Journal of Experimental Psychology:
Learning, Memory, and Cognition
1999, Vol. 25, No. 3, 721–742

0278-7393/99/$3.00

Dual-Coding, Context-Availability, and Concreteness Effects in Sentence Comprehension: An Electrophysiological Investigation

Phillip J. Holcomb
Tufts University

John Kounios
University of Pennsylvania

Jane E. Anderson
Harvard Medical School

W. Caroline West
Tufts University

Event-related potentials were recorded in 2 experiments while participants read sentences in a word-by-word congruency judgment task. Sentence final words were either congruent, semantically anomalous (Experiments 1 and 2), or neutral (Experiment 2) with respect to sentence context. Half of all final words referred to concrete and half to abstract concepts. A different scalp distribution of the N400 to concrete and abstract final words was found for anomalous and neutral, but not congruent sentences. Although the interaction of context and concreteness is consistent with the context-availability model, the differential scalp distribution of effects for concrete and abstract words, as well as larger context effects for concrete words, was interpreted as being more consistent with an extended dual-code account of semantic processing.

Theories of how knowledge is stored and processed in the brain have generally fallen into one of two camps. The first proposes that all meanings for objects, events, and concepts are stored and processed by a common amodal semantic system (e.g., Caramazza, Hillis, Rapp, & Romani, 1990; Gernsbacher, 1985; Kroll & Potter, 1984; Pylyshyn, 1984). Conversely, the second class of theories posits that multiple semantic systems independently store and process semantic information, often redundantly (e.g., Paivio, 1971, 1986, 1991; Shallice, 1988, 1993).

A salient example of this distinction, and the focus of the present study, has been the debate over the origin of *concreteness effects,* which is the observation that words representing concrete concepts (e.g., *table*) are processed more quickly and accurately than words representing comparatively abstract concepts (e.g., *aptitude*). Unfortunately, in spite of the relative consistency of the experimental findings, there has been considerable disagreement concerning the source of concreteness effects, the two major theoretical contenders being dual-coding theory and the context-availability model (see reviews by Paivio, 1991; Schwanenflugel, 1991).

Phillip J. Holcomb and W. Caroline West, Department of Psychology, Tufts University; John Kounios, Institute for Research in Cognitive Science, University of Pennsylvania; Jane E. Anderson, Department of Psychiatry, Harvard Medical School.

This research was supported by National Institute of Child Health and Development Grant HD25889 and National Institute of Mental Health Grant MH57501.

Correspondence concerning this article should be addressed to Phillip J. Holcomb, Department of Psychology, Tufts University, 490 Boston Avenue, Medford, Massachussetts 02155. Electronic mail may be sent to pholcomb@tufts.edu.

Dual-Coding Theory

Dual-coding theory (Paivio, 1986, 1991) explains concreteness effects by recourse to modality-specific systems for representation and processing. According to this theory, a variant of the multiple semantic-systems view, concrete words are associated with information stored in both a verbal "linguistic" semantic system and a nonverbal "imagistic" semantic system. Abstract words, however, are associated primarily with information stored in the linguistic system. When one encounters a concrete word, the word initially activates linguistic information, but shortly thereafter it also begins to activate imagistic information by means of referential links that interconnect the linguistic and image systems. Abstract words, on the other hand, lack or have many fewer referential connections between systems and predominantly activate linguistic representations. Concrete words have distinct processing advantages over abstract words because they have access to information from multiple systems. So, for example, in a lexical-decision task participants can classify *hand* as a word faster than *idea* because *hand* is processed and represented in both systems whereas *idea* is processed and represented only in the linguistic system. This additional semantic activity from dual systems allows participants to quickly differentiate concrete words from pseudowords (pseudowords presumably generate little semantic activation). The relatively lower semantic activity from a single system makes abstract words more difficult to differentiate from pseudowords, resulting in relatively slower reaction times (RTs).

Context-Availability Model

In contrast, the context-availability model (Bransford & McCarrell, 1974; Kieras, 1978) denies the existence of different types of informational codes or processing systems as determinants of concreteness effects. This theory, a variation of the single semantic-system view, argues that comprehension is heavily reliant on context supplied by either the preceding discourse or the comprehender's own mental knowledge base (semantic memory). Concrete words are thought to be more closely associated to relevant contextual knowledge in semantic memory than are abstract words, because concrete words exhibit stronger or more extensive associative links to this stored material. However, the underlying nature of the representations for the two word types and the processes that operate on these representations do not differ according to this account. So, participants can classify *hand* as a word faster than *idea* because *hand* activates more semantic information. Where context availability differs from dual coding is in how and where this additional information is stored and processed. Context availability argues for a simpler quantitative difference between word types within a single system, whereas dual coding argues for a qualitative difference based on activity in different systems.

Numerous studies have sought to empirically invalidate one or the other of these explanations (Paivio, 1986, 1991; Schwanenflugel, 1991). Unfortunately, many of these experiments have exhibited what hindsight suggests to be methodological and theoretical limitations. At least in the realm of semantic processing (the focus of this report), neither view can claim, on the basis of previous findings, to be the complete explanation of concreteness effects. Below, we briefly discuss some of the relevant research supporting both accounts. Following this, we describe a paradigm and methodology designed to circumvent some of the difficulties inherent in previous studies with the aim of providing data capable of discriminating between the competing theories.

Context and Concreteness

Schwanenflugel and her colleagues have been the most outspoken proponents of the context-availability interpretation of concreteness effects in semantic processing (e.g., Schwanenflugel, 1991; Schwanenflugel, Harnishfeger, & Stowe, 1988; Schwanenflugel & Shoben, 1983; Schwanenflugel & Stowe, 1989). They have presented two sources of evidence that they argue favor this model. First, Schwanenflugel and colleagues (e.g., Schwanenflugel et al., 1988; Schwanenflugel & Shoben, 1983) found that participants' estimates of the relative difficulty of retrieving associated contextual information for isolated abstract and concrete words (context availability ratings) were correlated with concreteness ratings. Moreover, they found that these context-availability ratings were a better predictor of lexical-decision performance than rated concreteness or imageability. When concrete and abstract words were equated on this variable, the advantage normally seen for concrete words was no longer significant. However, one potentially serious problem with these studies is that it is not clear how participants actually made context-availability ratings. In particular, the authors of these studies apparently did not check to see if participants might have sometimes used some type of imagery strategy. For example, it might be that for concrete words, and even some abstract words, many participants used mental images to help determine how easy or how many different contexts a word can be used in. Thus, partialing out rated concreteness might have missed an important residual dimension of concreteness or imagery. To eliminate this possibility participants' actual generated contexts would have to be monitored and controlled for image-based intrusions.

In a second series of experiments Schwanenflugel and colleagues (e.g., Schwanenflugel et al., 1988; Schwanenflugel & Shoben, 1983; Schwanenflugel & Stowe, 1989) have more convincingly demonstrated that when sufficient supportive context is provided, either in the form of several or even a single prior sentence, concreteness effects on accuracy and RT diminish or even vanish in a variety of tasks including lexical decision, naming, and sentence meaningfulness judgments. This effect takes the form of context producing large changes in performance on abstract items and little or no change in performance on concrete items. Schwanenflugel and colleagues argued that this implies that concreteness effects are reducible to differences in the availability of context. In other words, when abstract words are provided with an external context, such as a supportive sentence stem, of equivalent potency to that normally available to concrete words from within semantic memory, then they are processed as efficiently as concrete words. Concrete words do not benefit as much from an external context because they already have strong built-in contexts, so an external context does little to change how these items are processed. Therefore, according to this view, there is no need to postulate a more architecturally complex separate system for representing and processing imagistic information.

We believe that the above conclusion may be premature. This is because the basic dual-coding theory as proposed by Paivio (1986) does not argue that context cannot facilitate semantic processing, nor does it argue that such contextual facilitation could not supersede or mask concreteness effects. It simply states that there are separate imaginal and linguistic systems, and that the consequences of this structural configuration can sometimes be manifested in behavior. With minor augmentation dual-coding theory could be made to account for context effects such as those reported by Schwanenflugel and colleagues, while retaining its central feature of multiple systems. One possibility is that a supportive sentence frame works by "priming" relevant sentence final words (e.g., Stanovich & West, 1983). The resulting semantic facilitation within the linguistic system may be sufficient to overcome the added benefit concrete words normally exhibit in isolation due to referential processing by means of the imagistic system.[1] This could happen, for example, because context works faster or earlier than concreteness.

[1] Here *priming* is used to refer to the notion of semantic integration. In this view, a prior context (a related word or supportive sentence stem) facilitates processing by establishing a discourse representation into which the meaning of the target word can easily be fit (e.g., Hess et al., 1995).

Note that at one level the above explanation is actually very similar to that offered by context availability: External context compensates for richer or stronger internal associations by means of priming in the linguistic system. However, upon closer inspection the two theories remain quite distinct. Although context availability argues that the biggest effects of linguistic context should occur for abstract words, context-extended dual-eoding theory predicts similar effects of linguistic context for abstract and concrete words within the verbal system. However, it also predicts larger effects of context for concrete than abstract words within the imagistic system. We return to the discussion of the predictions of each model in the General Discussion section, but for now, suffice it to say that the finding of a Context × Concreteness interaction does not constitute evidence against the existence of multiple systems. Rather it simply implies that there are two different factors that affect one or more processing stages in common (McClelland, 1979; Sternberg, 1969). If this is true, then one of the major tenets of the context-availability critique of dual-coding research would be rendered nugatory.

Event-Related Potentials (ERPs)

The ERP technique involves measuring the brain's electrical activity (with electroencephalograms [EEGs]) detectable by scalp electrodes after specific stimulus events. These individual EEG waveforms are then averaged across stimulus presentations to yield a waveform characterizing the measurable part of the brain's electrical response to the stimulus (i.e., minus the "noise" representing other brain activity). Researchers have linked the various components (roughly the peaks) in the ERP waveform to a number of cognitive processes (see Coles & Rugg, 1995, for a recent introduction and review).

The ERP component of most interest for present purposes is the N400, which is a negative-going wave peaking at approximately 400 ms after the onset of a stimulus. Numerous studies have implicated the N400 in some aspect of semantic processing. For example, Kutas and Hillyard (1980, 1984), in what is now the classic N400 paradigm, were the first to show that this component is larger in amplitude in response to sentence final words that are semantically anomalous (e.g., "He takes cream and sugar in his *attention.*"), and is greatly reduced or even absent in response to high probability congruent sentence endings (e.g., "He takes cream and sugar in his *coffee.*"). In contrast, the N400 has been shown to be unaffected by physically anomalous but semantically congruent stimuli (e.g., a congruent final word in a different type font). Subsequent studies revealed that semantic anomalies throughout a sentence elicit N400s (e.g., Kutas & Hillyard, 1983) as do anomalies occurring in sentences presented at very rapid rates (up to 10 words/s; Kutas, 1987).

An important study by Kutas, Lindamood, and Hillyard (1984) showed that robust N400s could be obtained without the occurrence of a semantic anomaly and that under these conditions its amplitude is a monotonic, decreasing function of the cloze probability of the final words of sentences. For example, N400 amplitude was large in response to less predictable, but nonanomalous words (e.g., "Captain Sheir wanted to stay with the sinking *raft.*") and was smaller in response to more predictable words (e.g., "She called her husband at his *office.*"). On the basis of findings such as these and others it has been argued that the N400 reflects the process of integrating semantic information into a relatively high level discourse or mental model type of representation (e.g., Brown & Hagoort, 1993; Holcomb, 1993). In this formulation, larger N400s are taken as being indicative of a more effortful or involved integration process. A common thread in virtually every account of the N400 is that it is sensitive to contextual and semantic manipulations, which would appear to make it an ideal choice of dependent variables for use in searching for the locus of interactions between concreteness and context availability.

