

Research report

Event-related potentials during discourse-level semantic integration of complex pictures¹

W. Caroline West²

*Neuropsychology Laboratory and NMR Center, Department of Neurology,
Massachusetts General Hospital, Harvard Medical School*

Phillip J. Holcomb

Department of Psychology, Tufts University, Medford, MA

Abstract

This study examined event-related potentials (ERPs) elicited in response to semantic processing of nonverbal stories. ERPs were recorded from 29 scalp electrodes on 16 participants while they viewed series of complex gray-scale pictures, each of which relayed a simple story. The final picture of each story was either congruous or incongruous with the preceding context. Participants made delayed meaningfulness judgments for each story. Averaged ERPs time-locked to the onset of the final picture were more negative for incongruous than congruous pictures. Two distinct components were sensitive to congruency. The first component peaked at about 325 ms (N300) and was distributed over central and frontal sites. The second component peaked at about 500 ms and also had a centro-frontal maximum but was more widespread than the earlier component (anterior N400). The distinct scalp topographies of these two negativities provide strong evidence that the N300 and N400 are separate and distinguishable components. Furthermore, the presence of the N300 in this exclusively pictorial task suggest that the N300 is specific to the semantic processing of nonverbal stimuli and is not due to linguistic mediation. This study also revealed that the N400 can be modulated by discourse-level coherence manipulations with pictures. Finally, the different patterns of ERP effects observed during the semantic processing of verbal and nonverbal information suggest that non-identical neuronal sources, and thus non-identical representational systems, are accessed by these different types of materials. These findings strongly support at least partial modularity of semantic representations and processing mechanisms in the human brain.

Keywords: N300, N400, event-related potential, discourse, pictures, semantic representation

1. Introduction

The human brain is unique in that it has the ability to store and process conceptual representations in response to both direct sensory stimuli (e.g., pictures) as well as symbolic stimuli (written and spoken language). However, the degree to which the cognitive and neural semantic memory systems for sensory and symbolic stimuli overlap is as yet unknown. Two major viewpoints are currently under contention. Multiple-code theories of

conceptual representation (e.g., dual-coding theory) advocate qualitatively different and separate semantic systems for pictures and words (e.g., [34-36,39,40]). Conversely, common-code theories maintain that pictures and words differ only in their perceptual surface-level representations and are handled by the same common semantic system (e.g., [1,8,13,24,37]). Behavioral measures in normal subjects and in neuropsychological patients have produced a large body of data on this topic. However, most findings can be interpreted from

¹ This research was supported by NICHD grant HD25889 to P.H.

² W. Caroline West, Ph.D. E-mail: caroline@nmr.mgh.harvard.edu

either perspective, and thus no consensus has been reached. Fortunately, studies using modern brain imaging techniques are beginning to shed some light on this debate. The experiment presented here utilized event-related potentials (ERPs) recorded while subjects viewed congruous and incongruous picture stories to characterize the neural activity associated with discourse-level semantic processing of pictorial information.

Several previous studies have examined event-related potentials to picture stimuli in item-level semantic priming paradigms. In these studies, two distinct negative ERP components were observed to be sensitive to semantic manipulations. In a relatedness judgment task using pairs of line drawings of objects, Barrett and Rugg [4] observed a frontally distributed component that was more negative to unrelated than related pictures which peaked at 300 ms (N300). This was followed by a more widely distributed component that was also more negative to unrelated pictures and which peaked at 450 ms (N450). Similarly, Holcomb and McPherson observed both a frontally distributed N300 and a more widely distributed (except for the occipital sites) N400 semantic priming effect in an object decision task [18] and an object identification task [30]. In addition to the larger N400 effect for unrelated than related target pictures, non-objects (the picture equivalent of pseudowords) and unidentifiable real objects elicited an even larger N400 than unrelated objects. The time course of these N400 effects closely paralleled the N400 effects observed in analogous word tasks (e.g., [19]). However, the scalp distributions differed. The N400 relatedness effect for pictures was largest over the frontal midline site (Fz) rather than posterior sites and showed no difference between related and unrelated targets over occipital sites. Also in contrast to the N400 for words, the effect for pictures was larger over the left hemisphere than the right. These findings corroborate the hypothesis that words and pictures are activating at least partially non-overlapping semantic systems. Furthermore, while the picture N400 may be similar to the N400 observed in linguistic tasks, there has been no observable verbal correlate to the N300. The presence of this early component therefore supports a

distinct processing mechanism for pictures.

In another experiment using a relatedness judgment task, McPherson and Holcomb [30] observed that the N300 and N400 were distinguishable both by their scalp distributions and by their sensitivity to degree of relatedness. The frontally distributed N300 was larger for unrelated than related pictures, although it did not differ for moderately related and highly related pictures. The more widely distributed N400, on the other hand, showed a graded effect in that moderately related pictures had an amplitude in between highly related and unrelated pictures. This effect also had a posterior maximum that was similar to N400 effects in comparable word tasks (e.g., [16]). These differences in scalp distribution and relatedness sensitivity imply further that the N300 and N400 are distinguishable and separate components. This functional dissociation also suggests that the N300 is not simply an earlier manifestation of a subcomponent of the N400.

ERPs have also been examined to pictures in sentence contexts [10,11,31] using the anomalous sentence task in which the N400 was originally observed [26,27]. In these studies, the final word of each sentence was replaced with a line drawing of the corresponding object. Both anomalous final pictures and anomalous final words generated a larger N400 than congruous final pictures and words. Also, the time courses of the effects were similar for both pictures and words. However, in the studies that utilized electrode sites covering the entire scalp [10,11], the effects for pictures displayed a different scalp distribution than the effects for words. Specifically, at occipital and parietal sites the N400 congruity effects were larger for words than for pictures. Conversely, at frontal sites the N400 congruity effects were larger for pictures than for words. These results suggest that the N400 reflects a semantic processing mechanism that is functionally similar for pictures and words. However, the brain regions responsible for the storage and processing of semantic representations for pictures and words may be partially nonoverlapping, resulting in slightly different scalp distributions [10,25].

While the N400 for pictures generally has a more anterior spatial distribution than that

for words, the scalp distributions associated with pictures and words are not entirely consistent across studies. For example, pictures presented at the ends of sentences in one study produced an N400 with substantial activity at anterior sites and virtually no activity over posterior sites [11], and in another similar study produced an N400 that was broadly distributed over both anterior and posterior sites, including occipital leads [10]. The N400 for words typically has a centro-parietal maximum, but has occasionally been found to have a more anterior distribution. Of particular interest, *concrete* words generally elicit a more anterior N400 than *abstract* words [17,23,46]. Because of the inconsistency of the N400 distribution across studies, caution must be exercised in comparing the scalp distributions of the different stimulus types across studies. Bearing that in mind, the similar anterior focus of the N400s for concrete words and for pictures is intriguing. It suggests that concrete words and pictures may activate similar semantic processing mechanisms. The anterior distribution of the N400 for concrete words and pictures may be due to activation of a mechanism that processes image-based information.

