 4.3 sec. The test statement disappeared from the screen after the participants' responses. After the response time window of 4.3 sec, a central fixation cross was presented for a pseudorandomly varied interval of 7 to 11 sec. In total, one trial lasted for 27 sec.

The experiment was preceded by seven training trials. Subsequently, a total of 108 trials followed (18 trials each for explicit, paraphrase, inference, unrelated, filler, and pseudoword condition). After blocks of 36 trials short resting periods of up to 3 min were given upon demand of the participants. The whole experimental procedure took approximately 65 min.

Results

As the primary purpose of the behavioral study was to verify that the modified sentence material and the presentation technique lead to results that are comparable to similar behavioral experiments (e.g., McDaniel et al., 2001), we will only mention the most important aspects here. For all statistical tests, we assumed $\alpha = .05$ and in pairwise comparisons we adjusted α according to the Bonferroni–Holm procedure. Average response latencies and endorsement rates are presented in Table 2. An analysis of variance (ANOVA) on the response times revealed significant differences between the sentence conditions [$F(4, 156) = 19.7$, $p < .01$]. Descriptively, there was an increase in latencies from the pseudoword condition to explicit, paraphrase, inference, and unrelated conditions. Planned pairwise comparisons were carried out for the following conditions: (1) pseudowords versus explicit, (2) explicit versus paraphrases, (3) paraphrases versus inferences, and (4) inferences versus unrelated. Significant differences were found between the explicit and paraphrase condition [$t(39) = 3.45$, $p < .01$], as well as between paraphrases and inferences [$t(39) = 4.59$, $p < .01$]. The differences between inference and unrelated condition [$t(39) = 2.04$, $p = .05$], as well as between pseudowords and explicit sentences [$t(39) < 1$], were not significant on the corrected α-level. An ANOVA on the response frequencies (yes responses in explicit, paraphrase, and inference trials; no responses in unrelated trials; correct responses in pseudoword trials) also revealed significant differences [$F(4, 156) = 9.87$, $p < .01$]. Most notably, the endorsement rates for explicit and paraphrase trials were slightly higher than in the other conditions. The difference between paraphrases and inferences was significant [$t(39) = 5.46$, $p < .01$]. All other comparisons were not significant.

Discussion

Table 2 depicts that, overall, participants showed the expected response pattern in the verification test. Specifically, the proportion of yes responses was greater for paraphrases than for inferences. This presumably reflects the influence of the propositional representation that is accessible for paraphrases but not for inferences. This result is also supported by the faster response times to paraphrases compared to inferences.

EXPERIMENT 2

The main goal of the present study was to investigate the functional neuroanatomy of inference generation in a sentence verification task. Therefore, we conducted an event-related fMRI study with the same material and design we used in Experiment 1. We were especially interested in the comparison of the processing of the verification statements in the inference and paraphrase conditions, as the contrast of these two conditions supposedly reflects the generation of inferences from the situation-level representation. Research on the neural correlates of sentence and discourse processing proposes that, during the reading of context sentences, a widespread network of brain areas is likely to be activated. This ‘‘extended language network’’ comprises the left inferior frontal cortex, anterior and posterior aspects of the left MTG and STG, and (although to a smaller extent) homologue areas of the right hemisphere (Ferstl, 2007; Ferstl, Neumann, Bogler, & von Cramon, 2007). As we have reviewed in the general introduction, several of these areas have been associated with inference processing. By utilizing the levels of representation approach, we expect to be able to differentiate between regions that primarily subserve processing on the lexical level (paraphrase condition > explicit condition) and regions that subserve constructive, episodic memory-based processes (inference condition > paraphrase condition).

Methods

Participants

Thirteen participants, who did not take part in Experiment 1 volunteered in the fMRI study for course credit

Table 2. Mean Response Latencies and Proportion of Responses from the Behavioral and Imaging Experiment

	Experiment 1 (n = 40)		*Experiment 2 (n = 13)*	
	Response Times (SE), msec	*Rel. Frequency (SE)*	*Response Times (SE), msec*	*Rel. Frequency (SE)*
Pseudowords (correct)	826 (43)	0.94 (0.02)	828 (37)	1 (0.00)
Explicit (‘‘Yes’’)	850 (28)	0.99 (0.00)	961 (54)	0.99 (0.01)
Paraphrases (‘‘Yes’’)	886 (29)	0.98 (0.01)	999 (57)	0.98 (0.01)
Inference (‘‘Yes’’)	994 (45)	0.89 (0.02)	1085 (61)	0.89 (0.04)
Unrelated (‘‘No’’)	1058 (38)	0.93 (0.01)	1207 (65)	0.90 (0.02)

For the pseudoword condition, all correct responses were analyzed. In the explicit, paraphrase, and inference conditions, only ‘‘yes’’ responses were included, whereas in the unrelated condition, ‘‘no’’ responses were considered.

or payment. Seven participants were women, and the average age was 22.8 years. All participants were right-handed native speakers of German. They were healthy, had no history of neurological illness, and all had normal or corrected-to-normal vision (contact lenses). Informed written consent for participation in the fMRI study was obtained from all participants.

Materials and Procedure

The fMRI study was, in all aspects, identical to the behavioral study except for the different laboratory setting. Stimuli were generated with E-prime (Psychology Software Tools) and were presented onto a screen within the MR-cabin with an LCD projector (JVC, DLA-G15, Yokohama, Japan) that was located outside of the RF-shielded room. The participants lay inside the magnet and viewed the screen via a coil-mounted mirror. Responses were recorded with a button box (LUMItouch button box, Photon Control, Burnaby, Canada) which the participants held in their right hand (index finger for "yes," middle finger for "no").

The main scanning session was preceded by seven training trials inside the magnet. Afterward, three experimental sessions with 36 trials (16.3 min) each followed. Between the sessions, short resting periods of up to 3 min were scheduled and were given upon demand of the participants. The experiment ended with the acquisition of a structural image, lasting for approximately 5 min. After the experiment, all participants were informed about the aim of the study.

fMRI Image Acquisition

Magnetic resonance images were acquired in a 1.5-T Siemens Sonata whole-body MRT equipped with an eight-channel head array coil (MRI-Devices Europe, Würzburg, Germany). Head fixation was achieved by using soft pads. Foam ear plugs and sound damping headphones were used for noise shielding.