Experiment 1

Kounios and Holcomb (1994) demonstrated that ERPs are sensitive to concreteness, particularly when the task involves deep semantic processing. Specifically, concrete words presented in a list elicited a more negative ERP between 300 and 500 ms after stimulus onset than did abstract words, and this difference was larger in a semantic categorization task (Experiment 2) than in a lexical-decision task (Experiment 1). This suggests that the effects were at least partially mediated by semantic properties of the words. The ERP negativity coincided temporally with the classic N400, although the topographic distribution of the concreteness effect differed from the typical N400 effect in that it was maximal at anterior scalp locations, whereas the typical N400 effect tends to be centro-parietal (e.g., Kutas & Van Petten, 1988). However, more "anterior N400" distributions have been observed previously in single-word tasks (e.g., Boddy, 1986; Nobre & McCarthy, 1994) and in picture-processing tasks (e.g., Ganis, Kutas, & Sereno, 1996; Holcomb & McPherson, 1994; McPherson & Holcomb, in press). The most likely explanation for these differing distributions is that the N400 is actually the product of several underlying neural generators whose weights vary according to task demands and stimulus properties (Kounios, 1996; Nobre & McCarthy, 1994). In other words, the N400 probably does not reflect a unitary process, but rather several functionally related, but neurally distinct ones. Regardless of whether this or some other explanation of the differing scalp distributions for the N400 is correct, the differential scalp distributions of ERPs reported by Kounios and Holcomb (1994) for concrete and abstract words suggest the involvement of nonidentical neural and cognitive processing systems, which if correct, would be inconsistent with single-code theories such as the context-availability model.

The logic of the above conclusion is based on what we call the "spatial distinctiveness principle," which assumes that two or more different cognitive systems will tend to be more spatially distinct within the brain than will a single cognitive system. For example, if concrete and abstract words activate the same basic cognitive processes, representations, or both (a common semantic system) then, on average, the same population of neurons should be active when these two types of words are processed and these neurons should produce, on average, a similar spatial pattern of electrical activity for both types of events. With electrodes placed on the scalp this activity should show up as a difference in the size of potentials (i.e., a significant main effect of word type), but not as a difference in the distribution of potentials across the scalp (i.e., an absence of a Word Type × Scalp Site interaction). If, however, the two types of words influence somewhat different cognitive processes and representations (separate semantic systems) then, on average, somewhat distinct populations of neurons should be active when these two types of events are processed and these different populations should produce electrical activity with different scalp distributions (a Word Type × Scalp Site interaction).

The spatial distinctiveness principle is quite robust and, in theory, can differentiate between a variety of possible neural organizations.[2] For example, a single system that encompasses an expansive cortical area or even multiple areas could be differentiated from separate systems with equivalent spatial distributions. This is because a large or multiarea single system need only be spatially homogeneous with respect to the factor of interest. Such a single large system would produce main effects for differences in strength of activity due to the factor of interest (e.g., word type) that extend across a relatively wide region of scalp, but no interaction between factor and site. Separate systems of equivalent spatial distribution would produce the Factor × Site interaction. This logic works for much smaller brain regions as well, although an adequate number of electrodes placed close enough together to differentiate the systems is necessary.

Another important finding in the Kounios and Holcomb (1994) study was that repetition had a greater effect on the ERPs to concrete than abstract words. Furthermore, large repetition effects (decreased N400 amplitude to repeated words) were observed over both hemispheres in both experiments for concrete words, but were not observed for abstract words in Experiment 1 and only over the left hemisphere in Experiment 2. These results are in direct contrast with the context-availability model, which predicts that added context (in this case in the form of repetition) should make a bigger difference in the processing of abstract words than concrete words (Schwanenflugel & Shoben, 1983). This finding could be interpreted as being consistent with the context-extended dual-coding theory described above. This model predicts similar effects of linguistic context for concrete and abstract words within the linguistic system, and larger effects of context for concrete than abstract words within the imagistic system.

The authors did, however, acknowledge some caveats concerning the interpretation of their results. First, the repetition effects reported in Kounios and Holcomb (1994) may not have tapped the same type of contextual processes as the sentence level context effects studied by Schwanenflugel and colleagues (e.g., Schwanenflugel & Shoben, 1983). Second, the frontal 300–500 ms negativity that differentiated concrete and abstract words may not have been a "true" N400 effect (neither of the tasks used by Kounios & Holcomb, 1994, were the classic N400 paradigm) and thus may reflect something other than semantic processing (see Neville, Kutas, Chesney, & Schmidt, 1987).

The present study was an attempt to address both of these issues and to replicate the Word Type × Scalp Site interaction that Kounios and Holcomb (1994) interpreted as evidence for dual-coding theory and against the context-availability model. To ensure that effects of context were being measured, a sentence processing task was used rather than a single word task. Furthermore, to ensure that what was being observed was, in fact, an N400, the classic anomalous sentence paradigm used to elicit an N400 response (Kutas & Hillyard, 1980) was used.

Sentences in which the final word was either congruent or incongruent (anomalous) were presented to participants whose task was to decide if each sentence made sense. In the present experiment, the concreteness of the sentence final words was manipulated in addition to their congruency; thus, it was designed to examine effects of both concreteness and context. The design is much like one used by Schwanenflugel and Stowe (1989), in which they found clear differences in meaningfulness-judgment RTs between concrete and abstract items when these words formed an anomalous ending of a sentence (concrete faster than abstract), but no differences in RT when the two word types were congruent final words. They interpreted this finding as further evidence for the context-availability model.

The results of Kounios and Holcomb (1994) suggest the following predictions with regard to ERPs in the current experiment. First, concrete words should produce ERPs that are more negative-going between 300 and 500 ms (the N400 window) than abstract words. Second, this difference should be due, at least in part, to the modulation of the traditional N400 component, which should be revealed by an interaction between context (a variable known to influence the N400) and concreteness. Third, supporting the dual-code model, but not the context-availability model, the distribution across the scalp of the context-sensitive N400 effect should be different for concrete and abstract words. Fourth, consistent with extended dual-coding theory, but not with the context-availability model, the biggest effects of context on the N400 should be for concrete words.

Method

Participants. Sixteen students (9 women and 7 men) between 18 and 20 years of age ($M = 18.63$ years) from the Tufts University

---

[2] The one condition under which the spatial distinctiveness principle might not prove valid would be if the different cognitive systems of interest are completely spatially intertwined in the brain. In this case, scalp potentials might not be sensitive enough to detect multiple systems. Note that this does not apply to functional intertwining. If systems are functionally intertwined then it becomes difficult to call them separate. Moreover, the type of complete structural intertwining that would be necessary for multiple systems to masquerade as a single system when monitored by scalp ERPs seems unlikely. Over 100 years of neuropsychological studies support a more "localist" view of brain organization (e.g., Kolb & Whishaw, 1996).

Table 1
Sample Sentences Used in Experiment 1

Final word and sentence type	Sample sentence
Concrete and Congruent	Armed robbery implies that the thief used a *weapon.*
Abstract and Congruent	Lisa argued that this had not been the case in one single *instance.*
Concrete and Anomalous	Armed robbery implies that the thief used a *rose.*
	Lisa argued that this had not been the case in one single *rose.*
Abstract and Anomalous	Armed robbery implies that the thief used a *fun.*
	Lisa argued that this had not been the case in one single *fun.*
Concrete and Neutral	They said it was because of the *rose.*
	Robert said it was due to the *weapon.*
Abstract and Neutral	They said it was because of the *fun.*
	Robert said it was due to this *instance.*

community served as participants. All were right-handed, native speakers of English with normal or corrected-to-normal visual acuity.

Stimuli and procedure. The stimuli for this experiment were 160 sentences, 80 of which ended in an abstract word (mean concreteness rating = 2.52), the remaining 80 of which ended in a concrete word (mean concreteness rating = 6.19). The sentences for the two final-word types did not differ in cloze probability (concrete = .61 [*SD* = .22], abstract = .59 [*SD* = .27]) and the final words did not differ in frequency (concrete = 65.1 per million [*SD* = 77], abstract = 69.7 per million [*SD* = 92]; Francis & Kučera, 1982) or length (concrete = 5.8 letters [*SD* = 1.75], abstract = 5.9 letters [*SD* = 1.65]). Concreteness and cloze probability were assessed by having separate groups of participants rate the materials. For concreteness, 35 participants rated the pool of 160 final words on a scale from 1 (*very abstract*) to 7 (*very concrete*). Several examples of each category were given prior to the test. Cloze probability was calculated by having another group of 15 participants fill in what they thought was the most appropriate final word for each of the 160 sentence stems. These participants were told to read each sentence stem quickly and to fill into the final word position the first word that came to mind. No word generated by any of the 15 cloze participants was used in forming the anomalous sentence endings, thus effectively making the cloze values for the anomalies zero.

From these sentences, four lists were formed such that half of the sentences in each list (80) ended in an appropriate or semantically congruent final word, whereas the other half ended in a semantically anomalous final word (see Table 1). The anomalous sentences were counterbalanced such that half of the 40 sentences that predicted concrete final words were completed with anomalous abstract words and half were completed with anomalous concrete words. For the 40 anomalous sentences that predicted abstract final words, half were completed with anomalous concrete words and the other half were completed with anomalous abstract words. So, each of the four lists had 40 sentences with semantically congruent concrete final words, 40 sentences with semantically congruent abstract final words, 40 sentences with semantically anomalous concrete final words (20 predicting an abstract word and 20 predicting a concrete word), and 40 sentences with semantically anomalous abstract final words (20 predicting an abstract word and 20 predicting a concrete word). Each participant read each of the 160 sentence frames and each of the 160 final words only once. However, across participants, each final word and each sentence frame appeared in both the congruent and anomalous conditions.

The experiment was self-paced, each trial beginning after the participant responded to the final word from the previous sentence. Participant responses ("yes" or "no") were registered with a small two-button panel resting in their lap. Three seconds following the participant's response, a fixation cross was displayed for 500 ms in the center of the computer monitor. This served as a warning that the next trial was about to begin. This was followed by a 300 ms blank screen, after which the first word of the sentence was displayed. The first word and each subsequent word in the sentence was sequentially displayed for 200 ms each. Consecutive words were separated by a 300 ms blank-screen interstimulus interval (ISI) for a total word-to-word onset interval of 500 ms. The final word of each sentence was displayed with a period to indicate that it concluded the sentence, and was followed by a 1,300 ms blank-screen ISI. All words were centered on the display monitor. After the final-word ISI, the message "RESPOND NOW" was presented in the center of the display until the participant made his or her response. Participants were instructed to press the "yes" button if the sentence made sense and the "no" button if it did not. They were also told not to move or blink their eyes from the onset of the fixation cross until the "RESPOND NOW" message was presented. Response hand was counterbalanced across participants. Each participant was given 10 practice trials prior to the run of 160 experimental sentences.

EEG procedure. Tin electrodes (Electro-Cap International, Eaton, Ohio) were placed at several scalp sites (see the head schematic in Figure 1 for the relative locations on the head), on the right mastoid bone, below the left eye, and to the right of the right eye (all referenced to the left mastoid). The scalp sites included standard International 10–20 System locations: occipital left (O1) and right (O2); frontal left (F7) and right (F8); midline frontal (Fz), central (Cz), and parietal (Pz). Six electrodes were also placed at nonstandard locations over left and right temporo-parietal cortex (30% of the interaural distance lateral to a point 13% of the nasion–inion distance posterior to Cz: Wernicke's left [WL] and right [WR]), temporal right and left (TR and TL; 33% of the interaural distance lateral to Cz) and anterior temporal right and left (ATR and ATL; 50% of the distance from T3/4 to F7/8).