Taken together, the functional and temporal similarities but spatial variability of the N400 suggests that the N400 may involve contributions from multiple brain areas that perform analogous cognitive operations on different types of input. The hypothesis that the N400 has multiple underlying subcomponents is supported by findings from intracranial recordings that show N400-like modulations in multiple brain areas, including medial and lateral temporal lobe, hippocampus, and ventrolateral prefrontal cortex [9,14,15,28,32,42]. The scalp-recorded N400 may be the result of coordinated activity in a large network of brain areas, with different types of input (e.g., pictorial vs. symbolic, concrete vs. abstract) and task demands or processing strategies (e.g., imagistic vs. linguistic) inducing preferential activity in different subsets of regions. The combined contribution of these weighted sources would give rise to N400s with varying scalp distributions in different task situations. The N400 may, in fact, reflect an amodal semantic process in which information from a variety of

sources is integrated into a higher-level conceptual representation [25].

On the other hand, the N300 may reflect a semantic process specific to object representations. As described above, no ERP component with latency and scalp distribution similar to the N300 has been reported in any language studies. Furthermore, there is some evidence that the N300 has different functionality than the N400 [30]. The presence of a distinct ERP component that is sensitive to semantic manipulation of picture stimuli is evidence for a modality-specific semantic mechanism. Another possibility is that the N300 reflects a mechanism whose function is to translate imagistic representations into linguistic representations. All of the ERP studies described above used line drawings or photographs of simple objects as stimuli. These objects are easily labeled and it is likely that subjects covertly named each object as they performed the tasks. It is unknown whether an N300 would be elicited to pictures in a situation that discourages covert naming.

The main goal of the current study was to characterize the neural activity associated with nonverbal semantic processing. We examined the ERP effects elicited during semantic processing of pictures in a task designed to limit linguistic mediation. Our chosen task utilized sequences of pictures that depicted simple stories in a discourse-level version of the anomalous sentence task. In this task, the final picture of each story was either congruous or incongruous (anomalous) with the preceding context and the subjects' task was to decide if each story made sense (meaningfulness judgment). The anomalous sentence tasks previously performed with picture stimuli [10,11,31] used a "rebus" format. In those studies, the verbal sentence context as well as the simple object line drawings likely encouraged picture naming. To prevent a covert object naming strategy and to encourage the exclusive use of nonverbal semantic processing mechanisms, the stimuli in the current study were composed entirely of pictures depicting complex scenes. These scenes involved various numbers of characters and/or objects and represented actions and interactions between and among them. These stimulus materials should

not encourage covert naming since no single word could be used to adequately label their content. This “anomalous story task” allowed comparison with previous ERP studies of contextual congruity as well as examination of more naturalistic and exclusive picture processing. We predicted that complex pictorial scenes processed with minimal linguistic mediation would elicit (1) a frontally distributed N300 congruity effect and (2) an N400 congruity effect with a scalp distribution similar to the anterior N400 for concrete words. The presence of an N300 effect in this study would (1) indicate that the N300 that has been observed in previous studies is not due to linguistic mediation (naming) and (2) provide further evidence that the N300 reflects activation of an object-specific semantic mechanism.

The story congruity manipulation also allowed us to examine discourse-level comprehension of pictures similar to the kind of comprehension that occurs when we read or listen to a narrative story. At this level, semantic information is integrated into the overall theme or “macrostructure” of a discourse and involves what is known as *global coherence* [21]. The N400 has been shown previously to be sensitive to global coherence using verbal stories with congruous or incongruous final sentences [45] and ambiguous paragraphs presented with or without titles that are essential for comprehension [44]. The N400 coherence effects observed for verbal stories were remarkably similar to the N400 effects observed for words in sentences (*local coherence*) in spatial distribution, latency and duration. This suggests that the N400 reflects a general mechanism for semantic integration that is not specific to the level of processing (i.e., word, sentence or discourse). It is unclear whether these levels of processing in the linguistic domain can be straightforwardly mapped onto the picture domain. It seems reasonable to equate single words and objects. However, this mapping becomes more uncertain at higher levels. The pictures used in the current experiment often contain more information than a typical sentence, but less than a complete discourse. It is furthermore unclear whether the N400 should be sensitive to discourse-level semantic processing of picture as it is for

discourse-level processing of words or whether this response should be similar to that for picture processing at the item and sentence levels.

While there have been no ERP or functional neuroimaging studies to date that have examined discourse-level processing with pictures, data from a few behavioral studies suggest that processing of verbal and nonverbal stories may involve similar mechanisms. For example, subjects were able to make inferences after viewing sequences of line drawings that depicted simple everyday events [2]. Also, subjects’ recall was similar for stories they had viewed as movies without dialogue or had read as text [3]. Furthermore, Gernsbacher [13] found that surface (e.g., left/right orientation) information is lost during comprehension of picture stories in a manner similar to the surface (e.g., word order) information loss that occurs for sentences. Gernsbacher points out that verbal and nonverbal narratives share several basic components: a setting, a main character or group of characters, and a plot or event sequence [13]. It follows that integration of semantic information and establishment of global coherence might proceed in a similar fashion for the two modalities. Thus, we expected that, like text stories, picture stories would show ERP global coherence effects, but that these effects would have spatial and temporal characteristics similar to the local coherence effects for pictures at the item and sentence level. In other words, discourse-level semantic processing of picture stimuli will be functionally similar to that for words, but will engage a network of brain regions that is nonidentical to the network of regions engaged for words.

2. Materials and Methods

2.1 Participants

Sixteen (11 female, 5 male) right-handed undergraduates from Tufts University aged 18 to 21 (mean = 18.5) participated in the experimental portion of this study. Selection criteria required all participants to have normal or corrected-to-normal vision and be native speakers of English. Four participants had at least one left-handed relative in their immediate families. Written informed consent was obtained from all participants prior to the start of the experiment.

2.2 Stimulus Materials

Stimulus materials were extracted from commercially available animated full color videos using the Snappy video snapshot software (Play, Inc., Rancho Cordova, CA). Sequences of pictures (four to ten frames in length) were constructed, each of which relayed a simple story (Fig. 1). All frames were transformed to 8-bit gray-scale images to eliminate palette differences across sequences. Each frame consisted of a complex scene and was 4 by 5 1/2 inches in size (subtending approximately five degrees of visual angle) and centered on a black background. The final frame of each story was distinguished as such by a white border (providing a closure cue similar to a period at the end of a sentence).

Two initial sets of picture stories were created. First, stories were constructed in which the final frame of each sequence was congruous with the preceding context. These stories were assessed by 10 participants for general comprehensibility on a 7-point scale. Only stories with a rating of 4.6 or higher (mean = 6.03 ± 0.60) were used in the final set of 80 congruous stories. A second set was created by taking the congruous sequences and pseudo-randomly mixing the final frames (Fig. 1) to produce stories with incongruous endings. The number of characters and their gender in the incongruous final frames were matched to those in the corresponding congruous final frames.

Finally, two final lists of stories were formed. List A was composed of 40 congruous stories and 40 incongruous stories. There was no repetition of either story stems or final frames and the order of the stories was random. List B was composed of the remaining 40 congruous and 40 incongruous stories. Thus, stimuli were counterbalanced such that a story with a congruous final frame in List A appeared in List B with an incongruous final frame and a story with an incongruous final frame in List A appeared with a congruous final frame in List B. In short, every story was presented in both congruous and incongruous versions. However, each subject saw only one version of each story. Ten naïve participants rated the congruity of the final frame of each story on a 7-point scale. Raters viewed either List A or List B on a computer monitor in a manner identical to that

in which stimuli were later viewed by participants in the experimental portion of the study (see below). Congruous stories had an average congruity rating of 6.23 ± 0.52 . Incongruous stories had an average congruity rating of 1.81 ± 0.90 .