During the functional scans, the blood oxygen level-dependent (BOLD) response was measured using a T2*-weighted gradient EPI sequence (TR = 3 sec, TE = 50 msec, flip angle = 90°, resolution 3×3 mm^2, number of slices = 35, interleaved acquisition sequence, slice thickness = 3 mm, distance factor: 0–10%). The acquired slices were rotated approximately 10° relative to the AC–PC line in order to cover prefrontal, parietal, and temporal regions in full and the majority of the occipital cortex, sometimes excluding the ventral extent of V1. A total of three functional sessions of 326 images each was recorded. The onset of the stimulus presentation was jittered with 0, 1, or 2 sec relative to the onset of the image acquisition to reach a virtual time resolution of 1 sec. Structural images were acquired for each participant using a T1-weighted MP-RAGE sequence (TR = 1900 msec, TE = 3.93 msec, resolution $1 \times 1 \times 1$ mm^3) at the end of the experiment.

Data Preprocessing and Analysis

All magnetic resonance data were preprocessed and statistically analyzed using SPM2 (Wellcome Department of Imaging Neuroscience, London, UK; www.fil.ion.ucl.ac.uk/spm/). Functional images were corrected for acquisition delays and realigned to the first image. A mean image was computed, and the structural image was coregistered to the mean of the functional images. After normalizing all images to the MNI-152 template, realigned functional images were resampled to $2 \times 2 \times 2$ mm^3 and spatially smoothed using an isotropic Gaussian kernel with 10 mm full width at half maximum.

A general linear model was fitted to each individual dataset, modeling the presentation of the headline, the subsequent sentence, and the verification probe for each condition separately. The modeling of the sentence presentation was split into three regressors corresponding to six words each. The verification process was modeled using a block with a duration of 1.8 sec starting from the onset of the test statement. This duration was chosen based on the participants' largest average response time to inference statements. Individual t-test contrasts were calculated between verification tasks in the inference, explicit, paraphrase, and unrelated conditions as well as for the pseudoword condition. For statistical analysis, a random-effects model was used to test for the presence of significant activation clusters (t statistics). Statistical maps were thresholded with an uncorrected p value of $p < .001$ ($t = 3.93$). Unless otherwise noted, only clusters surpassing a corrected p value of $<.05$ on cluster level are reported as significantly activated. This corresponds to a minimal cluster volume of approximately 1200 mm^3. The same statistical analysis was also executed for the regressor that represented the reading of the last six words of the context sentence.

Regions of interest (ROI) were functionally defined using the group comparison of reading real words (explicit, paraphrase, inference, and control conditions) versus reading pseudowords (see Table 4, Figure 1). Within the five clusters that showed significantly differential activation, we chose all local maxima that surpassed a threshold of $z > 4.0$ and were at least 18 mm apart from each other. A sphere of 10 mm radius was drawn around each local maximum and intersected with the original cluster. This resulted in 10 ROIs (see Table 3). Statistical analysis within the ROIs was conducted using the MARSBAR toolbox for SPM (Brett, Anton, Valabregue, & Poline, 2002), which averages over all voxels within one ROI. The modeling and contrasts used were exactly the same as the ones used for the whole-brain analysis, the only difference being in the correction for multiple comparison. For the ROI analysis, a Bonferroni correction for the 10 ROIs was applied. Time courses of the BOLD

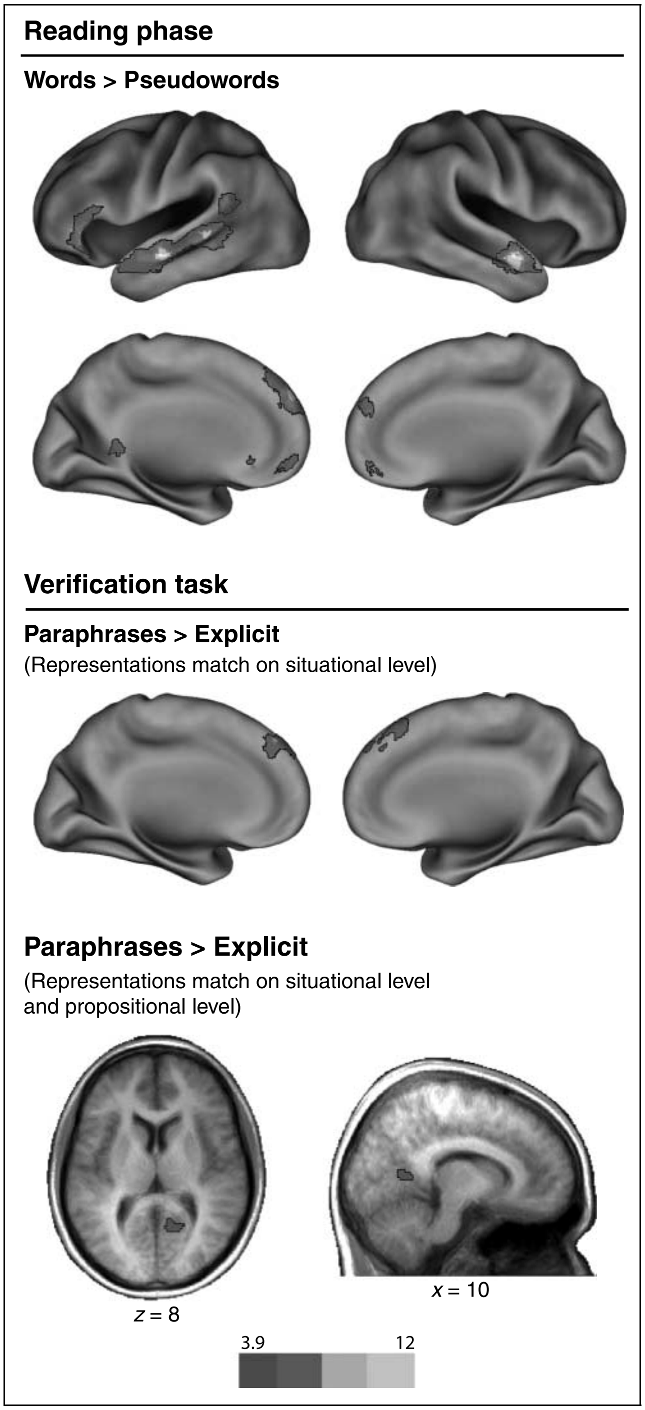

Figure 1. Results of whole-brain analysis. To characterize activations during the reading of the context sentences, the mean activation of the explicit, paraphrase, inference, and unrelated conditions is contrasted against the activation of the pseudoword condition. The contrasts of inference > paraphrase condition and paraphrases > explicit condition are displayed with respect to the regressor for the verification test. No significant activation was found in the contrast of unrelated versus inference condition. Statistical activation maps are superimposed on inflated brain surfaces using the Caret software and surface-based brain atlases from http://brainvis.wustl.edu/caret/ (Van Essen, 2002; Van Essen et al., 2001). Brain slices for the illustration of the paraphrase > explicit contrast were created from the averaged anatomical images of all participants.

response, starting with the presentation of the headline and lasting 28 sec, were extracted for each participant in each ROI for the four sentence conditions. Time-course plots were created by first normalizing each time course to percent signal change (relative to the mean signal of that region). Subsequently, the individual time courses were averaged over participants separately for each ROI and condition.