The EEG was amplified with Grass Model 12 amplifiers (3dB cutoffs at .01 and 30 Hz; Grass, Quincy, MA) and was digitized on-line at 200 Hz. Average ERPs were formed off-line from correct-response trials free of ocular and movement artifacts (which averaged less than 10% across conditions).

Data analysis. The ERP data to sentence final words were quantified by calculating the mean amplitudes (relative to a 100-ms prestimulus baseline) in three latency windows (150–300 ms, 300–500 ms, and 500–800 ms). These windows were chosen because they roughly correspond to the latency ranges of the P2, N400, and the late positivity reported in previous language studies (see Kutas & Van Petten, 1988; Osterhout & Holcomb, 1995) and because they were also used in our earlier study comparing concrete and abstract words (Kounios & Holcomb, 1994).

The approach to statistical analysis involved the use of a

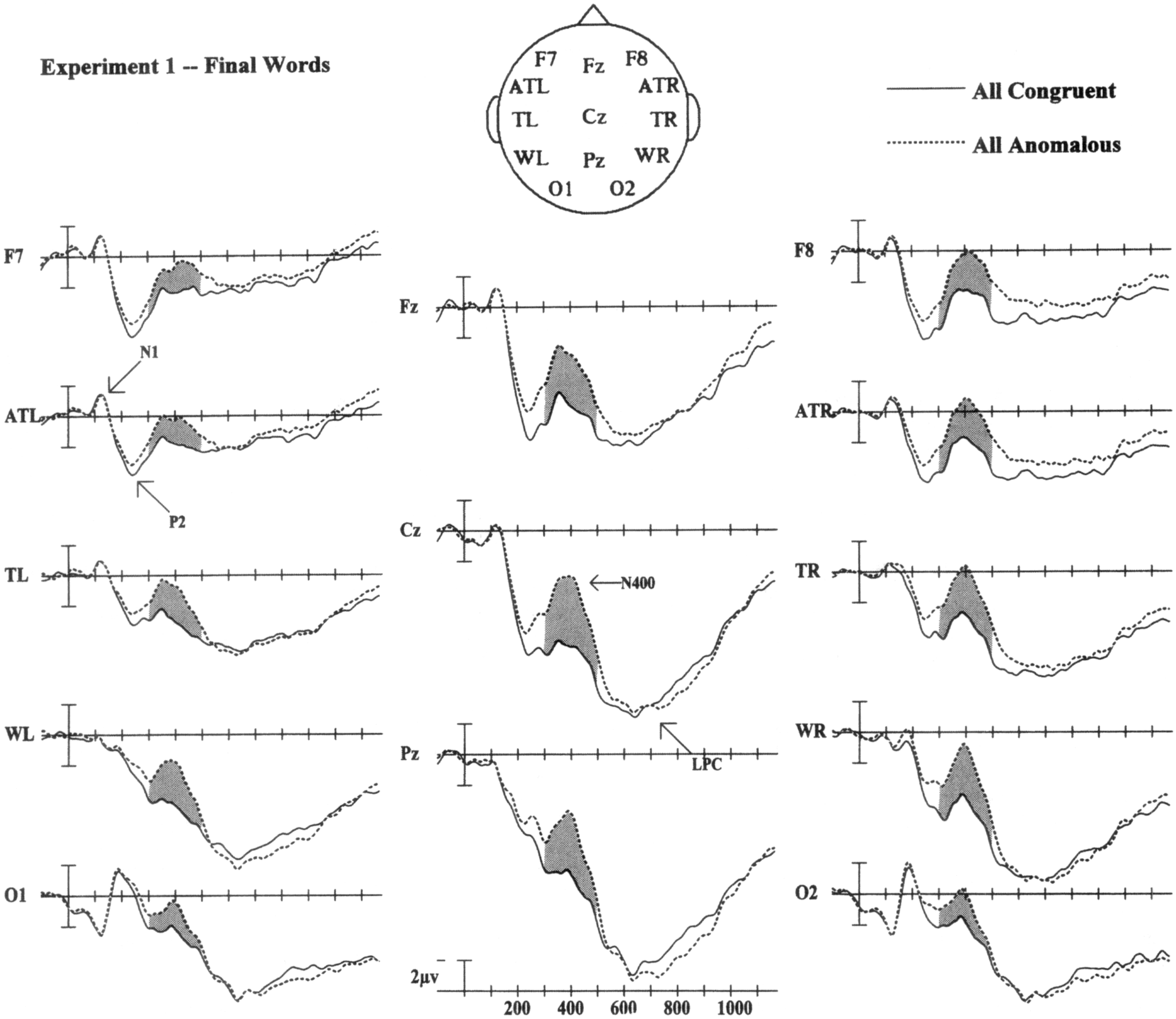

Figure 1. Grand mean event-related potentials (ERPs) from 13 scalp sites for congruent and anomalous final words in Experiment 1. Three sites down the middle of the head are plotted in the middle of the figure. The approximate locations of these sites can be seen in the head schematic located at the top of the figure (note that this is a view looking down at the top of the head with the nose pointed toward the top of the figure). The *x*-axes display time with the vertical calibration bar placed at the time of the onset of the stimulus. Each *x*-axis tick represents 100 ms. Note that 100 ms of activity prior to stimulus onset is displayed. This was used as a baseline for equating the poststimulus portion of each waveform. The *y*-axes represent voltages on a microvolt scale, with negative voltages plotted up, according to convention. Note that the peaks of the N1, P2, N400, and late positive components have been labeled at representative sites. Finally, the filled area between the ERPs for the congruent and anomalous final words represents the area used to quantify the N400 component (300 to 500 ms). F7/8 = frontal left/right; ATL/R = anterior–temporal left/right; TL/R = temporal left/right; WL/R = Wernicke's left/right; O1/2 = occipital left/right; Fz = frontal zero; Cz = central zero; Pz = parietal zero.

Table 2
Experiment 1—Sentence Judgments

	Final-word type			
	Concrete		Abstract	
Sentence type	% correct	*SD*	% correct	*SD*
Congruent	97.3	3.5	95.2	3.9
Anomalous	98.8	1.6	98.8	1.6

repeated measures analysis of variance (ANOVA) followed by simple effects tests in the case of significant interactions. There were two levels of sentence type (congruent vs. anomalous) and two levels of final word-type (concrete vs. abstract). ERPs at midline and lateral sites were analyzed in separate ANOVAs so that laterality effects could be assessed. Midline-site analyses included an electrode site variable (Fz vs. Cz vs. Pz). Lateral-site analyses included the variables of electrode site (frontal vs. anterior-temporal vs. temporal vs. Wernicke's vs. occipital), and hemisphere (right vs. left). The Geisser–Greenhouse correction (Geisser & Greenhouse, 1959) was applied to all repeated measures containing more than one degree of freedom in the numerator. Finally, analyses with significant interactions of stimulus variables with a topographic variable (e.g., electrode site or hemisphere) were repeated after amplitude values were normalized (using *z* scores) separately within each level of the word- or sentence-type variable (see McCarthy & Wood, 1985). Only interactions significant after normalization are reported.

Results

Accuracy data. Participants were very accurate in deciding if sentences made sense, averaging 97.5% correct responses (see Table 2). Overall, they were significantly more accurate in their decisions about anomalies than about congruent sentences: main effect of sentence-type, $F(1, 15) = 13.71$, $MSE = 0.01$, $p < .01$. Furthermore, participants were more accurate in their responses to concrete than abstract final words in the congruent-sentence condition, but performed equally well for the two word types in the anomalous-sentence condition: Sentence Type × Word Type interaction, $F(1, 15) = 4.87$, $MSE = 0.01$, $p < .05$.

Overview of ERPs. The grand-mean ERPs (averaged across all 16 participants) for congruent and anomalous final words are plotted in Figure 1. The ERPs in this figure show that there were several relatively early (less than 400 ms) components elicited in both conditions. These potentials are generally thought to reflect sensory and early perceptual processes (e.g., Rugg & Coles, 1995). They included a broadly distributed early negativity (N1) that peaked around 125 ms at all but the most posterior sites (i.e., O1, O2). At the posterior sites, there was an early positivity between 100 and 125 ms (P1) followed by a later N1 with a peak near 200 ms. At most sites, the N1 was followed by a positivity between 200 and 300 ms (P2). Note that, with the possible exception of the P2, none of the early components appeared to be differentiated according to sentence type.

There were also several later ERP components visible in the waveforms (see Figure 1). Following the P2, there was a negative-going wave that peaked around 400 ms. This negativity, which exhibited a broad scalp distribution, was clearly larger (i.e., was more negative-going) for anomalous than congruent final words, suggesting that it was a traditional N400 wave (cf. Kutas & Hillyard, 1980). Following the N400, there was a late positive component (LPC), which peaked between 600 and 800 ms over central and posterior sites. At these sites, the LPC was slightly larger and peaked somewhat later for anomalous final words than for congruent ones.

Analyses by epoch. Differences between anomalous and congruent final words started in the 150–300 ms time window, with anomalous endings producing more negative-going ERPs than congruent endings: main effect of sentence type, midline, $F(1, 15) = 13.60$, $MSE = 10.99$, $p < .01$; lateral, $F(1, 15) = 14.49$, $MSE = 9.25$, $p < .01$. However, there was not a significant main effect for word type or interaction between word type and electrode site in this epoch.

In the 300–500 ms epoch, which typically encompasses most of the activity of the N400 component (e.g., Kutas & Hillyard, 1980), anomalous sentences continued to elicit more negative ERPs than congruent sentences: main effect of sentence-type, midline, $F(1, 15) = 67.65$, $MSE = 20.42$, $p < .0001$; lateral, $F(1, 15) = 70.55$, $MSE = 20.33$, $p < .0001$; and concrete final words elicited more negative ERPs than abstract words: main effect of word type, midline, $F(1, 15) = 10.46$, $MSE = 10.06$, $p < .01$; lateral, $F(1, 15) = 24.32$, $MSE = 7.62$, $p < .0001$. As can be seen in Figure 1, the difference between congruent and anomalous sentences was largest at centro-parietal scalp sites: Sentence Type × Electrode Site interaction, midline, $F(2, 30) = 11.86$, $MSE = 0.08$, $p < .002$; lateral, $F(4, 60) = 8.67$, $MSE = 0.33$, $p < .01$; and was also larger over the right hemisphere than the left: Sentence Type × Hemisphere interaction, $F(1, 15) = 13.47$, $MSE = 0.41$, $p < .01$.

The difference between concrete and abstract final words was larger over anterior lateral sites than posterior lateral sites: Word Type × Electrode Site interaction, $F(4, 60) = 14.72$, $MSE = 0.16$, $p < .0001$. There were also significant Sentence Type × Word Type × Electrode Site interactions: midline, $F(2, 30) = 4.64$, $MSE = 0.07$, $p < .05$; lateral, $F(4, 60) = 7.36$, $MSE = 0.26$, $p < .01$, indicating that the effects of concreteness (word type) were different for congruent and anomalous final words (see the difference waves in Figure 2). Therefore, this latter interaction was followed up by separate ANOVAs for the two types of sentences. Analyses of the anomalous sentences produced a significant main effect of word type: midline, $F(1, 15) = 6.38$, $MSE = 20.63$, $p < .05$; lateral, $F(1, 15) = 13.88$, $MSE = 21.27$, $p < .01$, with concrete words more negative than abstract words (see Figure 3). Of particular importance, there was also a significant Word Type × Electrode Site interaction: midline, $F(2, 30) = 7.09$, $MSE = 0.09$, $p < .01$; lateral, $F(4, 60) = 18.50$, $MSE = 0.22$, $p < .0001$, indicating that the difference between concrete and abstract words was larger at more anterior sites. In contrast, for congruent sentences (see Figure 4), there were no significant main effects or interactions involving word type (*ps* > .5), indicating that the concreteness effects revealed by the overall analyses in the 300–500 ms window were due almost exclusively to the anomalous sentences (compare Figures 3 and 4).