Fig.1. Example of congruous and incongruous picture stories. The center column shows the series of pictures in a typical picture story. The bottom left picture shows the congruous final picture for this story and the bottom right frame shows the corresponding incongruous final picture for this story.

2.3 Procedure

Experimental subjects were randomly

assigned to List A or List B. Subjects sat in a comfortable chair in a dimly lighted room separate from the experimenter and computers. Picture stories were presented to each subject on a computer monitor. Each trial (one story) began with the presentation of a fixation cross at the center of the screen for 1500 ms. After a 300 ms ISI the first frame of the story appeared. The duration of each frame was 1500 ms with an ISI of 300 ms (these stimulus presentation parameters were empirically determined to be optimal with test subjects). The final frame was followed by a 500 ms ISI at which time the words "RESPOND NOW" appeared. This cue remained on the screen until the subject made her response. The subject's task was to decide whether the story did or did not make sense by pushing one of two buttons on a response box with either her left or right thumb. "Yes" and "no" response hands were counterbalanced across subjects. After the subject made her response, the words "PRESS FOR NEXT TRIAL" appeared on the screen, thus, the experiment was self-paced. Each subject was given 12 practice trials at the start of the experiment.

2.4 Electrophysiological Recording

Twenty-nine active tin electrodes were held in place on the scalp (Fig. 2) by an elastic cap (Electro-Cap International, Inc., Eaton, OH). Several of these electrodes were placed in standard International 10-20 System locations. These included five sites along the midline (Fpz, Fz, Cz, Pz, and Oz) and eight lateral (four over each hemisphere) sites (F3/F4, C3/C4, T3/T4, and P3/P4). Eight extended 10-20 system sites were also used (FC1/FC2, FC5/FC6, CP1/CP2, and CP5/CP6). Eight additional 10-20 sites were altered to form a circle around the perimeter of the scalp. These altered sites included Fp1'/Fp2' (33% of the distance along the circle between Fpz and T3/T4), F7'/F8' (67% of the distance between Fpz and T3/T4), T5'/T6' (33% of the distance between T3/T4 and Oz), and O1'/O2' (67% of the distance between T3/T4 and Oz). Electrodes were also placed below the left eye and beside the right eye to monitor vertical and horizontal eye movements. All active electrodes were referenced to an electrode placed on the left

mastoid. The right mastoid was recorded to detect any activity at or asymmetry between the mastoids.

The EEG signal was amplified by an Isolated Bioelectric Amplifier System Model H&W-32/BA (SA Instrumentation Co., San Diego, CA) with a bandpass of 0.01 to 40 Hz (3dB cutoffs) and was continuously sampled at 200 Hz by an analogue-to-digital converter. The stimuli presented to each subject and the subject's behavioral responses were simultaneously monitored by the digitizing computer.

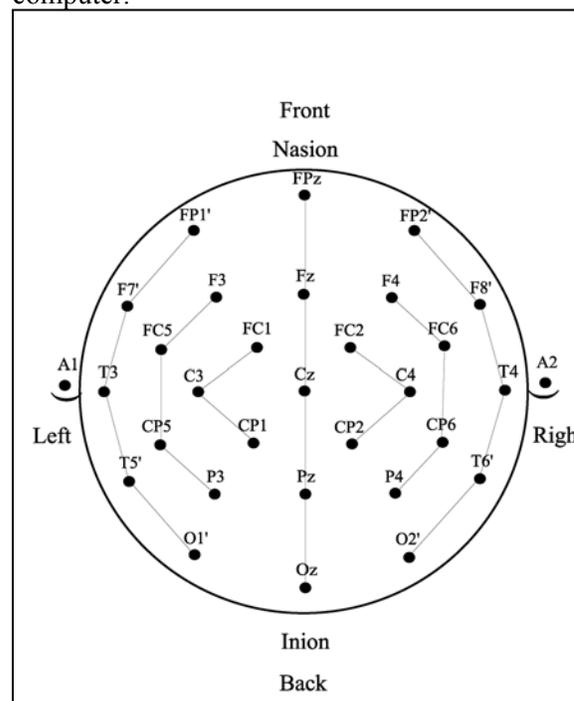


Fig. 2. Montage of electrode placement on the scalp. A1 is the left mastoid reference electrode. A2 is the right mastoid recorded site. Also shown are the five parasagittal columns used for the data analysis. The midline analysis included the column along the central axis. The medial analysis included the two columns directly to the left and right of the central column. The lateral analysis included the next two left and right columns, and the peripheral analysis included the final two left and right columns.

2.5 Data Analysis

Averaged ERPs were formed off-line from trials free of ocular and muscular artifact. All ERPs were time-locked to the onset of the final frame of each story. The averaged ERPs were quantified by calculating the mean amplitude values (relative to a 100 ms prestimulus baseline) for the voltage points in three time windows (300-400 ms, 400-600 ms,

and 600-900 ms after stimulus onset.). The resulting data for each time window were analyzed with analyses of variance (ANOVAs) for repeated measures having two levels of *congruity* (congruous and incongruous). Four separate analyses were performed in order to examine parasagittal columns of scalp electrodes along the anterior-posterior axis of the head (Figure 2). The **midline** analysis had five levels of *electrode site* (FPz, Fz, Cz, Pz, Oz). The **medial** analysis had three levels of *electrode site* and two levels of *hemisphere* (FC1/FC2, C3/C4, P1/CP2). The **lateral** analysis had four levels of *electrode site* and two levels of *hemisphere* (F3/F4, FC5/FC6, CP5/CP6, P3/P4). The **peripheral** analysis had five levels of *electrode site* and two levels of *hemisphere* (FP1'/FP2', F7'/F8', T3/T4, T5'/T6', O1'/O2'). The Geisser-Greenhouse correction [12] was applied to all repeated measures with more than one degree of freedom. ANOVAs were also performed after a z-score transformation to normalize the data [29]. Only electrode site interactions significant after normalization are reported.

3. Results

3.1 Behavioral Data

Subjects responded quite accurately to the meaningfulness of the picture stories. For congruous stories the accuracy rate was $96.3\% \pm 5.6$. For incongruous stories the accuracy rate was $93.1\% \pm 5.7$. The difference in accuracy between congruous and incongruous stories did not reach conventional significance levels [$t(15) = 1.559, p = .140$]

3.2 Event-Related Potential Data

The averaged ERPs time-locked to congruous and incongruous final pictures are shown in Figure 3. Waveforms for congruous and incongruous final pictures were equivalent from stimulus onset until approximately 275 to 300 ms. At which time they became more negative-going for incongruous final pictures than for congruous final pictures. The ERPs elicited by congruous final pictures displayed a small negativity which peaked at about 275 ms and then a large extended positivity (P3). The ERPs elicited by incongruous final pictures

displayed two separate negative components. The first negativity was similar to that observed for congruous pictures but had a longer peak latency (325 ms) and larger amplitude. The second negativity peaked at about 500 ms. This component was followed by a long slow positivity (P3) similar to that observed for congruous final pictures.

Visualization of the congruity effect is aided by examination of the difference waves (Figure 4) and voltage maps (Figure 5) produced when ERPs for congruous final pictures are subtracted from ERPs for incongruous final pictures. The amplitude difference for ERPs elicited by congruous and incongruous final pictures was largest at central and frontal scalp sites (focused around Fz and Cz) for both the earlier (N300) and later (N400) negativities. At the most anterior sites (FP1, FPz, and FP2) and at posterior sites there was little difference. The divergence was also longer lasting at more anterior sites and appeared to be slightly larger over the right hemisphere than the left hemisphere.