Results from the fMRI Study

Behavioral Data

For all statistical tests, we assumed $\alpha = .05$, and in pairwise comparisons, we adjusted α according to the Bonferroni–Holm procedure. An ANOVA revealed that the participants' response times differed significantly across sentence conditions (explicit, paraphrase, inference, unrelated, and pseudowords) [$F(4, 48) = 20.5$, $p < .001$]. The effect of the sentence conditions was also significant for the response frequencies [$F(4, 48) = 6.3$, $p < .001$]. The response latencies increased descriptively from the pseudoword condition to explicit, paraphrase, inference, and unrelated conditions. The performance in inference and unrelated conditions was slightly lower than in the other conditions (see Table 2). Planned pairwise comparisons were carried out for the following conditions: (1) pseudowords versus explicit, (2) explicit versus paraphrases, (3) paraphrases versus inferences, and (4) inferences versus unrelated. The response time difference between the pseudoword and explicit condition [$t(12) = 3.08$, $p = .01$] reached significance. All other pairwise comparisons were not significant on the corrected α-levels: explicit versus paraphrases, $t(12) = 1.19$, $p = .26$; paraphrases versus inferences, $t(12) = 2.25$, $p = .04$; inferences versus unrelated, $t(12) = 2.56$, $p = .03$.

Functional Imaging Data

The results of the whole-brain fMRI analysis are depicted in Figure 1 and Table 4. As a general validation of our design and data analysis, we first report the differential activations of the mean of all language trials, namely, the explicit, paraphrase, inference, and unrelated conditions, against the pseudoword condition for the reading phase of the experiment. This contrast was also used to define ROIs. We then carried out pairwise comparisons between the following conditions for the verification test phase of the experiment and for the regressor that represented the reading of the last six words of the context sentence. These contrasts were motivated by the levels of representation theory: (1) paraphrases versus explicit, (2) inferences versus paraphrases, and (3) unrelated versus inferences. (We also analyzed the reversed direction of contrasts, i.e., inference versus unrelated,

Table 3. Specification of Regions of Interest

ROI	Description	L/R	Brodmann's Area	MNI Coordinates	Size (mm^3)
1	Ventromedial prefrontal cortex	L/R	10	2, 56, −14	1432
2	Ventrolateral prefrontal cortex	L	47	−44, 32, −12	2592
3	Inferior frontal gyrus	L	45	−50, 18, 14	368
4	Middle temporal gyrus, mid part	L	21, 22	−54, −38, 2	2760
5	Anterior middle temporal gyrus	L	21	−58, −12, −14	3160
6	Posterior superior temporal gyrus	L	22, 40	−60, −50, 18	2144
7	Posterior cingulate cortex	L	30	−20, −52, 14	696
8	Anterior middle temporal gyrus	R	21	56, 0, −18	3120
9	Dorsomedial prefrontal cortex	L	9, 8	−8, 44, 46	2256
10	Dorsomedial prefrontal cortex	L	10, 9	−10, 56, 32	2936

Regions of interest were functionally defined using the contrast "Reading words > Reading pseudowords." Spheres of 10 mm radius around local maxima were intersected with the original clusters. Refer to the Methods section of Experiment 2 for details.

paraphrase versus inference, and explicit versus paraphrase, but found no significant differences.) Additionally, we calculated the same contrasts within our ROIs and completed these analyses with estimations of activation time courses for the different sentence conditions (Table 5, Figures 2 and 3). Lastly, we also compared the unrelated and inference conditions with the explicit condition as a common baseline in the ROIs.

Reading real words versus reading pseudowords. A widespread network of primarily left hemisphere areas was more activated during the reading of the context sentences (explicit, paraphrase, inference, and unrelated conditions) than during reading the pseudoword sequences (Table 4, Figure 1). We found activation in the left MTG and STG, extending into the left inferior parietal lobe. Further clusters were located in the left

Table 4. Whole-brain Analysis

Contrast/Region	Side	Brodmann's Area (BA)	Size (mm^3)	p_{corr}	z_{max}	MNI Coordinates (x, y, z)
Sentence Reading Regressor						
Words > Pseudowords						
MTG, IFG, STG	L	21, 47, 22	2776	<.01	5.74	−58, −12, −14
Dorsomedial PFC	L/R	9, 10, 8	8544	<.01	5.11	−10, 56, 32
MTG, ITG, STG	R	21	4760	<.01	5.47	56, 0, −18
Ventromedial PFC	L/R	11	3120	<.01	4.53	2, 56, −14
PCC	L	29	1216	.03	4.32	−20, −52, 14
Verification Test Regressor						
Paraphrase > Explicit (*ns*)						
PCC, cuneus	R	30	904	.06	3.92	10, −60, 8
Inference > Paraphrase						
Dorsomedial PFC	L/R	8/9	5472	<.01	4.91	−2, 46, 46
Unrelated > Inference						
No significant activation	–	–	–	–	–	–

Shown are specifications of clusters with cluster-level $p < .1$. MNI coordinates denote the voxel with highest z-value within the cluster. Reverse comparisons concerning the verification test regressor (explicit > paraphrase, paraphrase > inference, inference > unrelated) did not result in any significant clusters.

IFG = inferior frontal gyrus; ITG = inferior temporal gyrus; MTG = middle temporal gyrus; PCC = posterior cingulate cortex; PFC = prefrontal gyrus; STG = superior temporal gyrus.

Table 5. Regions-of-interests Analysis

		Inference > Paraphrase		*Inference > Explicit*		*Unrelated > Explicit*	
ROI	*Area*	*t*	p_{corr}	*t*	p_{corr}	*t*	p_{corr}
1	L/R ventromedial PFC	–	–	–	–	2.86	.07
2	L ventrolateral PFC	2.95	.06	4.88	<.01	5.78	<.01
3	L IFG	–	–	4.12	<.01	4.53	<.01
4	L MTG, mid part	–	–	3.57	.02	4.58	<.01
5	L anterior MTG	–	–	–	–	3.75	.01
6	L posterior STG	–	–	2.97	.06	7.33	<.01
7	L posterior CC	–	–	–	–	–	–
8	R anterior MTG	–	–	3.15	.04	4.51	<.01
9	L dorsomedial PFC	4.49	<.01	3.33	.03	4.58	<.01
10	L dorsomedial PFC	3.01	.05	3.04	.05	4.25	<.01

The table reports the three contrasts in which the statistical analysis resulted in *t* values with $p_{corr} < .1$ in at least one ROI. No other comparison met this criterion.