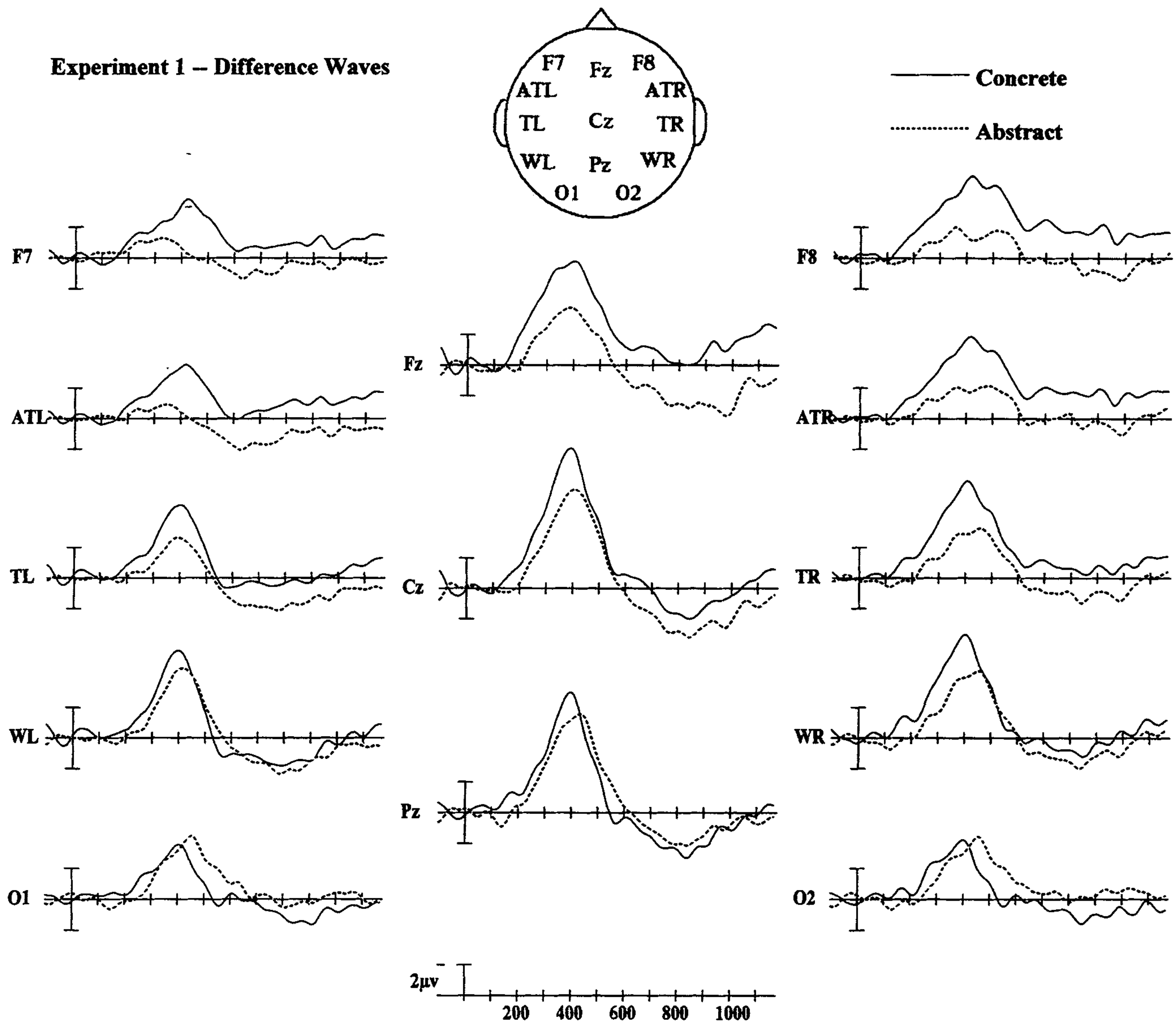

Figure 2. Plotted in this figure are difference waves that were produced by subtracting event-related potentials (ERPs) to congruent sentences from ERPs to anomalous sentences, for concrete and abstract final words in Experiment 1. The area under the large negative deflection between 200 and 600 ms represents the N400 effect (the difference between the anomalous and congruent final words). Note that this effect is larger for concrete than abstract words, especially at the most anterior sites.

As can be seen in Figure 1, the effects of sentence type had dissipated by the 500–800 ms temporal window at most electrode sites. There was, however, a small residual effect at right anterior sites that was revealed in the Sentence Type × Electrode Site × Hemisphere interaction, $F(4, 60) = 4.97$, $MSE = 0.07$, $p < .05$. However, as in the previous epoch, there continued to be robust effects of concreteness (word type—see Figure 3) at all anterior sites for anomalous sentences: Sentence Type × Word Type × Electrode Site interaction midline, $F(2, 30) = 13.25$, $MSE = 0.05$, $p < .001$; lateral, $F(4, 60) = 7.33$, $MSE = 0.14$, $p < .01$. This was confirmed by separate follow-up analyses on the two types of sentences. For the anomalous sentences, there was a significant Word Type × Electrode Site interaction: midline, $F(2, 30) = 9.38$, $MSE = 0.06$, $p < .01$; lateral, $F(4, 60) = 11.71$, $MSE = 0.10$, $p < .0001$. For the congruent sentences, again, there were no significant main effects or interactions involving word type (compare Figures 3 and 4).

Discussion

In this experiment, participants verified whether congruent and anomalous sentences with either concrete or abstract final words made sense. As in previous studies (e.g., Kutas &

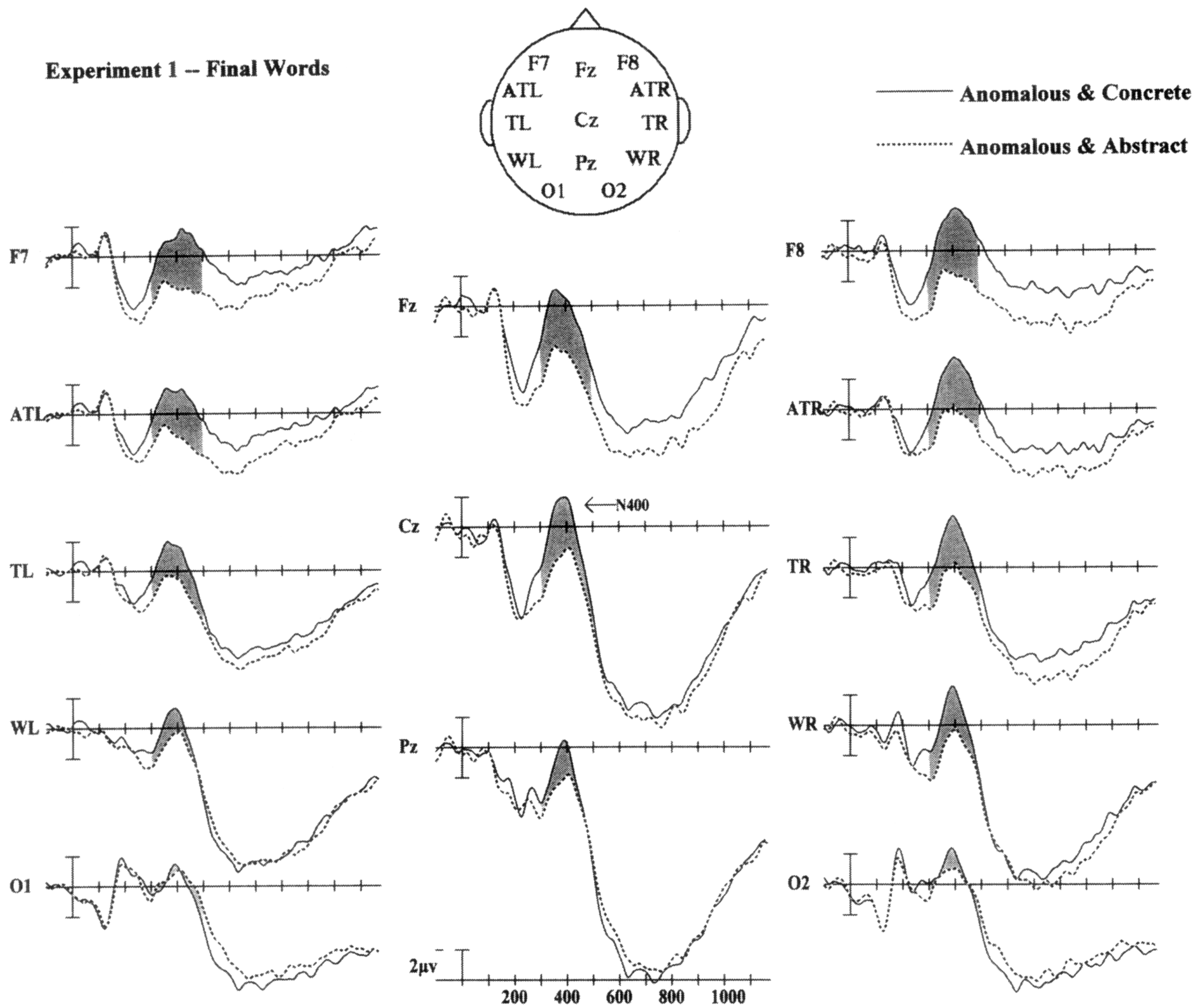

Figure 3. Grand mean event-related potentials for anomalous final words that were concrete or abstract (Experiment 1).

Hillyard, 1984) ERPs time-locked to the onset of final words revealed large effects of sentence context. Sentences with anomalous endings produced more negative-going ERPs than sentences with congruent endings, particularly in the 300–500 ms window. These effects of context had a central–posterior distribution and were slightly larger over the right than left hemisphere, which is consistent with the known distribution of the N400 component (Kutas & Hillyard, 1984). The most widely accepted view of the N400 is that it reflects differences in the ease of integrating semantic information into a higher level discourse or mental model type of representation (e.g., Brown & Hagoort, 1993; Holcomb, 1993); the larger the N400, the more effortful or involved the integration process. According to this view participants in the current study had more difficulty integrating the meaning of anomalous final words into the representation established by the prior sentence context.