3.2.1 300-400 ms epoch

During this epoch, incongruous pictures elicited more negative ERPs than congruous pictures at medial [main effect of congruity: $F(1,15) = 9.651, p < .01$] and lateral sites [main effect of congruity: $F(1,15) = 6.871, p < .05$]. Planned comparisons revealed that this congruity effect was significant at all medial and lateral sites except the posterior lateral sites P3 and P4. At midline sites, the main effect of congruity only approached conventional significance levels [$F(1,15) = 3.400, p = .085$]. However, ERPs were more negative for incongruous than congruous pictures at anterior-central midline sites [congruity by electrode site interaction: $F(4,60) = 4.792, p < .05$]. Planned comparisons revealed that a significant effect was present at Fz and Cz, but not at the most frontal site (FPz) or at posterior sites (Pz and Oz). Furthermore, at medial sites the congruity effect was largest over the anterior right hemisphere [congruity by electrode site by hemisphere interaction: $F(2,30) = 3.531, p < .05$].

Fig. 3. Grand average ERPs elicited by congruous and incongruous final pictures.

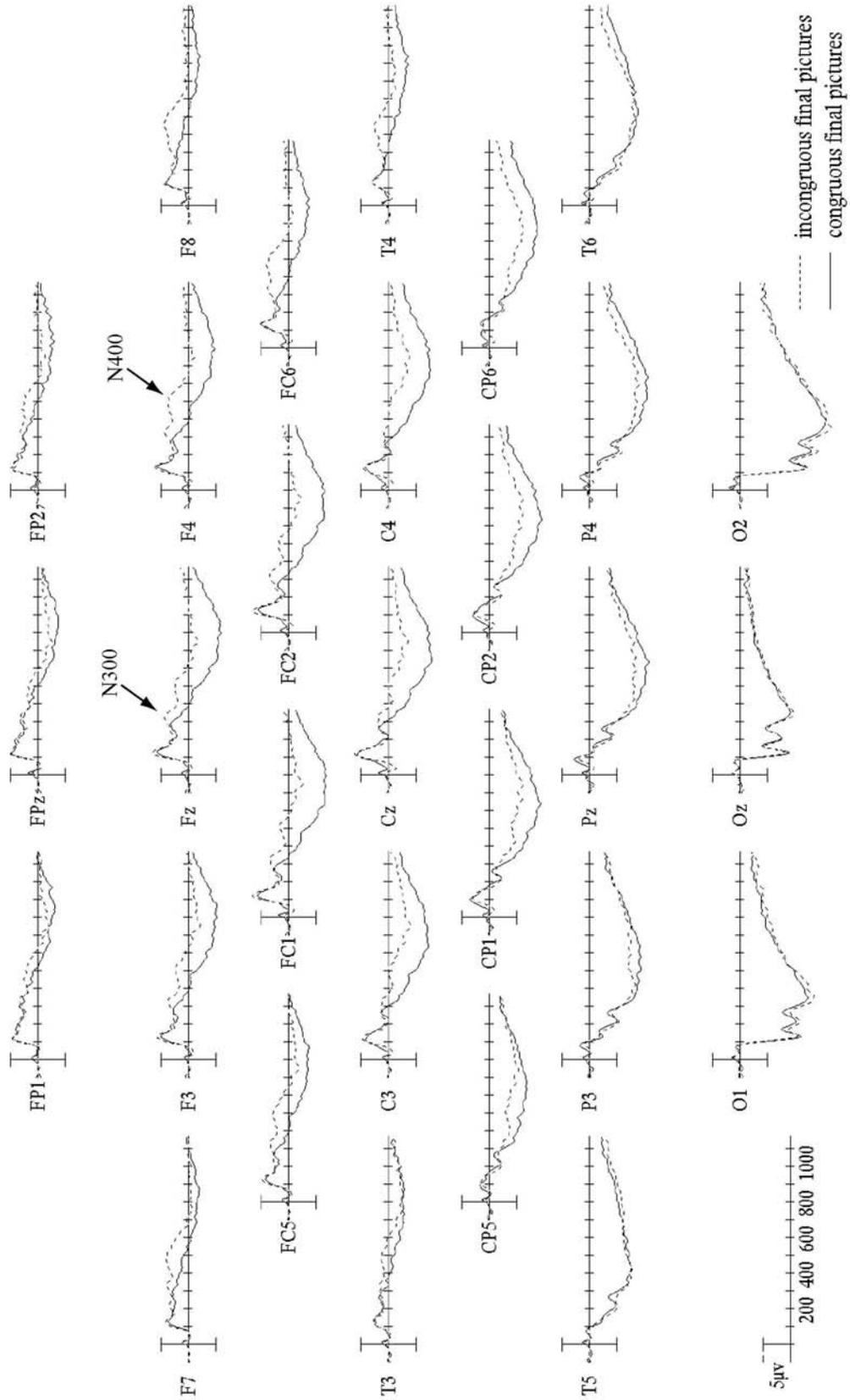
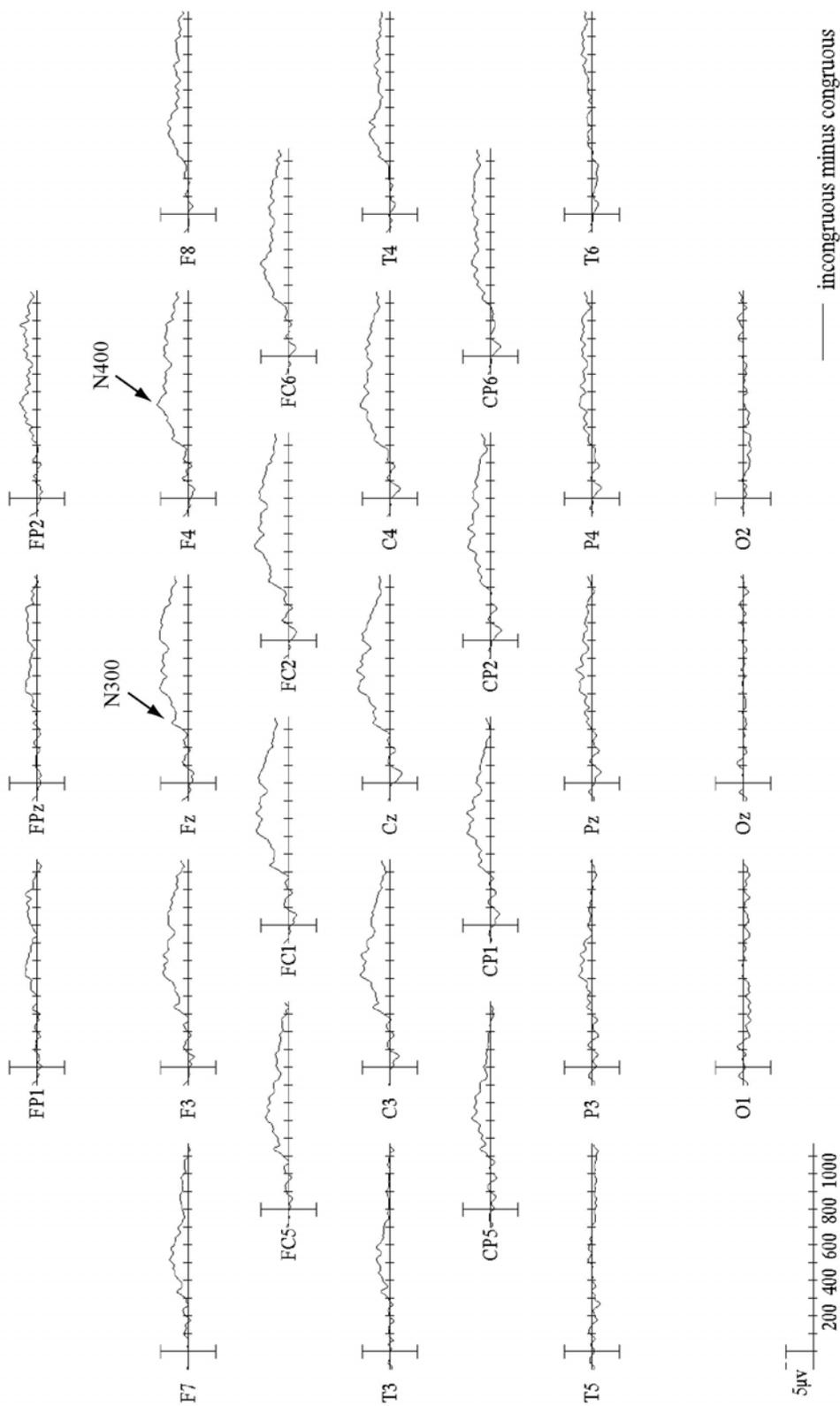