L = left; R = right; CC = cingulate cortex; IFG = inferior frontal gyrus; MTG = middle temporal gyrus; PFC = prefrontal cortex.

inferior and middle frontal gyri, the medial parts of left superior frontal gyrus, and a region near the left posterior cingulate cortex and the precuneus. Moreover, there were active areas in the right middle, inferior, and superior temporal gyri.

Whole-brain analysis. None of the predefined contrasts showed significant results with respect to the reading phase of the experiment. Concerning the verification test and the primary interest of this study, the following results were found. Although not significant according to the conventional criterion, there was a tendency for an area in the right posterior cingulate gyrus and the precuneus that was more activated in paraphrase trials than in explicit trials ($p = .06$). In the contrast of inference versus paraphrase conditions, one cluster in the bilateral superior and medial frontal gyri (dmPFC), corresponding to Brodmann's areas (BA) 8 and 9, was significantly activated (Table 4, Figure 1).

Regions-of-interest analysis. The results of the ROI analysis are presented in Table 5, Figure 2, and Figure 3. Again, there was no area in which paraphrases elicited significantly more activation than explicit statements. Likewise, the unrelated condition was not dissociated from the inference condition. The inference versus paraphrase contrast was significant in the posterior dmPFC (ROI 9) and close to significance in the more anterior dmPFC (ROI 10) as well as in the left anterior inferior frontal gyrus [IFG] (ROI 2). Descriptively, the activation time courses indicate that unrelated and inference conditions on the one hand, and explicit and paraphrase conditions on the other, elicited similar activation patterns in most ROIs. Except for the posterior cingulate area (ROI 7), unrelated and inference trials elicited more activation than paraphrase and explicit trials. This observation was confirmed by contrasting unrelated and inference conditions with the explicit condition as common baseline. Unrelated trials produced significantly more activation than explicit trials in all ROIs except for ROI 7. The inference condition was associated with more activation than the explicit condition in the left IFG (ROI 2 and 3), in the dmPFC (ROI 9), and in the left and right MTG (ROI 4 and 8).

Discussion

In this study, participants verified short statements after reading a context sentence. In the inference condition, the verification statement could be related to the context statement by an inference process. Moreover, the design of the sentence material was guided by the theoretical notion that language is represented on different levels (Schmalhofer et al., 2002; McDaniel et al., 2001). Comparisons between our conditions therefore reflect the processing of the verification statements at the three different representational levels to varying degrees.

In the paraphrase and explicit conditions, the verification statements and context sentences supposedly share the same situation-level representation as well as the propositional representation, and they differ only in the surface representation. As the literal wording of the test statement is not decisive for the verification task, the average proportion of yes responses for the explicit and the paraphrase conditions did not differ significantly in Experiment 1 and Experiment 2. The response latencies, however, were faster for explicit statements than for

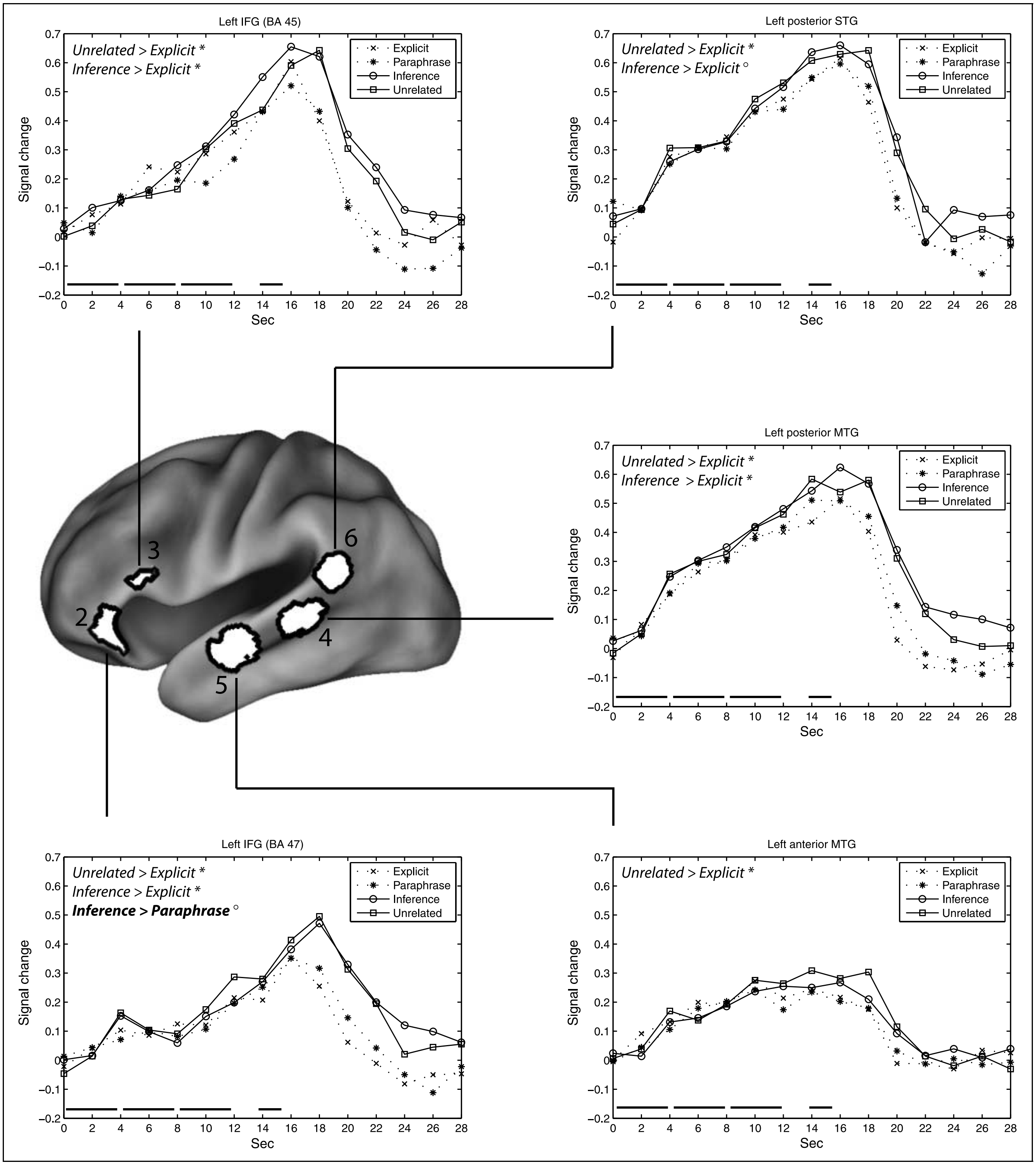

Figure 2. Results in lateral ROI of the left hemisphere. ROIs were projected on an inflated brain surface (see details in caption of Figure 1). Time-course diagrams show the estimated signal change in each ROI for the four sentence conditions from the onset of the headline to the end of the trial. Horizontal lines just above the *x*-axis of each diagram approximately indicate the timing details of the GLM regressors for sentence reading and verification test. In the upper left corner of each diagram, significant results with respect to the GLM verification regressor contrasts are indicated (*$p < .05$; °$p < .1$).