As predicted, there were also clear effects of concreteness. Concrete final words elicited more negative-going ERPs than abstract final words, most notably within the window of the N400 (300–500 ms), but also extending into the later measurement epoch (500–800 ms). Also as predicted, the concreteness effect had a more anterior scalp distribution than the overall effects of context. However, of most importance to the goals of this study, in the 300–500 ms window, there was a clear interaction of concreteness and context that strongly suggests modulation of at least some of the neural generators associated with the N400 component. Given the large literature linking the N400 to the processing of meaning (see Osterhout & Holcomb, 1995, for a review),

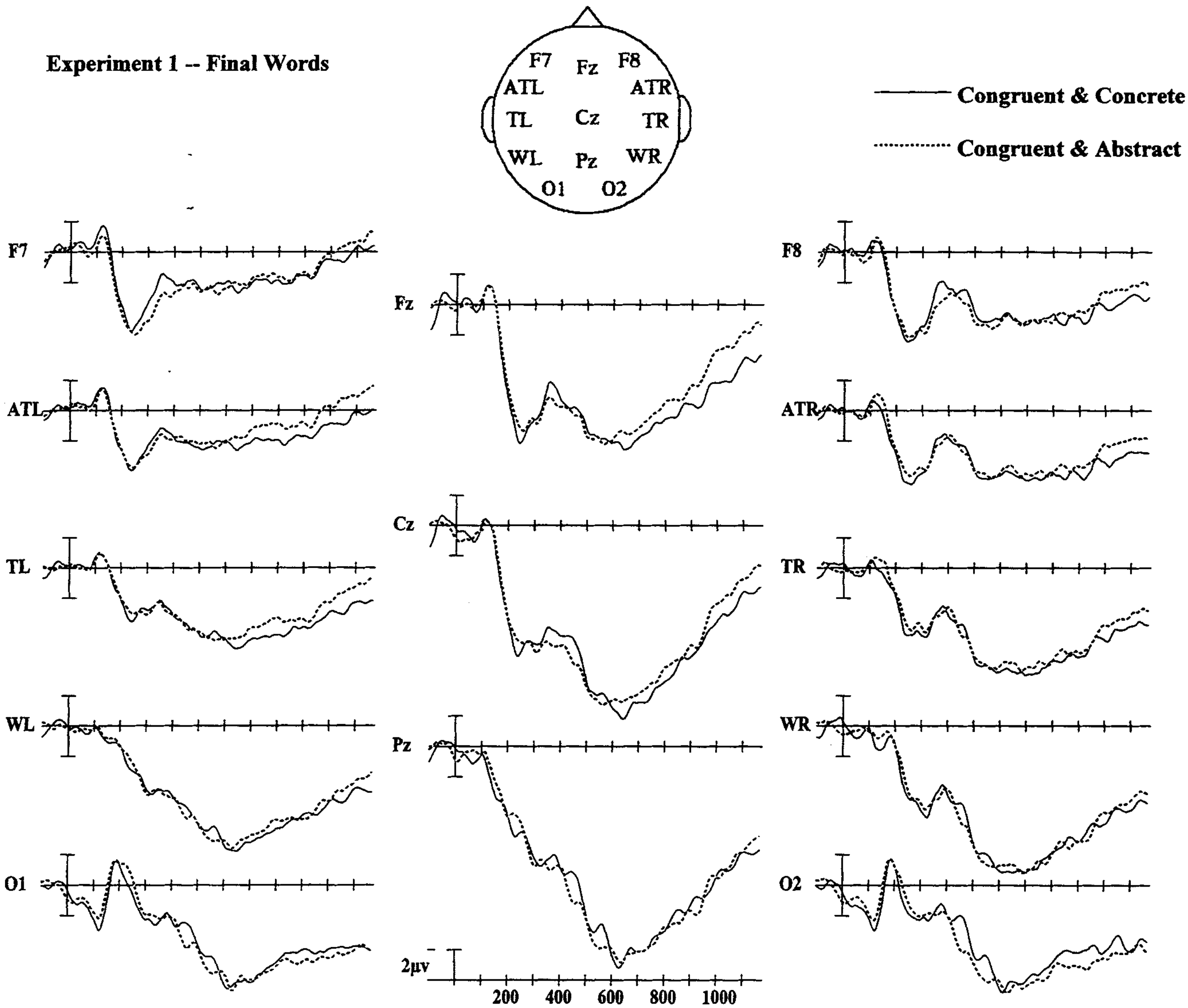

Figure 4. Grand mean event-related potentials for congruent final words that were concrete or abstract (Experiment 1).

this finding adds credence to the argument that the observed effects of concreteness and context are indeed semantic.

Interestingly, there were no discernible ERP differences in any window between concrete and abstract words in the congruent sentences. In other words, a supportive sentence context wiped out all ERP evidence of concreteness effects, a finding that would appear to be consistent with the central prediction of the context-availability model. However, two other findings, which replicate effects reported by Kounios and Holcomb (1994), are inconsistent with the predictions of this theory. First, in the anomalous sentences, the differences between concrete and abstract words varied systematically across the scalp. This concreteness effect was largest at the most anterior scalp sites, with concrete words producing much larger negativities than abstract words. This difference gradually decreased, moving toward the back of the head until at the most posterior sites (O1 and O2) there were virtually no concreteness amplitude differences (see Figure 3). Under the assumptions of the spatial distinctiveness principle outlined in the introduction, this pattern of results strongly suggests that nonidentical neural–cognitive systems are responsible for processing the two word types, at least when they are read outside of a supportive context. In this case, the labile topography of the N400 also supports the notion that this ERP component is generated by at least two different neural sources whose relative contributions to scalp topography vary across tasks and materials (Kounios, 1996; Nobre & McCarthy, 1994).

The above pattern of results is most consistent with multiple system theories such as dual-coding theory (Paivio, 1986), which argue that semantic information is represented and processed by two separate systems, one linguistic and one image based, and that although concrete words are processed in both systems, abstract words are processed primarily in the linguistic system. On the other hand, this pattern of results is clearly at odds with the predictions of single system accounts of concreteness, such as the context-availability model (Schwanenflugel, 1991), which argue that both word types are processed and represented within a single system. This latter type of model predicts only main effects of concreteness and a flat distribution for the N400 across the scalp (i.e., no interaction of scalp site and concreteness).

A second finding that was at odds with the predictions of the context-availability model and data reported to support it, was that the effects of context in the N400 window were larger for concrete than abstract words. In other words, when placed in a supportive context, concrete words revealed a more dramatic decline in N400 amplitude than abstract words. The context-availability model predicts larger context effects for abstract words.

In contrast to the above findings, prior findings using RT as the dependent measure have been interpreted as supporting the context-availability predictions. For example, in lexical decision and meaningfulness judgments, abstract words have been shown to produce large declines in RT going from a nonsupportive to a supportive sentence context, whereas concrete words, which produce relatively fast nonsupportive context RTs to begin with, show only small or no changes in RT when placed in a supportive context (Schwanenflugel & Shoben, 1983; Schwanenflugel & Stowe, 1989). Context availability proposes that the same process that underlies faster out-of-context concrete word RTs (built-in context), also accounts for concrete words' fast responses within a supportive sentence (external context) and that it is only the source of the context that differs. In essence, participants simply trade one source of context for another, and thus, RTs remain relatively unchanged.

This apparent contradiction in findings between behavioral measures and ERPs is not without precedence. Kounios and Holcomb (1992) also found that RT and N400 amplitude were not correlated in a sentence verification task and concluded that this is because these measures are not necessarily sensitive to the exact same set of underlying cognitive operations. In particular, they argued that RT is much more sensitive to participants' decision processes and task-dependent strategies than is N400 amplitude. Accordingly, one way to explain the apparent discrepancy between the RT and N400 Concreteness × Context interactions is to assume that the behavioral judgments in the Schwanenflugel studies (e.g., Schwanenflugel & Shoben, 1983) were relatively more sensitive to participants' decision strategies (in particular a two-step decision process), whereas the N400s in the current study and in the Kounios and Holcomb (1994) study were primarily sensitive to changes in a single semantic integration process and were relatively immune to decision processes. In this view, N400s were larger to out-of-context concrete than abstract words because concrete words have relatively more semantic information (linguistic and imagistic) that readers must attempt to integrate into an unrelated context and therefore integration was more effortful.[3] N400s were equivalently small to both word types following supportive contexts because semantic information from the final word was easily integrated into the representation established by the context. The observation that the N400s were equally small for congruent sentences fits with the fact that the concrete and abstract sentence stems were matched to produce equivalent levels of contextual support.

In the case of RT, this view proposes that when participants make lexical decisions or meaningfulness judgments they first use the strategy of determining if a target item can easily be integrated into the prior sentence context. If it can, then it's likely a word (pseudowords cannot be semantically integrated) or a meaningful item (nonmeaningful items cannot easily be integrated) and RT is similarly fast for both concrete and abstract words. If, however, the context is nonsupportive and integration does not easily occur, then participants must fall back on a second strategy to make their judgments. Here it is proposed that participants now use integration difficulty to aid their decision process. For example, in a meaningfulness judgment task they might respond to concrete words in an anomalous context more quickly than abstract words, because they detect the relatively greater effort involved in the semantic integration process for concrete words. A similar strategy would work for lexical decision, as it should be relatively easier to discriminate out-of-context concrete words from pseudowords than out-of-context abstract words from pseudowords because of the greater effort involved in integrating concrete words into a nonsupportive context. This type of explanation is similar to one offered by Neely and colleagues (e.g., Neely, 1991) to explain the pattern of RT effects in semantic priming lexical-decision tasks.[4]

[3] For convenience sake, we refer to semantic integration as a "single" process, although this may or may not be the case. For example, it is possible that there are separate integration processes for the linguistic and imagistic systems and that the more anterior N400 effect for concrete words reflects the relatively greater activity in the imagistic integrator for this word type, whereas the more posterior N400 effect reflects activity in the linguistic integrator for both word types. Alternatively, there may be a single integration process that serves both the imagistic and linguistic systems and the differential N400 distributions for concrete and abstract words may reflect the anatomically distinct representational inputs into this common process. In either case, one aspect of the task of integration(s) may be to decide which parts of the relatively greater amount of information (linguistic and imagistic) are relevant. With a poor or absent context, all of the semantic information for a given word may be important and therefore may need to be integrated or used to form a new context.

[4] To account for the effects of faster naming latencies for concrete words, a somewhat different explanation is required. Concrete words out of context might be named more quickly than abstract words because their greater semantic activation, in addition to making integration more effortful, also simultaneously feeds back to the lexical level speeding processes necessary for pronunciation. A similar architecture has been proposed in models such as interactive activation (e.g., Rummelhart & McClelland, 1982), where higher level information (e.g., semantic activation) starts to accrue very quickly, well ahead of complete word recognition.

One difference between the current results and those of Kounios and Holcomb (1994) is that the earlier study reported a lateral asymmetry for the N400 effect between concrete and abstract words (larger differences over the right hemisphere) in their second experiment. No such asymmetry was found here, although there was a slight (nonsignificant) trend in this direction (compare the WL and WR sites in Figure 2). The most likely reason for this discrepancy is differences in the task used in the two studies. Kounios and Holcomb (1994) found the asymmetry in a concreteness judgment task with single word contexts. No asymmetry was found here in a sentence meaningfulness judgment task. One possibility is that the larger right hemisphere effect was due to a more explicit use of imagery in the concreteness judgment task, although West and Holcomb (1998) also failed to find a hemispheric asymmetry between word types in a task where participants were asked about the ease of imageability.

In summary, the findings of this experiment support and extend Kounios and Holcomb (1994), showing that the concreteness variable yields a topographic distribution of ERPs consistent with the structural assumptions of extended dual-coding theory, but which is not reconcilable with the notion that concreteness effects are reducible to the availability of supportive context.

Experiment 2

In Experiment 1, no difference was observed in the ERPs elicited by congruent concrete and congruent abstract words. However, anomalous concrete words produced substantially more negative values in the time window of the N400 than did abstract words. Dual-coding theory proposes that this larger effect for concrete words is due to the activity of two qualitatively different systems (verbal and imagistic) processing the concrete words in parallel while only one system (the verbal system) processes abstract words. There is, however, an alternate explanation, namely, that in the case of the anomalies, the concrete final words were somehow more anomalous than the analogous abstract final words. Perhaps the abstract words were inherently more vague and seemed less anomalous than their concrete counterparts. By this same logic, the concrete words can be readily identified as sharply incongruent, because they refer to more specific entities. According to this view, the larger N400 seen for the concrete words in Experiment 1, and perhaps the more anterior distribution of this effect may have had nothing specifically to do with the concreteness of the word, but rather to the degree of anomaly. This is possible, because although the congruent sentences were carefully matched for cloze probability across the two word types, anomalies were formed by attaching final words from high-cloze sentences to another sentence stem. In forming anomalies, two rules were followed. First, none of the 15 individuals that participated in the cloze procedure filled in the target word for the anomalous sentence stem (i.e., cloze probability = 0). Second, in selecting items for anomalous sentence stems the experimenters used their best intuitions in attempting to pair anomalous endings with sentence stems so that each ending was not related to any acceptable ending for the sentence. This is probably not the best way to equate level of anomaly between the two word types. Ideally the concrete and abstract anomalies should have been equated on the basis of some more objective method of assessment.