Fig. 4. Difference waves produced by subtracting the ERPs for congruous final pictures from the ERPs for incongruous final pictures.



3.2.2 400-600 ms epoch

During this epoch, incongruous pictures generally elicited more negative waveforms than congruous pictures across the scalp [main effect of congruity, midline: $F(1,15) = 11.272$, $p < .01$; medial: $F(1,15) = 22.531$, $p < .001$; lateral: $F(1,15) = 20.271$, $p < .001$; peripheral: $F(1,15) = 12.031$, $p < .01$]. At midline sites, this effect was larger at central sites [congruity by electrode site interaction: $F(4,60) = 6.379$, $p < .01$]. Planned comparisons revealed a pattern similar to that observed in the 300 to 400 ms time window. Sites Fz and Cz displayed a significant effect of congruity. In addition, the effect at Pz approached significance. At medial, lateral, and peripheral sites, planned contrasts revealed a significant effect of congruity at all sites except the most posterior sites (T5/T6 and O1/O2). Furthermore, at lateral sites the congruity effect was larger over the right hemisphere than over the left hemisphere [congruity by hemisphere interaction: $F(1,15) = 5.181$, $p < .05$].

3.2.3 600-900 ms epoch

Incongruous pictures continued to elicit more negative waveforms than congruous pictures [main effect of congruity, midline: $F(1,15) = 17.776$, $p < .01$; medial: $F(1,15) = 29.197$, $p < .001$; lateral: $F(1,15) = 17.241$, $p < .01$; peripheral: $F(1,15) = 5.532$, $p < .05$]. Planned comparisons revealed marginally significant effects at FP1/FP2 but no effect at FPz. Also, there were no significant effects at the most posterior regions (Pz, P3/P4, T5/T6, Oz, O1/O2). In addition, the congruity effect was largest at fronto-central locations in both midline and medial site analyses [congruity by electrode site interaction, midline: $F(4,60) = 9.701$, $p < .01$; medial: $F(2,30) = 4.910$, $p < .05$].

4. Discussion

This study was designed to address three major issues. The first aim was to enhance the characterization of the ERP components that are sensitive to the semantic processing of pictures. Second, the experiment was designed to determine whether and in what way discourse-level semantic processing influences ERPs for

complex pictures. Finally, by comparing ERPs in this task with previous linguistic tasks, this study addressed the question of whether there is a single common system or multiple modality-specific systems in the brain for storing and manipulating semantic representations.

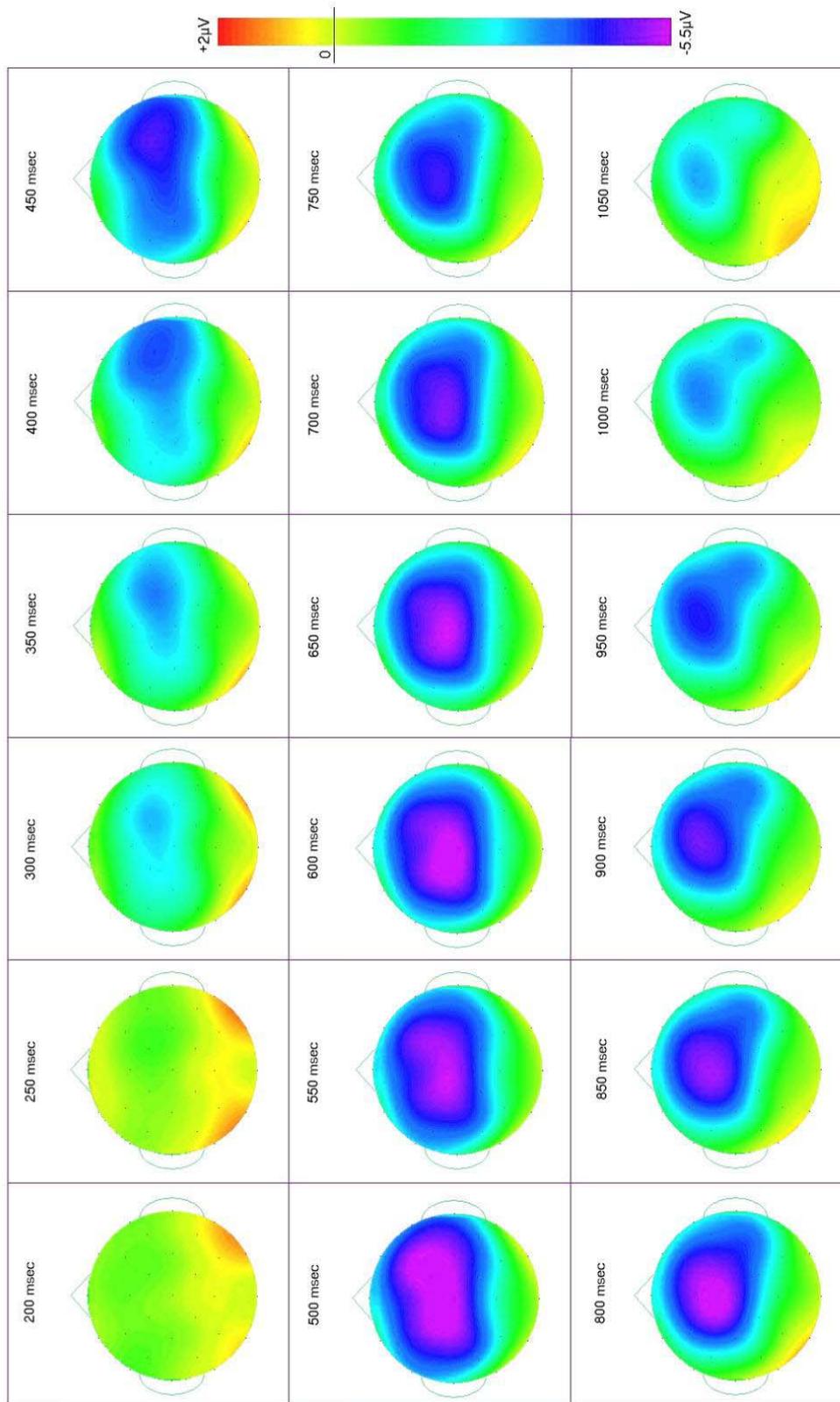
Event-related potentials recorded while subjects viewed picture stories revealed that the ERP is sensitive to the global coherence of a picture with prior context. In this anomalous picture story task, both a frontally distributed N300 congruity effect and a more widespread N400 congruity effect were observed. By using stimuli designed to discourage covert naming, this study demonstrated that the anterior ERP effects for pictures are not due to linguistic mediation but rather reflect distinct mechanisms for semantic processing of pictorial information. This study also demonstrated that both the N300 and N400 ERP components are sensitive to discourse-level conceptual integration. The discourse-level picture congruity effects were similar in spatial and temporal characteristics to picture effects at the item and sentence levels, suggesting the involvement of similar semantic mechanisms for pictorial stimuli at various levels of processing. However, the spatiotemporal pattern of the nonverbal congruity effect differed from the typical verbal effect, suggesting that non-identical neural mechanisms are responsible for the semantic analysis of symbolic and direct sensory information.