paraphrase statements in Experiment 1. Although in the whole-brain fMRI analysis, one cluster in the right posterior cingulate cortex almost reached significance in the corresponding contrast of paraphrase and explicit condition, no significant differences were found in any of the ROI for this contrast. Activation in the right posterior cingulate and precuneus area (ROI 7) has been found in some studies on language comprehension and

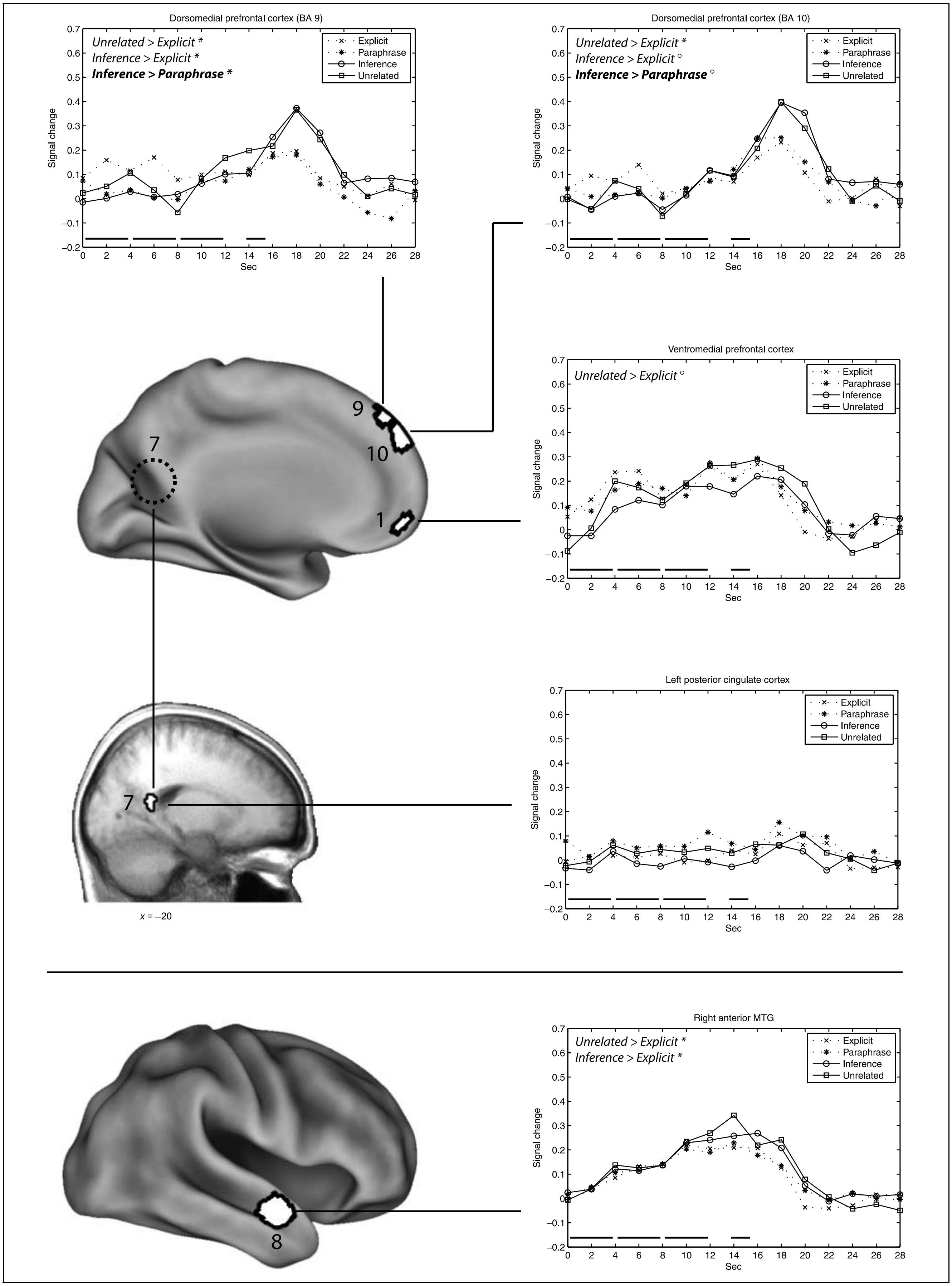

Figure 3. Results in the medial ROI of the left hemisphere and in the right temporal lobe (for details, see caption of Figure 2).

is assumed to reflect memory retrieval processes (Ferstl & von Cramon, 2002; Bottini et al., 1994).

Inference and paraphrase statements differ with respect to the surface level and the propositional level from the representations of the context sentences. As the verification tasks requires an evaluation of the meaning of the test statement in relation to the context sentence, the lack of a corresponding propositional representation in the inference condition requires an elaboration of the situation model. This is reflected by significant differences in the response times and endorsement rates to inference and paraphrase statements in Experiment 1. The behavioral results of Experiment 2 point into the same direction, although the corresponding comparisons are not significant at the corrected α-level. In the whole-brain fMRI analysis, we found that the dmPFC is the only area where significantly more activity is found in the inference trials than in the paraphrase trials. The ROI analysis confirmed this result (ROI 9 and 10) and additionally pointed to a stronger involvement of the left IFG (BA 47, ROI 2) in inference trials.

The representations of unrelated statements do not match the representations of the context sentences on neither level. Due to the verification instruction, the participants had to evaluate whether the unrelated statement could be integrated into the situation model that was built according to the context sentence, and if not, the unrelated statement had to be rejected. The behavioral results of Experiment 1 and Experiment 2 show no significant differences between these conditions. The interpretation of these results is difficult though because unrelated statements require a no response, whereas only yes responses are considered in the inference condition. Interestingly, the lack of significant activation in the comparison of unrelated and inference condition in the fMRI analyses seems to suggest that there is considerable overlap between the areas that are involved in the inference generation process on the one hand and in the evaluation of implausible statements on the other hand. This is also illustrated by the activation time-course diagrams that show closely corresponding curves for inference and unrelated trials in most ROIs.