Experiment 2 was designed to circumvent the possibility that the concrete anomalous endings were somehow more anomalous than the abstract anomalies and that this difference rather than differences in concreteness was responsible for the N400 concreteness findings in Experiment 1. However, rather than attempting to match the anomalous concrete and abstract final words on degree of anomaly, a neutral condition was added instead. This was done for two reasons. First, equating for degree of anomaly would undoubtedly require using a different set of sentences than those used in Experiment 1, making any between-experiment differences difficult to interpret. Second, anomaly ratings and ERPs would have had to have been collected in different groups of participants performing somewhat different tasks (anomaly detection vs. anomaly rating). This would make it difficult to know if degree of anomaly had been adequately controlled in the ERP participants. To circumvent these problems, we took a different approach. Rather than equating anomaly levels, we added a neutral condition to the design. In this condition, concrete and abstract words occurred at the ends of sentences where they were not anomalous, but also where they would not have been easily predicted. Schwanenflugel and Stowe (1989) used neutral sentences (e.g., "The next word will be . . .") in the out-of-context condition of one of their experiments contrasting concrete and abstract words. These sentences had the desired effect in their study of producing large differences in RT between concrete and abstract words. Similar low cloze probability sentences have been demonstrated by Kutas and colleagues (Kutas et al., 1984) to produce robust N400s. If, under this condition, concrete words do not elicit a larger N400 than abstract words, it could be concluded that the results observed in Experiment 1 may have been due to the concrete words having been perceived as more anomalous than the abstract words. Conversely, if concrete words still elicit a larger negativity in the neutral condition, then it can be concluded that the results of Experiment 1 were not simply due to the degree of anomaly, but rather reflect inherent differences in the processing of the two word types.

Neutral sentences were constructed such that the final words were congruent in that they fit into the sentences, but were of low cloze probability in that the sentence context provided no evidence on which to predict the final word or even to predict its concreteness (e.g., "Larry said it must have been the *wine.*"; "It happened because of her *mood.*"). Consistent with the results of Experiment 1 it was predicted that concrete words at the ends of neutral and anomalous sentences would produce larger N400s with a more anterior distribution than abstract words at the ends of these same sentences.

Method

Participants. Twenty-four naive undergraduates (21 women and 3 men) between 18 and 22 years of age (M = 18.67) from Tufts

Table 3
Experiment 2—Sentence Judgments

	Final-word type			
	Concrete		Abstract	
Sentence type	% correct	*SD*	% correct	*SD*
Congruent	95.8	7.2	95.5	7.0
Anomalous	97.5	3.0	96.3	4.0
Neutral	82.3	15.3	88.5	10.9

University served as participants. All were right-handed native speakers of English with normal or corrected-to-normal visual acuity.

Stimuli and procedure. Stimulus materials for congruent and anomalous sentences were taken from Experiment 1. However, half of the congruent sentence stems were replaced with neutral stems. Thus, Experiment 2 consisted of 80 anomalous sentences, 40 congruent sentences, and 40 neutral sentences. Neutral sentence frames were composed in such a way as to be semantically acceptable but contextually ambiguous (see Table 1 for an example). To assure that this was the case, we used a separate group of 15 participants to assess the cloze probabilities of these materials. These participants read the neutral stems and filled in the first sentence-ending word that came to mind. The mean cloze value for neutral sentences was .007 (1 participant each filled in the appropriate final word for each of 4 of the 40 neutral sentences).

As in Experiment 1, participants were instructed to press the "yes" button if the sentence made sense (congruent and neutral) and the "no" button if it did not (anomalies). Four lists were created so that anomalous and congruent–neutral sentence frames were each completed with both a concrete and an abstract final word. The lists were counterbalanced such that each word appeared in each type of sentence frame. Thus, each participant saw eight sentence types. These consisted of the same four anomalous sentence types from Experiment 1, congruent sentences with concrete or abstract final words, and neutral sentences with concrete or abstract final words. Both the experimental procedure and the ERP procedure were identical to that in Experiment 1. Data analysis was also identical to that in Experiment 1, except for the inclusion of three levels of sentence type (congruent vs. anomalous vs. neutral). These conditions were compared in a single omnibus ANOVA and were followed up with separate pairwise analyses where warranted.

Results

Accuracy data. In this experiment, participants were generally quite accurate in deciding whether sentences made sense, averaging 91.3% correct responses (see Table 3). However, participants were less accurate in correctly judging the neutral sentences than the congruent or anomalous sentences: main effect of sentence type, $F(2, 46) = 19.36$, $MSE = 104.57$, $p < .0001$. They also responded more accurately to sentences with abstract final words than to sentences with concrete final words: main effect of word-type, $F(1, 23) = 5.58$, $MSE = 30.87$, $p < .05$; but only in the neutral-sentence condition: Sentence Type × Word Type interaction, $F(2, 46) = 5.05$, $MSE = 33.01$, $p < .05$.

Overview of ERPs. The grand-mean ERPs (averaged across all 24 participants) for all congruent, anomalous, and neutral final words are plotted in Figure 5. In all conditions, the ERPs show early components (P1, N1, P2) similar to those observed in Experiment 1 (see Figure 1). Again, only the P2 component appeared to be affected by sentence type.

The later ERP components were also similar to those in the previous experiment. There was again a negative-going wave that peaked at approximately 400 ms (N400) that was broadly distributed. This component was clearly larger for anomalous and neutral final words than for congruent final words. Following the N400, there was again a positive-going wave (LPC) peaking at 600–800 ms over central and posterior sites. This late positivity was larger for anomalous final words and smaller for neutral final words.

Analyses by epoch. In the 150–300 ms time window, there was a main effect of sentence type: midline, $F(2, 46) = 5.46$, $MSE = 6.4$, $p < .01$; lateral, $F(2, 46) = 5.83$, $MSE = 10.3$, $p < .01$. Follow-up analyses indicated that anomalous final words produced more negative ERPs than either congruent or neutral final words: congruent versus anomalous, midline, $F(1, 23) = 10.54$, $MSE = 6.6$, $p < .01$; lateral, $F(1, 23) = 5.41$, $MSE = 9.2$, $p < .05$; anomalous versus neutral: midline, $F(1, 23) = 4.43$, $MSE = 4.9$, $p < .05$; lateral, $F(1, 23) = 14.96$, $MSE = 7.8$, $p < .001$. The main effect of concreteness was not significant in the 150–300 ms epoch, but there was a significant interaction between word type and sentence type: midline, $F(2, 46) = 6.09$, $MSE = 13.1$, $p < .01$; lateral, $F(2, 46) = 3.42$, $MSE = 15.8$, $p < .05$. Follow-up analyses indicated that concrete words yielded more negative ERPs than did abstract words for both anomalous and neutral sentences when compared with congruent sentences: anomalous versus congruent, midline, $F(1, 23) = 12.12$, $MSE = 10.5$, $p < .002$; lateral, $F(1, 23) = 5.96$, $MSE = 14.1$, $p < .03$; neutral versus congruent, midline, $F(1, 23) = 5.60$, $MSE = 19.7$, $p < .03$; lateral, $F(1, 23) = 3.79$, $MSE = 20.7$, $p < .06$; anomalous versus neutral, $ps > .7$.

In the 300–500 ms time-window, there was a main effect of sentence type: midline, $F(2, 46) = 34.57$, $MSE = 25.8$, $p < .0001$; lateral, $F(2, 46) = 28.50$, $MSE = 33.9$, $p < .0001$; and a Sentence Type × Electrode Site interaction: midline, $F(2, 46) = 11.2$, $MSE = 0.10$, $p < .0001$; lateral: $F(4, 92) = 25.38$, $MSE = 0.31$, $p < .0001$. Follow-up analyses indicated that both anomalous and neutral final words elicited more negative ERPs than did congruent final words, and that these effects tended to be larger over more posterior sites: Sentence Type × Electrode Site interaction, congruent versus anomalous, midline, $F(2, 46) = 5.52$, $MSE = 0.1$, $p < .05$; lateral, $F(4, 92) = 7.01$, $MSE = 0.3$, $p < .01$; congruent versus neutral, midline, $F(2, 46) = 19.60$, $MSE = 0.1$, $p < .0001$; lateral, $F(4, 92) = 43.95$, $MSE = 0.3$, $p < .0001$ (see Figure 5).

During the 300–500 ms epoch, there was also a main effect of word type: midline, $F(1, 23) = 23.48$, $MSE = 19.0$, $p < .001$; lateral, $F(1, 23) = 36.28$, $MSE = 23.7$, $p < .0001$; a Sentence Type × Word Type interaction: midline, $F(2, 46) = 9.30$, $MSE = 11.8$, $p < .001$; lateral, $F(2, 46) = 9.35$, $MSE = 23.6$, $p < .0001$; and a three-way interaction between word type, sentence type, and electrode site: midline, $F(4, 92) = 2.96$, $MSE = 0.1$, $p < .05$; lateral, $F(8, 184) = 9.03$, $MSE = 0.2$, $p < .0001$. Follow-up analyses indicated that concrete words produced significantly more

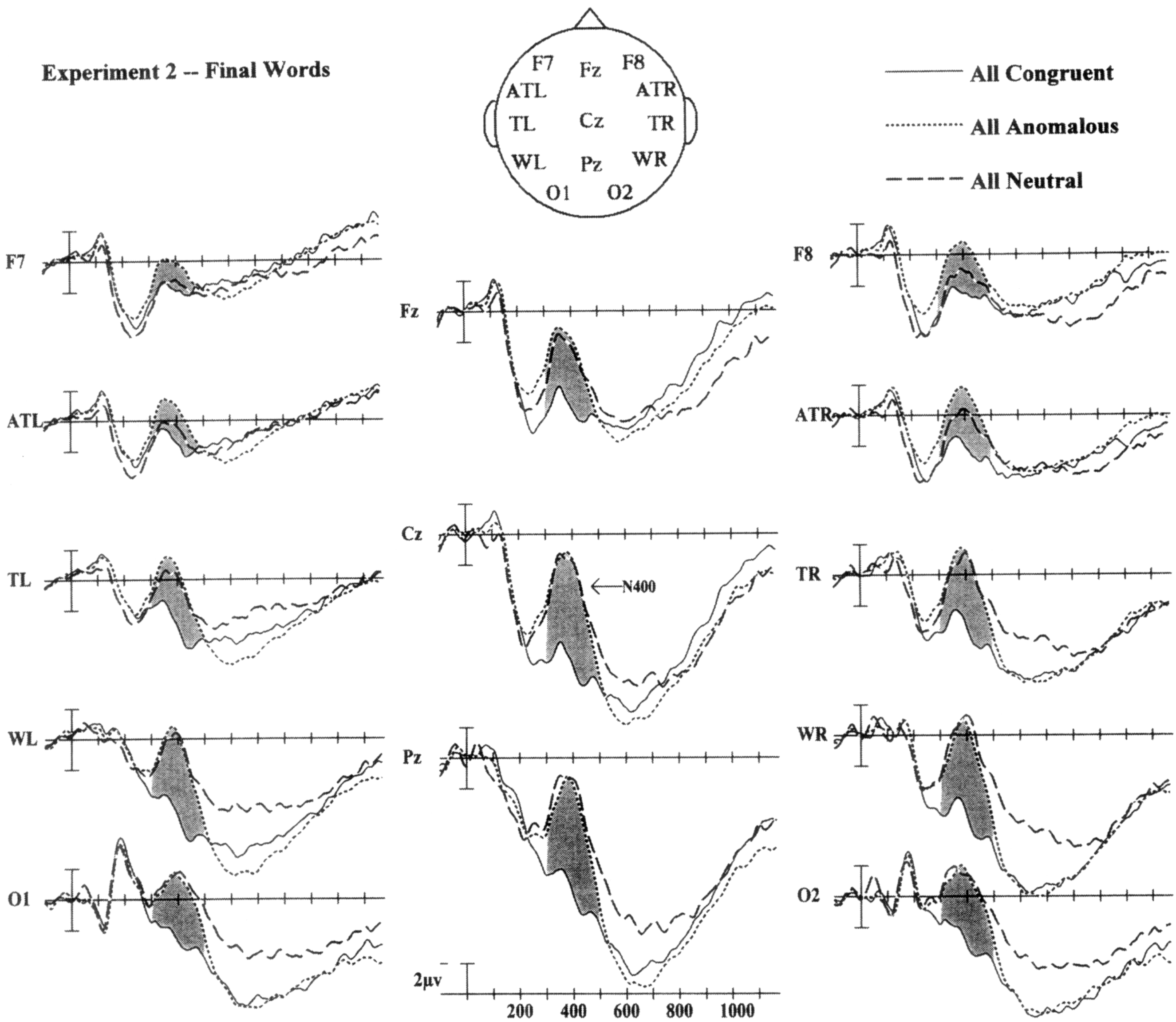

Figure 5. Grand mean event-related potentials for congruent, anomalous, and neutral final words in Experiment 2.