4.1 Characterization of ERPs Elicited by Pictures

A final picture that was incongruous with the context established in the preceding series of pictures elicited a more negative waveform than a final picture that was congruous with the preceding context. Waveforms for congruous and incongruous pictures diverged between 250 and 300 ms. This divergence continued until the end of the epoch (1180 ms) and appeared to contain two distinct peaks that were more negative for incongruous pictures.

Between 300 and 400 ms, a congruity effect was evident in a negative component which peaked at about 325 ms after stimulus onset.

Fig. 5. Voltage map distributions for the difference waves (created with the EMSE Data Editor software Source Signal Imaging),



The effect in this time window was focused around central and frontal midline scalp sites. There was little difference at this latency at the most lateral sites and no difference at the most posterior sites. Furthermore, at medial fronto-central sites where the effect size was at its greatest, there was a hemispheric asymmetry. In this region, the congruity effect was larger over the right hemisphere than over the left.

Between 400 and 600 ms, a second negativity peaking at about 500 ms after stimulus onset was also sensitive to semantic congruity. The magnitude of this effect was greater than for the earlier component. The congruity effect during this time window also had a more widespread spatial distribution. While the effect was still focused around fronto-central regions, it was not restricted to the midline, but extended to the most lateral frontal and central sites. There was also a small congruity effect at parietal sites although there was no effect at posterior temporal and occipital sites. Furthermore, the right hemisphere maximum that was observed for the earlier negativity shifted to more lateral regions for this later negativity. After 600 ms the effect began to taper off at parietal sites, while still displaying a large effect at anterior regions until about 900 ms.

To summarize, the negativities between 300 and 400 ms and between 400 and 600 ms had somewhat different distributions over the scalp. The earlier epoch was focused over frontal and central regions at the midline and was larger over right than left medial sites. Potentials in the later epoch had a more widespread distribution, extending into parietal and particularly anterior lateral sites. Also, the right hemisphere bias during the later epoch shifted to more lateral sites than in the earlier epoch. The different scalp distributions for the effects of these two negativities suggest that they are subserved by at least partially non-overlapping sources. The general temporal and spatial patterns of the semantic congruity effects in this experiment were similar to the N300 and N400 effects observed in previous ERP studies with pictures [4,11,18,30]. The negativities in the earlier and later epochs observed here will be referred to as N300 and

N400 respectively for simplicity in comparison with the previous literature. There were several important differences in both the time course and the scalp distributions of the effects in this study and the previous studies. First, both negativities observed in this experiment had longer onset and peak latencies than in any of the previous picture studies. In the previous studies, congruity effects onset at about 200-250 ms with the N300 peaking at about 250-300 ms and the N400 peaking at about 400-450 ms, while in the current study the congruity effect onset about 275-300 ms with the N300 peaking at about 325 ms and the N400 peaking at about 500 ms. The similar anterior spatial distribution but longer latencies of the congruity effects in the picture story paradigm compared to the previous picture studies suggests that the anomalous story task engages similar mechanisms but may require more effortful or complex processing.

Also, the N400 effect in this study had a longer duration than in typical word tasks. The N400 effect may be temporally smeared because there is a lot of information, and perhaps more temporal jitter, associated with these stimuli. Such extended effects have been observed in spoken language (e.g., [19]) which produces ERP effects that do not have distinct peaks and are long-lasting. But they have also been observed in other picture studies (e.g., [4,11,18,30]). This fits the hypothesis that more complex stimuli produce more long-lasting "smeared" effects.

The longer latency and duration of the N300 and N400 congruity effects for picture stories could be due to several factors. First, they may be due to the increased processing load required for maintaining and integrating information into a discourse-level representation compared to that required for matching individual items or integrating information into a sentence context. It is known that N400 *amplitude* decreases systematically with the degree of semantic relatedness of recently presented items and with the item's predictability in the local context [25]. Perhaps the level of processing affects N400 *latency and duration*. This would be consistent with the hypothesis that the N400 is sensitive to the

difficulty involved in integrating information from lower-level sources (e.g., lexical, semantic, syntactic) into a higher-level representation [16,33,38].

Second, increased latencies and duration could be due to the greater complexity of the individual pictures, which depicted complex scenes as opposed to simple objects. These stimuli were more complex both in their visual perceptual characteristics and in their semantic content than the picture stimuli used in previous studies. While the factors controlling modulation of the N300 have only begun to be explored, it has been demonstrated that N400 latency can be modulated by perceptual-level processing. For example, when written words are degraded such that they are difficult to encode perceptually, N400 latency is shifted [16]. Thus, the longer latencies could be due to the increased visual complexity of the stimuli. These pictures were also more semantically complex than previous picture stimuli. They involved complex scenes that relayed information equal to or greater than the information that would be derived from a typical sentence. These picture stimuli were also more semantically complex than the stimuli examined in previous verbal discourse studies. For example, in the study by van Berkum, Brown, and Hagoort [45] the N400 global coherence (congruity across sentences) effect, had a latency and duration similar to the N400 local coherence (congruity within a sentence) effect. In that study, however, the averaged waveforms were time-locked to a single critical word in the final sentence and not to the sentence as a whole. Thus, the longer latency and duration of the N400 in the current study may have been due to the increased semantic complexity of the final pictures and consequently more extensive semantic evaluation. In short, either visual complexity or conceptual complexity of the individual stimuli could have influenced N400 and N300 latency and duration.

4.2 Effects of Global Coherence

The congruity effect in the present study also differed from the previous results in that anomalous picture stories elicited larger ERPs over the right hemisphere than over the left hemisphere. While the N400 for written words

presented in lists and in sentences has often been found to have a larger amplitude over the right hemisphere (e.g., [26,27]), the only other study using pictorial stimuli that has shown a right hemisphere bias was the anomalous sentence task [11] in which the final word of each sentence was replaced by a picture. This distributional difference implies the contribution of a somewhat different set of neural generators in the sentence and story tasks than in the picture-pair tasks.

Integration of semantic information into an ongoing discourse representation may recruit additional resources, perhaps located in the right hemisphere. The right hemisphere has been implicated to play an important role in comprehension of discourse-level information. For example, patients with right hemisphere damage have been shown to have difficulties making certain types of inferences [7] and appreciating the punch lines of jokes [5,6]. Using functional magnetic resonance imaging, St. George and colleagues found increased activation in the right hemisphere for untitled than titled paragraphs, when the paragraphs were difficult to understand without the title [43]. These data suggest that the right hemisphere is crucial for the ability to integrate semantically or temporally distant concepts and to establish the global coherence of a narrative. It is impossible to identify the specific brain regions underlying scalp-recorded ERPs in the current study. However, this experiment does demonstrate that the N400 is sensitive to the global coherence of a pictorial story.