Inference Processing and the Prefrontal Cortex

According to McDaniel et al. (2001) and Schmalhofer et al. (2002), inference and paraphrase statements share the same situation-level representations with the corresponding context sentence but differ with respect to the propositional- and surface-level representations. Thus, more activity in the inference trials than in paraphrase trials reveals comparison processes at the situational level, which are thought to rely on the episodic memory system (Magliano, Radvansky, & Copeland, 2007; Kintsch, 1988). The most prominent region we could associate with these processes was the dmPFC. There was one large significant cluster found in the whole-brain analysis (Figure 1). In the ROI analysis, significant activations were located in the posterior dmPFC (BA 9, ROI 9) and tendencies toward significance in the anterior dmPFC (BA 10, ROI 10) as well as in the left anterior inferior PFC (ROI 2).

Our finding that regions in the dmPFC seem to play a central role in inference processing is in close agreement with different lines of research. In the language domain, Ferstl and von Cramon (2001) have demonstrated that the dmPFC was activated when their participants evaluated the coherence of pragmatically related sentences. Mazoyer et al. (1993), for example, also found activity in the medial aspects of the superior frontal gyrus comparing the processing of meaningful stories versus distorted stories. A recent meta-analysis also pointed to the importance of the dmPFC for establishing coherence (Ferstl et al., 2007).

Moreover, the dmPFC seems to be involved in a variety of other higher-level cognitive processes such as theory of mind processing (e.g., Gallagher & Frith, 2003; Frith & Frith, 1999; Fletcher, Happe, Frith, & Baker, 1995), reasoning (Ruff, Knauff, Fangmeier, & Spreer, 2003; Goel, Gold, Kapur, & Houle, 1997), task-switching (Forstmann, Brass, Koch, & von Cramon, 2005), response inhibition (Li, Huang, Constable, & Sinha, 2006), and autobiographical and episodic memory (Svoboda, McKinnon, & Levine, 2006; Gilboa, 2004; Graham, Lee, Brett, & Patterson, 2003; Cabeza & Nyberg, 2000). To embrace these diverse findings, Ferstl and von Cramon (2002) suggest that the dmPFC might have a "domain-independent functionality related to volitional aspects of the initiation and maintenance of nonautomatic cognitive processes" (p. 1611). Recently, it has been proposed that this region is part of a network that generally subserves the reconstruction or simulation of past and future episodes based on prior experience (Addis, Wong, & Schacter, 2007; Buckner & Carroll, 2007; Schacter & Addis, 2007). In this view, the abilities to infer information and to predict events are core functions of the episodic memory system. In the field of language comprehension, Zwaan (2004) has presented a related framework in which comprehension processes are also viewed as simulations at the situational level. On the basis of these ideas and the studies reviewed above, we think the activations we found in the dmPFC reflect the intentional evaluation of the plausibility of the verification statements based on the participants' prior knowledge.

In addition to the significant activation in the dmPFC, the ROI analysis revealed some evidence for higher activation in the inference condition than in the paraphrase condition in the left anterior inferior PFC (ROI 2). Mostly based on subsentence-level studies, it has been argued that the left inferior PFC is critically involved in the selection of semantic representations (Thompson-Schill, D'Esposito, Aguirre, & Farah, 1997) or in the control of semantic retrieval (Wagner, Pare-Blagoev, Clark, & Poldrack, 2001). Jung-Beeman (2005) also highlights the importance of this region for the selection of semantic concepts and, thereby, for the integration into a wider context. Similarly, Hagoort (2005) suggested that an important function of the left IFG for language comprehension is "unification," that is, "integration of lexically retrieved information into a representation of multi-word utterances." Kuperberg et al. (2003), for example, showed that activity in these regions increases during the processing of pragmatically anomalous sentences relative to normal sentences. They suggest that a left temporo-frontal language network serves in mapping and conceptually integrating words onto information in semantic memory. In our experiment, increased activity in the anterior inferior PFC might also reflect the attempt to map and integrate inference and unrelated statements on a conceptual level. The absence of matching propositional and surface representations in these conditions probably requires increased integration efforts. However, as we did not control our sentence material for overall difficulty, our results remain somewhat inconclusive in this respect.

Semantic Processing and the Temporal Lobes

In the present study, the critical contrast between inference and paraphrase conditions did not yield significant results in temporal areas—neither in the whole-brain analysis nor in the ROI analysis. However, the ROI analysis provides evidence that these temporal areas did respond differentially to the test statements to some degree. In all temporal areas (ROI 4, 5, 6, 8)—including the right anterior MTG—the unrelated condition elicited more activity than the explicit condition. In the left middle portion of the MTG (ROI 4), the contrast between inference and explicit condition was also significant, and in the posterior left STG (ROI 6), this contrast was close to significance.

Most probably, the modulations of temporal activity in the anterior and middle temporal lobes reflect more intense semantic processing of unrelated and inference statements, as compared to explicit statements, because the former are not part of the propositional representation of the context sentence. Of course, unrelated and inference statements also differ with respect to the surface representation from explicit statements. As the paraphrase condition did not elicit higher activation than the explicit condition in any of our ROI, we suspect that the modulations of activity in temporal areas was rather due to semantic processes. The involvement of areas in the anterior and middle parts along the superior temporal sulcus in semantic processing of sentences and discourse is suggested by numerous studies (Virtue et al., 2006; Ferstl, Rinck, & von Cramon, 2005; Kuperberg et al., 2003; Ferstl & von Cramon, 2001, 2002; Vandenberghe, Nobre, & Price, 2002; St George, Kutas, Martinez, & Sereno, 1999; Bottini et al., 1994; Mazoyer et al., 1993). Furthermore, as in our study, activations in the anterior and middle parts of the temporal lobes often occur bilaterally during sentence comprehension (Ferstl, 2007; Crinion, Lambon-Ralph, Warburton, Howard, & Wise, 2003; Bottini et al., 1994).

Posterior left superior temporal areas, overlapping with regions in the inferior parietal lobe (supramarginal gyrus), have also been associated with phonological processing in the context of verbal working memory (e.g., Gitelman, Nobre, Sonty, Parrish, & Mesulam, 2005; D'Arcy, Ryner, Richter, Service, & Connolly, 2004; Gernsbacher & Kaschak, 2003). The pattern of activity we found in this area (ROI 6) might also reflect differences in the requirements of verbal working memory in our conditions. Although speculative, it seems possible that the significant differences in the posterior temporal lobe between unrelated and explicit conditions, as well as between inference and explicit conditions might be caused by the participants' attempts to rehearse the context sentences. Particularly, as the presentation of the context sentences was word-by-word, rehearsal might have been more necessary if the test statements differed from the surface and propositional representations of the context sentences.