negative-going ERPs than abstract words, but only for the anomalous and neutral conditions: main effect of word type, anomalous, midline, $F(1, 23) = 32.33$, $MSE = 8.7$, $p < .0001$; lateral, $F(1, 23) = 58.17$, $MSE = 10.5$, $p < .0001$; neutral, midline, $F(1, 23) = 27.91$, $MSE = 13.8$, $p < .0001$; lateral, $F(1, 23) = 28.52$, $MSE = 18.9$, $p < .0001$; congruent, $ps > .5$ (compare Figures 6, 7, and 8). Furthermore, for anomalous and neutral endings, these effects of concreteness tended to increase in magnitude toward more anterior sites: Word Type × Electrode Site interaction, anomalous, midline, $F(2, 46) = 5.13$, $MSE = 0.1$, $p < .05$; lateral, $F(4, 92) = 16.31$, $MSE = 0.2$, $p < .0001$; neutral, midline, $F(2, 46) = 3.43$, $MSE = 1.0$, $p < .05$; lateral, $p < .1$. However, for congruent endings, there were differences between the two word types only at the most posterior lateral sites (O1/O2) where concrete endings were more negative-going than abstract endings: Word Type × Electrode Site interaction, congruent, lateral, $F(4, 92) = 5.68$, $MSE = 0.2$, $p < .01$.

In the 500–800 ms epoch, there was again a main effect of sentence type: midline, $F(2, 46) = 8.58$, $MSE = 19.53$, $p < .001$; lateral, $F(2, 46) = 13.88$, $MSE = 30.17$, $p < .001$, and an interaction between sentence type and electrode site, midline, $F(4, 92) = 10.10$, $MSE = 0.10$, $p < .0001$; lateral, $F(8, 184) = 19.54$, $MSE = 0.30$, $p < .0001$. Follow-up analyses revealed that neutral sentence final words elicited more negative ERPs than either congruent or anomalous final words: sentence type, neutral versus congruent, midline, $F(1, 23) = 6.71$, $MSE = 21.81$, $p < .05$; lateral, $F(1, 23) = 12.66$, $MSE = 41.10$, $p < .01$; neutral versus

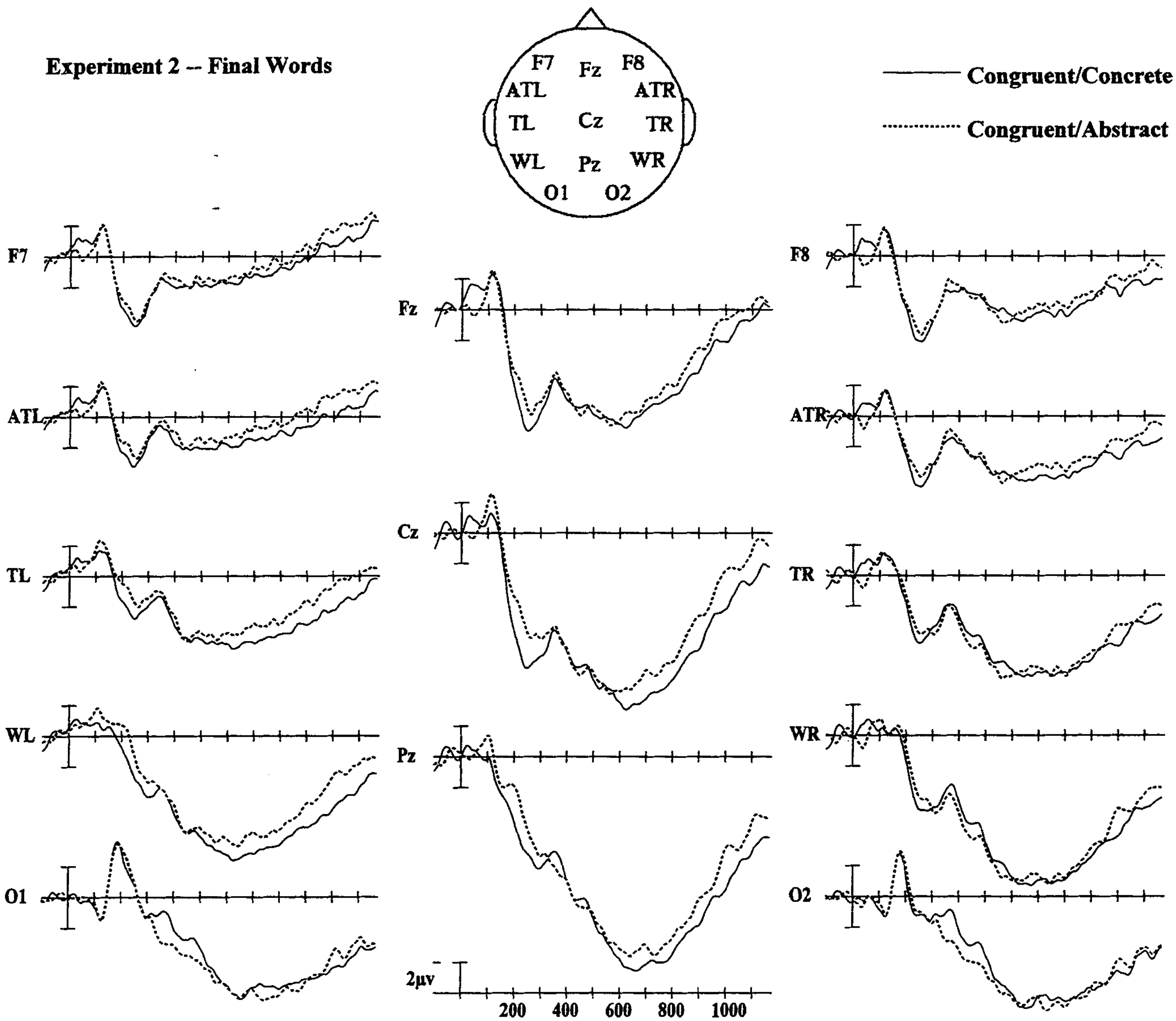

Figure 6. Grand mean event-related potentials for congruent final words that were concrete or abstract (Experiment 2).

anomalous, midline, $F(1, 23) = 18.16$, $MSE = 17.71$, $p < .001$; lateral, $F(1, 23) = 23.38$, $MSE = 30.79$, $p < .0001$; and these differences tended to have a posterior distribution: Sentence Type × Electrode Site interaction, neutral versus congruent, midline, $F(2, 46) = 12.62$, $MSE = 0.11$, $p < .001$; lateral, $F(4, 92) = 20.44$, $MSE = 0.31$, $p < .0001$; neutral versus anomalous, midline, $F(2, 46) = 18.14$, $MSE = 0.09$, $p < .0001$; lateral, $F(4, 92) = 39.51$, $MSE = 0.19$, $p < .0001$.

During the 500–800 ms epoch, there was also a main effect of word type: midline, $F(1, 23) = 10.75$, $MSE = 19.5$, $p < .01$; lateral, $F(1, 23) = 17.20$, $MSE = 28.0$, $p < .001$; a Sentence Type × Word Type interaction, midline, $F(2, 46) = 11.94$, $MSE = 11.9$, $p < .001$; lateral, $F(2, 46) = 11.50$, $MSE = 15.7$, $p < .001$; and, at lateral sites, a three-way interaction between sentence type, word type, and electrode site, $F(8, 184) = 5.28$, $MSE = 0.12$, $p < .01$. Follow-up analyses indicated that concrete words were more negative-going than abstract words for anomalous and neutral, but not for congruent final sentences: word-type main effect, anomalous, midline, $F(1, 23) = 18.83$, $MSE = 6.24$, $p < .001$; lateral, $F(1, 23) = 33.60$, $MSE = 9.75$, $p < .0001$; neutral, midline, $F(1, 23) = 16.95$, $MSE = 20.92$, $p < .001$; lateral, $F(1, 23) = 21.44$, $MSE = 23.76$, $p < .0001$; congruent $ps > .25$. Moreover, the concreteness effects for anomalous and neutral final words tended to be larger over more anterior sites: Word Type × Electrode Site interaction, anomalous, midline, $F(2, 46) = 9.59$, $MSE = 1.49$, $p < .01$; lateral, $F(4, 92) = 18.18$, $MSE = 1.27$, $p < .01$; neutral, midline, $F(2, 46) = 3.34$, $MSE = 2.13$, $p < .05$;