In the current study, however, it is not possible to definitively identify the specific aspect of coherence information responsible for the ERP congruity effects. Since this was the first ERP study to examine discourse-level semantic processing with pictures, we intentionally used frames with incongruous characters, objects, and scenes for the incongruous endings. This was done to maximize the effect of semantic congruity. Character, object, and scene, as well as action, are all integral parts of the discourse structure. Changes to any or all of these factors should affect integration at the global level. The relative importance of each of these different aspects of semantic information to the global coherence of

a picture story is at present unclear. To dissociate these factors, we have begun a series of studies which will manipulate these factors individually, while keeping all others constant.

In the current study, we did, however, attempt to minimize effects due to syntactic violations, by ensuring that the gender and number of characters in each incongruous final frame were consistent with the corresponding congruous final frame. We are confident that the observed effects, in fact, were *not* due to syntactic violations since (1) we found an increased N400, which is not known to be modulated by syntactic manipulations, and (2) we *did not* find an increased P600, which is known to be modulated by syntactic manipulations.

4.3 Common versus Multiple Semantic Systems

The N400 congruity effect for picture stories also had a more anterior scalp distribution than the N400 congruity effect typically observed for words. Although, as mentioned earlier, in some instances words produce N400 effects which are maximal at anterior scalp sites. Most notably, ERP differences between concrete and abstract words are largest over anterior regions [17,23,46]. The topography of the concreteness N400 effects is very similar to the topography of the picture N400 effect observed in the current study. For both concrete words and pictures the N400 effect has a widespread scalp distribution but has the largest amplitude at anterior scalp sites. Furthermore, in some studies the N400 concreteness effect has exhibited a long-lasting duration similar to that for complex pictures [46]. The similar spatial distributions of these two effects suggest that they may be subserved by identical or closely located sources. The similar time-course and sensitivity to experimental manipulations suggest that they involve similar cognitive functions. These similarities between the N400s for concrete words and for pictures are strong evidence that anterior N400 effects reflect activation in a brain region that is involved in the semantic integration of *image-mediated* information. This integration process is amodal, in that it handles both words and pictures. Moreover, the different N400 scalp distributions for pictures

and image-mediated words (anterior focus) and for non-image-mediated words (posterior focus), but similar N400 functionality, suggest a unitary but distributed cognitive process having multiple neural generators that respond to varying degrees to different types of information [25].

The earlier N300 observed for picture stories may also have a similarity to another concreteness ERP effect. In an explicit imagery task [46], concrete words displayed a larger frontal negativity between 550 and 800 ms (N700) than abstract words in addition to a widespread N400. West and Holcomb (2000) proposed that the N400 concreteness effect reflects activation of a semantic process that is common to concrete and abstract words, while the N700 concreteness effect reflects activation of a mental imagery process that is only available to imageable words. It is possible that the N300 for pictures reflects a similar imagery process as the N700 for concrete words. The N300 and N700 may both reflect activation of image-mediated representations or processes. This would, in fact, be consistent with the dual coding assumption [35] that image-mediated processing occurs faster for pictures than for words since pictures have direct access to imagistic representations while concrete words must first access a lexical representation before making a referential connection to the image-based system. In these ERP studies it is impossible to determine the precise loci of the generators underlying the N300 or N700 effects. The frontal scalp distributions of these components suggest that the sources reside in anterior brain regions, although this may not necessarily be the case.

Recent functional imaging studies support the hypothesis that certain mental imagery processes may take place in the frontal lobes. Activation of the prefrontal cortex during mental imagery has been observed with PET [22] and fMRI [20]. One framework for the role of the prefrontal cortex suggests that it may be involved in the organization and manipulation of information that is stored elsewhere in the brain [41]. The frontal functional imaging and ERP effects may reflect a “top-down” mechanism mediating the retrieval of imagistic representations in more posterior brain regions.

In summary, the earlier N300 picture effects may reflect activation of *image-based representations* while the later anterior N400 effects reflect activation of an amodal system which integrates both *image-based* and *conceptual representations* into a context. These data are supportive of an *extended dual-coding hypothesis* [17,23,46] in which there is a common amodal system or systems available to both pictures and words in addition to form-specific (verbal/nonverbal) systems for semantic representation and processing. As mentioned earlier however, it is difficult to make strong inferences based on comparisons of scalp distributions in different studies. While the present study is suggestive of these multiple systems, studies are needed which directly compare discourse-level processing of pictures and words. These follow-up studies are currently underway.

4.4 Conclusions

The present set of experiments contributed several important pieces of information in the characterization of the ERP components elicited by pictorial stimuli. First, the presence of a distinct N300 in the pure picture task suggests that the N300 is specific to the semantic processing of pictures and is not due to linguistic mediation. Second, the different scalp topographies observed for the N300 and N400 components in the picture story experiment provided further evidence that these two negativities in fact reflect separate and distinguishable components. Finally, the anterior scalp distribution of the N400 effects for pictures suggests pictures may access a similar "N400 mechanism" as concrete words for integrating image-mediated information into a prior context.

This set of experiments also revealed that the N400 is modulated by discourse-level (global) coherence manipulations with pictures. This finding has important implications for theories regarding the underlying mechanism of the N400. It strongly supports the theory that the N400 reflects the ease of integration of information from multiple sources into an ongoing discourse representation and opposes pre-lexical accounts of N400 function.

Finally, these experiments have important implications in the semantic systems debate. The similarity of the N400 for pictures and concrete words suggests that image-mediated (concrete) words and pictures both access an amodal semantic system which processes imagistic as well as abstract conceptual information. The presence of the strictly pictorial N300, however, suggests that a separate nonverbal semantic mechanism exists that is accessed by pictorial information, and may later be accessed by image-mediated words (N700). These findings strongly support at least partial modularity of semantic representations and semantic processing mechanisms in the human brain.

References

- [1] J.R. Anderson, G.H. Bower, *Human associative memory*, Wiley, New York, 1973.
- [2] P. Baggett, Memory for explicit and implicit information in picture stories, *Journal of Verbal Learning & Verbal Behavior*, 14 (1975) 538-548.
- [3] P. Baggett, Structurally equivalent stories in movie and text and the effect of the medium on recall, *Journal of Verbal Learning & Verbal Behavior*, 18 (1979) 333-356.
- [4] S.E. Barrett, M.D. Rugg, Event-related potentials and the semantic matching of pictures, *Brain and Cognition*, 14 (1990) 201-212.
- [5] A.M. Bihrlé, H.H. Brownell, J.A. Powelson, H. Gardner, Comprehension of humorous and nonhumorous materials by left and right brain-damaged patients, *Brain and Cognition*, 5 (1986) 399-411.
- [6] H.H. Brownell, D. Michel, J. Powelson, H. Gardner, Surprise but not coherence: sensitivity to verbal humor in right-hemisphere patients, *Brain and Language*, 18 (1983) 20-7.
- [7] H.H. Brownell, H.H. Potter, A.M. Bihrlé, H. Gardner, Inference deficits in right brain-damaged patients, *Brain and Language*, 27 (1986) 310-21.
- [8] A. Caramazza, A.E. Hillis, B.C. Rapp, C. Romani, The multiple semantics