Summary and Conclusions

We studied inference processing in an event-related fMRI experiment on the basis of theoretical assumptions derived from the text comprehension framework of Kintsch (1998) and Kintsch and van Dijk (1978). We compared the processing of context sentences and verification statements that contained the same situation model representation and differed in propositional and/or surface (explicit, paraphrase, and inference conditions) with unrelated sentences that differed in all levels of representation. Depending on the overlap of the different representations of context sentences and subsequently presented verification statements, differential activation patterns during the verification task could be observed. In summary, we have shown that the dmPFC is critically involved in the generation of inferences from the situation model. This is compatible with results from several lines of research that highlight the importance of this area for the allocation of prior knowledge. By contrast, some other imaging studies on language comprehension did not find significant activation in the dmPFC, but rather point to a special role of right hemisphere areas for comprehension at the sentence or discourse level. In these studies, participants typically read for comprehension, and they were not required to respond to an explicit task (e.g., Mason & Just, 2004). The results of our ROI analysis show that the right anterior MTG actually did contribute to the additional processing that was required to verify inference and unrelated statements, in comparison to the explicit statements. However, the critical contrast between inference and paraphrase condition was not significant in this ROI. Thus, our study provides some evidence for the involvement of the right anterior temporal lobe during inference processing, but it does not point to a pronounced function of this region in this task. Clearly, more research is needed to determine the exact conditions under which activities in the right hemisphere are found during language comprehension (see Ferstl, 2007, for a discussion of this topic). Overall, our results contribute to our understanding of the components of the extended language network, especially with respect to the functional role of the dmPFC.

Acknowledgments

We thank Evelyn Ferstl, Charles A. Perfetti, Mark Wheeler, Mark W. Greenlee, and two anonymous reviewers for their excellent comments on an earlier version of this manuscript. Experiment 1 was run by Carsten Griesel in partial fulfillment of his Bachelor thesis. Experiment 2 was conducted at the University of Oldenburg. This research was supported by a grant from the TransCoop-Program of the Alexander von Humboldt-Foundation to Franz Schmalhofer and Charles A. Perfetti.

Reprint requests should be sent to Uwe Friese or Franz Schmalhofer, University of Osnabrueck, Institute of Cognitive Science, D-49069 Osnabrueck, Germany, or via e-mail: uwe.friese@uni-osnabrueck.de; franz.schmalhofer@uni-osnabrueck.de.

REFERENCES

Addis, D. R., Wong, A. T., & Schacter, D. L. (2007). Remembering the past and imagining the future: Common and distinct neural substrates during event construction and elaboration. *Neuropsychologia, 45,* 1363–1377.

Bottini, G., Corcoran, R., Sterzi, R., Paulesu, E., Schenone, P., Scarpa, P., et al. (1994). The role of the right hemisphere in the interpretation of figurative aspects of language. A positron emission tomography activation study. *Brain, 117,* 1241–1253.

Brett, M., Anton, J.-L., Valabregue, R., & Poline, J.-B. (2002). *Region of interest analysis using an SPM toolbox*. Paper presented at the 8th International Conference on Functional Mapping of the Human Brain, June 2–6, 2002, Sendai, Japan. Available on CD-ROM in *Neuroimage, 16,* 497.

Buckner, R. L., & Carroll, D. C. (2007). Self-projection and the brain. *Trends in Cognitive Sciences, 11,* 49–57.

Cabeza, R., & Nyberg, L. (2000). Neural bases of learning and memory: Functional neuroimaging evidence. *Current Opinion in Neurology, 13,* 415–421.

Connell, L., & Keane, M. T. (2006). A model of plausibility. *Cognitive Science, 30,* 95–120.

Crinion, J. T., Lambon-Ralph, M. A., Warburton, E. A., Howard, D., & Wise, R. J. (2003). Temporal lobe regions engaged during normal speech comprehension. *Brain, 126,* 1193–1201.

D'Arcy, R. C., Ryner, L., Richter, W., Service, E., & Connolly, J. F. (2004). The fan effect in fMRI: Left hemisphere specialization in verbal working memory. *NeuroReport, 15,* 1851–1855.

Ferstl, E. C. (2007). The functional neuroanatomy of text comprehension: What's the story so far. In F. Schmalhofer & C. A. Perfetti (Eds.), *Higher level language processes in the brain: Inference and comprehension processes* (pp. 53–102). Mahwah, NJ: Erlbaum.

Ferstl, E. C., Neumann, J., Bogler, C., & von Cramon, Y. D. (2007). The extended language network: A meta-analysis of neuroimaging studies on text comprehension. *Human Brain Mapping, 29,* 581–593.

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. *Journal of Cognitive Neuroscience, 17,* 724–739.

Ferstl, E. C., & von Cramon, D. Y. (2001). The role of coherence and cohesion in text comprehension: An event-related fMRI study. *Cognitive Brain Research, 11,* 325–340.

Ferstl, E. C., & von Cramon, D. Y. (2002). What does the frontomedian cortex contribute to language processing: Coherence or theory of mind? *Neuroimage, 17,* 1599–1612.

Fletcher, P. C., Happe, F., Frith, U., & Baker, S. C. (1995). Other minds in the brain: A functional imaging study of ''theory of mind'' in story comprehension. *Cognition: International Journal of Cognitive Science, 57,* 109–128.

Forstmann, B. U., Brass, M., Koch, I., & von Cramon, D. Y. (2005). Internally generated and directly cued task sets: An investigation with fMRI. *Neuropsychologia, 43,* 943–952.

Frith, C. D., & Frith, U. (1999). Interacting minds—A biological basis. *Science, 286,* 1692–1695.

Gallagher, H. L., & Frith, C. D. (2003). Functional imaging of ''theory of mind''. *Trends in Cognitive Sciences, 7,* 77–83.

Gernsbacher, M. A., & Kaschak, M. P. (2003). Neuroimaging studies of language production and comprehension. *Annual Review of Psychology, 54,* 91–114.

Gilboa, A. (2004). Autobiographical and episodic memory—One and the same? Evidence from prefrontal activation in neuroimaging studies. *Neuropsychologia, 42,* 1336–1349.

Gitelman, D. R., Nobre, A. C., Sonty, S., Parrish, T. B., & Mesulam, M. M. (2005). Language network specializations: An analysis with parallel task designs and functional magnetic resonance imaging. *Neuroimage, 26,* 975–985.

Goel, V., Gold, B., Kapur, S., & Houle, S. (1997). The seats of reason? An imaging study of deductive and inductive reasoning. *NeuroReport, 8,* 1305–1310.

Graesser, A. C., Millis, K. K., & Zwaan, R. A. (1997). Discourse comprehension. *Annual Review of Psychology, 48,* 163–189.

Graesser, A. C., Singer, M., & Trabasso, T. (1994). Constructing inferences during narrative text comprehension. *Psychological Review, 101,* 371–395.

Graham, K. S., Lee, A. C., Brett, M., & Patterson, K. (2003). The neural basis of autobiographical and semantic memory: New evidence from three PET studies. *Cognitive, Affective & Behavioral Neuroscience, 3,* 234–254.

Hagoort, P. (2005). On Broca, brain, and binding: A new framework. *Trends in Cognitive Sciences, 9,* 416–423.

Jung-Beeman, M. (2005). Bilateral brain processes for comprehending natural language. *Trends in Cognitive Sciences, 9,* 512–518.

Keefe, D. E., & McDaniel, M. A. (1993). The time course and durability of predictive inferences. *Journal of Memory and Language, 32,* 446–463.

Kintsch, W. (1988). The role of knowledge in discourse comprehension: A construction-integration model. *Psychological Review, 95,* 163–182.

Kintsch, W. (1998). *Comprehension: A paradigm for cognition*. Cambridge, UK: Cambridge University Press.

Kintsch, W., & van Dijk, T. A. (1978). Toward a model of text comprehension and production. *Psychological Review, 85,* 363–394.

Kintsch, W., Welsch, D., Schmalhofer, F., & Zimny, S. (1990). Sentence memory: A theoretical analysis. *Journal of Memory and Language, 29,* 133–159.

Kuperberg, G. R., Holcomb, P. J., Sitnikova, T., Greve, D., Dale, A. M., & Caplan, D. (2003). Distinct patterns of neural modulation during the processing of conceptual and syntactic anomalies. *Journal of Cognitive Neuroscience, 15,* 272–293.

Kuperberg, G. R., Lakshmanan, B. M., Caplan, D. N., & Holcomb, P. J. (2006). Making sense of discourse: An fMRI study of causal inferencing across sentences. *Neuroimage, 33,* 343–361.

Lea, R. B., Mulligan, E. J., & Walton, J. L. (2005). Accessing distant premise information: how memory feeds reasoning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 31,* 387–395.

Li, C. S., Huang, C., Constable, R. T., & Sinha, R. (2006). Imaging response inhibition in a stop-signal task: Neural correlates independent of signal monitoring and post-response processing. *Journal of Neuroscience, 26,* 186–192.

Magliano, J. P., Radvansky, G. A., & Copeland, D. E. (2007). Beyond language comprehension: Situation models as a form of autobiographical memory. In F. Schmalhofer & C. A. Perfetti (Eds.), *Higher level language processes in the brain: Inference and comprehension processes*. Mahwah, NJ: Erlbaum.

Mason, R. A., & Just, M. A. (2004). How the brain processes causal inferences in text. *Psychological Science, 15,* 1–7.

Mazoyer, B. M., Tzourio, N., Frak, V., Syrota, A., Murayama, N., Levrier, O., et al. (1993). The cortical representation of speech. *Journal of Cognitive Neuroscience, 5,* 467–479.

McDaniel, M. A., Schmalhofer, F., & Keefe, D. E. (2001). What is minimal about predictive inferences? *Psychonomic Bulletin & Review, 8,* 840–846.

McKoon, G., & Ratcliff, R. (1986). Inferences about predictable events. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 12,* 82–91.

McKoon, G., & Ratcliff, R. (1992). Inference during reading. *Psychological Review, 99,* 440–466.

Potts, G. R., Keenan, J. M., & Golding, J. M. (1988). Assessing the occurrence of elaborative inferences: Lexical decision versus naming. *Journal of Memory and Language, 27,* 399–415.

Ruff, C. C., Knauff, M., Fangmeier, T., & Spreer, J. (2003). Reasoning and working memory: Common and distinct neuronal processes. *Neuropsychologia, 41,* 1241–1253.

Schacter, D. L., & Addis, D. R. (2007). The cognitive neuroscience of constructive memory: Remembering the past and imagining the future. *Philosophical Transactions of the Royal Society of London, Series B, Biological Sciences, 362,* 773–786.

Schmalhofer, F., & Glavanov, D. (1986). Three components of understanding a programmer's manual: Verbatim, propositional, and situational representations. *Journal of Memory and Language, 25,* 279–294.

Schmalhofer, F., McDaniel, M. A., & Keefe, D. (2002). A unified model for predictive and bridging inferences. *Discourse Processes, 33,* 105–132.

Singer, M. (2006). Verification of text ideas during reading. *Journal of Memory and Language, 54,* 574–591.

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.

Svoboda, E., McKinnon, M. C., & Levine, B. (2006). The functional neuroanatomy of autobiographical memory: A meta-analysis. *Neuropsychologia, 44,* 2189–2208.

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. *Proceedings of the National Academy of Sciences, U.S.A., 94,* 14792–14797.

Van Dijk, T. A., & Kintsch, W. (1983). *Strategies of discourse comprehension*. San Diego, CA: Academic Press.

Van Essen, D. C. (2002). Windows on the brain: The emerging role of atlases and databases in neuroscience. *Current Opinion in Neurobiology, 12,* 574–579.

Van Essen, D. C., Drury, H. A., Dickson, J., Harwell, J., Hanlon, D., & Anderson, C. H. (2001). An integrated software suite for surface-based analyses of cerebral cortex. *Journal of American Medical Informatics Association, 8,* 443–459.

Vandenberghe, R., Nobre, A. C., & Price, C. J. (2002). The response of left temporal cortex to sentences. *Journal of Cognitive Neuroscience, 14,* 550–560.

Virtue, S., Haberman, J., Clancy, Z., Parrish, T., & Jung Beeman, M. (2006). Neural activity of inferences during story comprehension. *Brain Research, 1084,* 104–114.

Wagner, A. D., Pare-Blagoev, E. J., Clark, J., & Poldrack, R. A. (2001). Recovering meaning: Left prefrontal cortex guides controlled semantic retrieval. *Neuron, 31,* 329–338.

Yang, C. L., Perfetti, C. A., & Schmalhofer, F. (2007). Event-related potential indicators of text integration across sentence boundaries. *Journal of Experimental Psychology: Learning, Memory and Cognition, 33,* 55–89.

Zwaan, R. A. (2004). The immersed experiencer: Toward an embodied theory of language comprehension. In B. H. Ross (Ed.), *The psychology of learning and motivation* (Vol. 44, pp. 35–62). New York: Academic Press.