- hypothesis: Multiple confusions?, *Cognitive Neuropsychology*, 7 (1990) 161-189.
- [9] C.E. Elger, T. Grunwald, K. Lehnertz, M. Kutas, C. Helmstaedter, A. Brockhaus, D. Van Roost, H.J. Heinze, Human temporal lobe potentials in verbal learning and memory processes, *Neuropsychologia*, 35 (1997) 657-667.
- [10] K.D. Federmeier, M. Kutas, Meaning and modality: influences of context, semantic memory organization, and perceptual predictability on picture processing, *J Exp Psychol Learn Mem Cogn*, 27 (2001) 202-24.
- [11] G. Ganis, M. Kutas, M.I. Sereno, The search for "common sense": An electrophysiological study of the comprehension of words and pictures in reading, *Journal of Cognitive Neuroscience*, 8 (1996) 89-106.
- [12] S. Geisser, S. Greenhouse, On methods in the analysis of profile data, *Psychometrika*, 24 (1959) 95-112.
- [13] M.A. Gernsbacher, Surface information loss in comprehension, *Cognitive Psychology*, 17 (1985) 324-363.
- [14] E. Halgren, P. Baudena, G. Heit, J.M. Clarke, K. Marinkovik, Spatio-temporal stages in face and word processing. 1. Depth-recorded potentials in the human occipital, temporal, and parietal lobes, *Journal of Physiology*, 88 (1994) 1-50.
- [15] E. Halgren, P. Baudena, G. Heit, J.M. Clarke, K. Marinkovik, P. Chauvel, Spatio-temporal stages in face and word processing. 2. Depth-recorded potentials in the human frontal and Rolandic cortices, *Journal of Physiology*, 88 (1994) 51-80.
- [16] P.J. Holcomb, Semantic priming and stimulus degradation: Implications for the role of the N400 in language processing, *Psychophysiology*, 30 (1993) 47-61.
- [17] P.J. Holcomb, J. Kounios, J.E. Anderson, W.C. West, Dual coding, context availability, and concreteness effects in sentence comprehension: An electrophysiological investigation, *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 25 (1999) 721-742.
- [18] P.J. Holcomb, W.B. McPherson, Event-related brain potentials reflect semantic priming in an object decision task, *Brain and Cognition*, 24 (1994) 259-276.
- [19] P.J. Holcomb, H.J. Neville, Auditory and visual semantic priming in lexical decision: A comparison using event-related brain potentials, *Language and Cognitive Processes*, 5 (1990) 281-312.
- [20] A. Ishai, L.G. Ungerleider, J.V. Haxby, Distributed neural systems for the generation of visual images, *Neuron*, 28 (2000) 979-90.
- [21] W. Kintsch, T.A. van Dijk, Toward a model of text comprehension and production, *Psychological Review*, 85 (1978) 363-394.
- [22] S.M. Kosslyn, W.L. Thompson, N.M. Alpert, Neural systems shared by visual imagery and visual perception: A positron emission tomography study, *NeuroImage*, 6 (1997) 320-334.
- [23] J. Kounios, P.J. Holcomb, Concreteness effects in semantic processing: ERP evidence supporting dual-coding theory, *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 20 (1994) 804-823.
- [24] J.F. Kroll, M.C. Potter, Recognizing words, pictures, and concepts: A comparison of lexical, object, and reality decisions, *Journal of Verbal Learning and Verbal Behavior*, 23 (1984) 39-66.
- [25] M. Kutas, K.D. Federmeier, Electrophysiology reveals semantic memory use in language comprehension, *Trends in Cognitive Science*, 4 (2000) 463-470.
- [26] M. Kutas, S.A. Hillyard, Reading senseless sentences: Brain potentials reflect semantic incongruity, *Science*, 207 (1980) 203-205.
- [27] M. Kutas, S.A. Hillyard, Brain potentials during reading reflect word expectancy and semantic association, *Nature*, 307 (1984) 161-163.
- [28] G. McCarthy, A.C. Nobre, S. Bentin, D.D. Spencer, Language-related field potentials in the anterior-medial

- temporal lobe: I. Intracranial distribution and neural generators, *Journal of Neuroscience*, 15 (1995) 1080-1089.
- [29] G. McCarthy, C.C. Wood, Scalp distributions of event-related potentials: An ambiguity associated with analysis of variance models, *Electroencephalography and Clinical Neurophysiology*, 62 (1985) 203-208.
- [30] W.B. McPherson, P.J. Holcomb, An electrophysiological investigation of semantic priming with pictures of real objects, *Psychophysiology*, 36 (1999) 53-65.
- [31] A. Nigam, J.E. Hoffman, R.F. Simons, N400 to semantically anomalous pictures and words, *Journal of Cognitive Neuroscience*, 4 (1992) 15-22.
- [32] A.C. Nobre, G. McCarthy, Language-related field potentials in the anterior-medial temporal lobe: II. Effects of word type and semantic priming, *Journal of Neuroscience*, 15 (1995) 1090-1098.
- [33] L. Osterhout, P.J. Holcomb, Event-related potentials elicited by syntactic anomaly, *Journal of Memory and Language*, 31 (1992) 785-806.
- [34] A. Paivio, *Imagery and verbal processes*, Holt, Rinehart, and Winston, New York, 1971.
- [35] A. Paivio, *Mental representations: A dual coding approach*, Oxford University Press, New York, 1986.
- [36] A. Paivio, Dual coding theory: Retrospect and current status, *Canadian Journal of Psychology*, 45 (1991) 255-287.
- [37] Z.W. Pylyshyn, Computation and cognition: Issues in the foundations of cognitive science, *Behavioral and Brain Sciences*, 3 (1980) 111-132.
- [38] M.D. Rugg, M.C. Doyle, J.S. Holdstock, Modulation of event-related brain potentials by word repetition: Effects of local context, *Psychophysiology*, 31 (1994) 447-459.
- [39] T. Shallice, Specialisation within the semantic system, *Cognitive Neuropsychology*, 5 (1988) 133-142.
- [40] T. Shallice, Multiple semantics: Whose confusions?, *Cognitive Neuropsychology*, 10 (1993) 251-261.
- [41] A.P. Shimamura, Memory and frontal lobe function. In M.S. Gazzaniga (Ed.), *The cognitive neurosciences*, MIT Press, Cambridge, MA, 1995, pp. 803-813.
- [42] M.E. Smith, J.M. Stapleton, E. Halgren, Human medial temporal lobe potentials evoked in memory and language tasks, *Electroencephalogr Clin Neurophysiol*, 63 (1986) 145-59.
- [43] M. St. George, M. Kutas, A. Martinez, M.I. Sereno, Semantic integration in reading: engagement of the right hemisphere during discourse processing, *Brain*, 122 (1999) 1317-1325.
- [44] M. St. George, S. Mannes, J. Hoffman, Global semantic expectancy and language comprehension, *Journal of Cognitive Neuroscience*, 6 (1994) 70-83.
- [45] J.J. van Berkum, P. Hagoort, C.M. Brown, Semantic integration in sentences and discourse: evidence from the N400, *Journal of Cognitive Neuroscience*, 11 (1999) 657-71.
- [46] W.C. West, P.J. Holcomb, Imaginal, semantic, and surface-level processing of concrete and abstract words: An electrophysiological investigation, *Journal of Cognitive Neuroscience*, 12 (2000) 1024-1037.