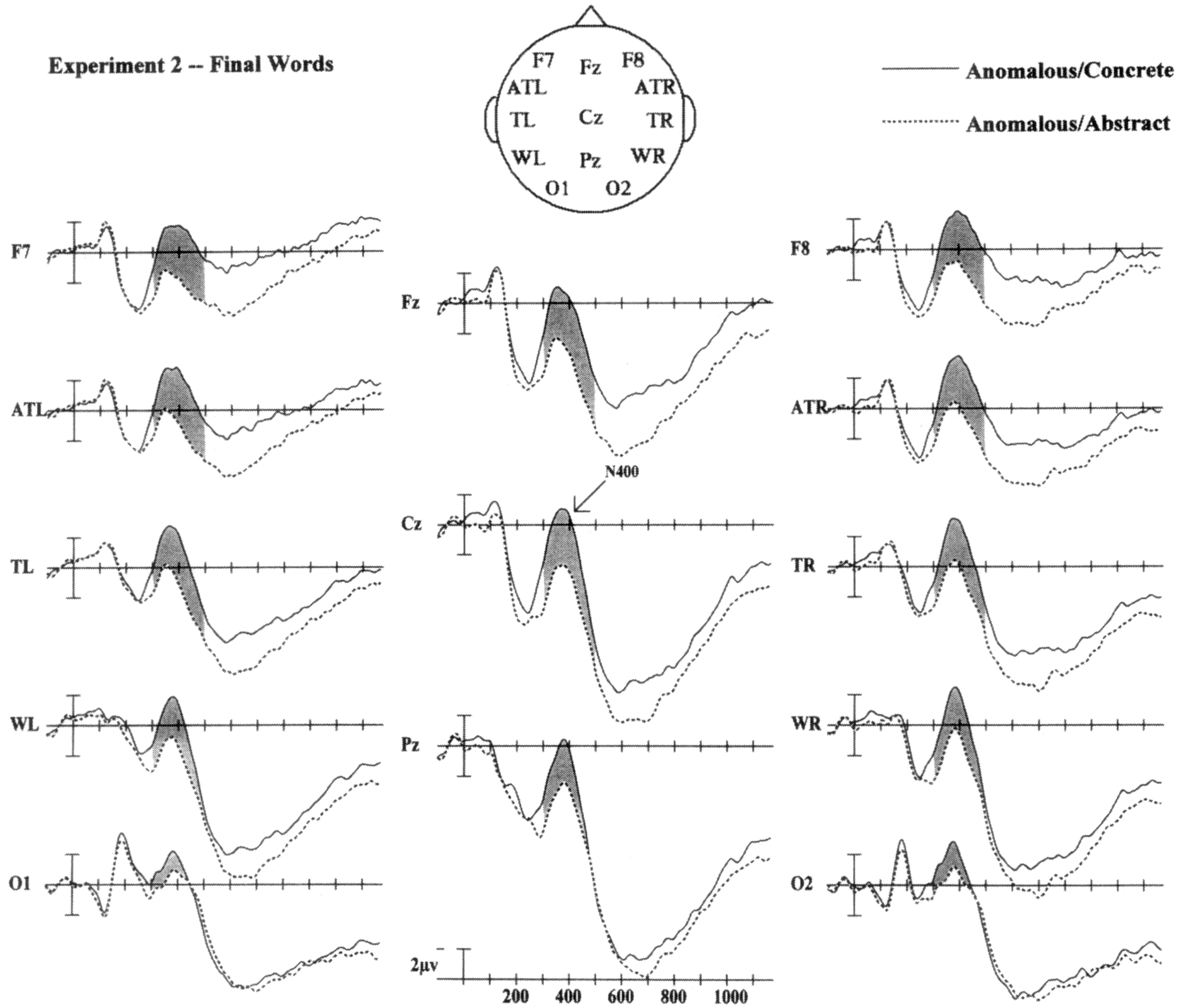

Figure 7. Grand mean event-related potentials for anomalous final words that were concrete or abstract (Experiment 2).

lateral, $F(4, 92) = 2.3$, $MSE = 2.69$, $p < .12$; congruent $ps > .13$.

Discussion

Experiment 2 was designed to test whether the concreteness effects observed for anomalous sentences in the first experiment were due to the concrete final words being more anomalous than abstract final words. Neutral sentences that were congruent but had a low cloze probability were compared to the congruent and anomalous sentences used in Experiment 1. ERPs to neutral sentences were similar to ERPs to congruent sentences during the early 150–300 ms window, which overlaps with the time course of the P2 component. During this epoch, anomalous final words elicited more negative waveforms, particularly at anterior sites, than either congruent or neutral final words. Conversely, during the 300–500 ms time window, ERPs to neutral final words resembled the ERPs to anomalous final words. Both neutral and anomalous final words elicited more negative waveforms than congruent final words during this epoch. Consistent with the topography of the N400, this effect for both neutral and anomalous words was maximal at central to posterior scalp sites. During the 500–800 ms time window, ERPs to neutral sentences diverged from both anomalous- and congruent-sentence ERPs, producing a more negative (i.e., less positive) waveform.

Relative to congruent sentences, both anomalous and neutral sentences produced concreteness effects during all three time windows and these effects tended to increase in magnitude toward more anterior scalp sites. Because ERP

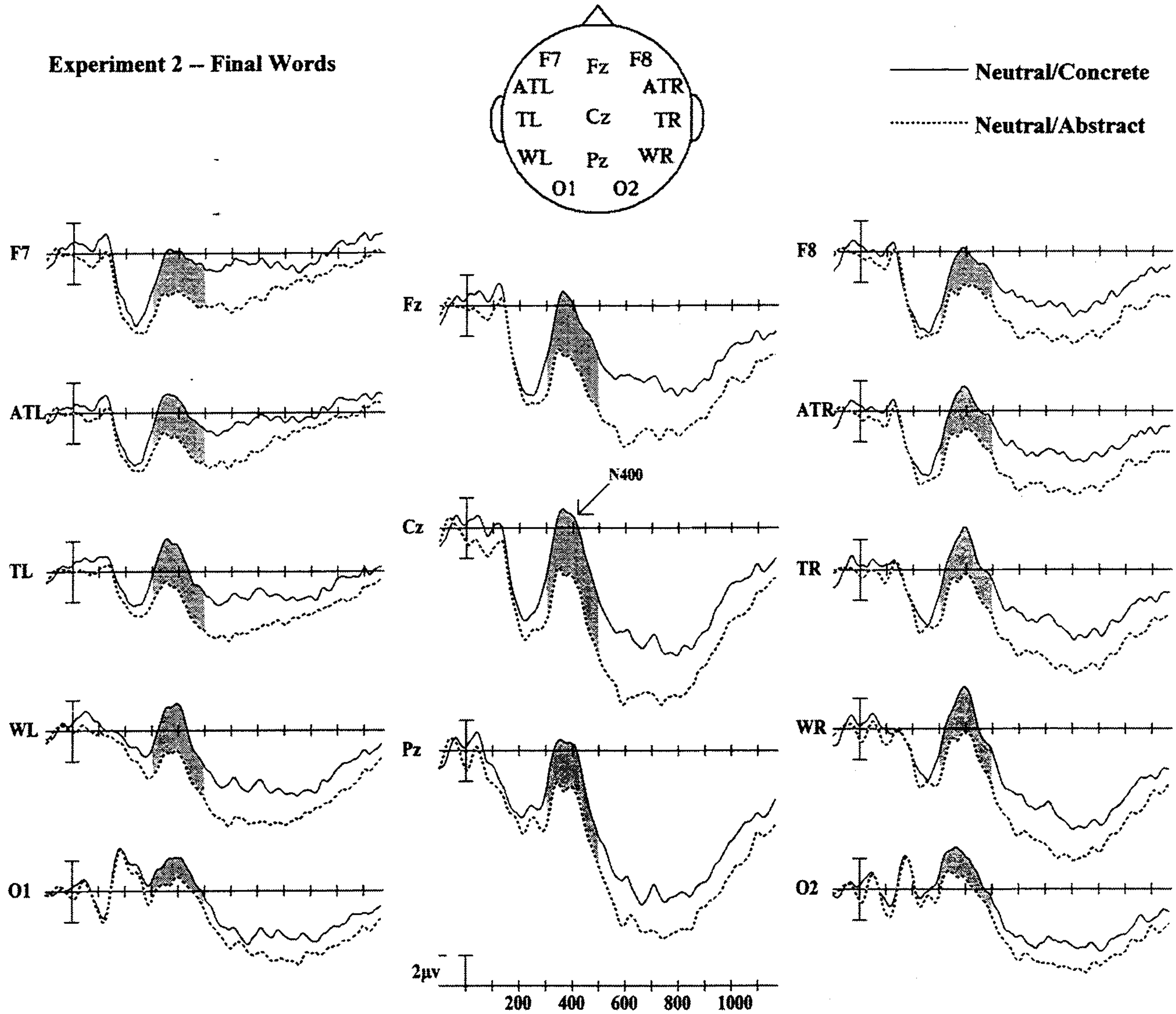

Figure 8. Grand mean event-related potentials for neutral final words that were concrete or abstract (Experiment 2).

responses to neutral sentences were similar to responses to anomalous sentences (both having low cloze probability for the final words), it seems reasonable to infer that the N400 concreteness effects observed in Experiment 1 did not occur because the concrete words seemed more anomalous than the abstract words. Rather, concreteness and cloze probability interacted, such that there was a difference in the processing of concrete words and abstract words in low cloze probability contexts. In contrast, as in Experiment 1, concreteness effects were eliminated in high cloze probability contexts.

The differences in the ERPs to anomalous and neutral sentences during the early (150–300 ms) epoch suggest that effects specifically related to congruency began before effects related to concreteness or cloze probability. In addition this finding suggests that congruency and concreteness are separate variables and implies that semantically anomalous words not only produce larger amplitude N400s than low cloze probability (but semantically congruous) words (Kutas et al., 1984), but that these N400s may also have an earlier onset latency.

Finally, both anomalous- and congruent-sentence final words displayed a large positivity during the 500–800 ms time window. However, the neutral-sentence final words had a greatly attenuated late positivity. The neutral sentences also were associated with significantly lower accuracy responses. One interpretation of both of these findings argues that the effects are due to differences in ambiguity resolution. This argument assumes that the late positivity in the congruent and anomalous sentences reflects, in part, activity associated with the P3 component (e.g., Donchin, 1981). It has been argued that this positivity is modulated by the degree of ambiguity resolution associated with a decision about task-relevant items (Johnson, 1986), the more ambiguity resolved the greater the positivity. Within this framework it could be argued that both congruent and anomalous final words resolved more ambiguity than did neutral final words. This could be because, for example, congruent and anomalous sentence stems provided a very constrained context, whereas neutral sentences did not.

One explanation for the higher error rates for neutral sentences in this experiment is that participants may have found some of these items relatively more anomalous than congruent sentences, and thus the "no" responses to these items were actually not errors, but rather reflected participants' comprehension problems. If true, this might suggest that our attempts to control for degree of anomaly were not successful and that the ERP results for the neutral sentences suffer from the same interpretive problem as the frank anomalies in Experiment 1. This possibility seems plausible given that significantly more concrete neutral items were responded to with a "no" response than were abstract neutral items. However, we think this interpretation is incorrect for at least two reasons. First, although there were more errors for the neutral conditions, the ERP results were based only on trials with correct responses. So, in the case of the neutral condition these were presumably reflecting only trials where participants found these sentences to make sense. However, it could be argued that with error rates hovering around 20%, even the correct trials might be contaminated by a significant number of lucky guesses or low confidence responses. Second, and more compelling, subsequent analyses splitting the 24 participants at the median (high accuracy vs. low accuracy) revealed that accuracy could not account for the obtained concreteness effects. Specifically, among the high accuracy group (mean accuracy for neutral concrete = 92.5, abstract = 96) there were no significant accuracy effects for sentence type, Sentence Type × Word Type, or between the concrete and abstract words within the neutral condition (all $ps > .35$). This suggests that this subgroup was classifying the neutral sentences in the manner we had intended. Therefore, to determine if the N400 to the neutral concrete condition might have been due to the significantly lower accuracy in the bottom 50% of participants (with means of 71.7% for neutral concrete and 81.7% for abstract), a separate set of analyses on the 300–500 ms (N400) window were run on the 12 participants with the highest accuracy and the 12 with the lowest accuracy. None of these analyses produced any significant concreteness differences between the high and low accuracy groups (all midline and lateral $ps > .40$). Moreover, both groups showed the same pattern of results as reported above for the combined groups. Most important, concrete neutral endings produced significantly larger, more anterior negativities than abstract neutral endings in both groups (see Figure 9). These supplementary analyses make it clear that the similarities in the N400 window between anomalous and neutral endings cannot be attributed to differences in response accuracy and buttress the claim that the concreteness effects reported in this experiment and in Experiment 1 are due to inherent differences in the processing of the two word types and not to degree of anomaly.

An alternative explanation of the higher error rates for neutral sentences is the integration difficulty strategy that was offered in the *Discussion* of Experiment 1 to account for the discrepancy between RT and N400 effects. In neutral sentences, final words were relatively difficult to integrate into the prior sentence context (as evidenced by the large N400). If participants were using the integration strategy, they may have based their relatively high proportion of erroneous responses on this information. In other words, integration difficulty may have initially signaled to participants that the appropriate response was "no—this is a meaningless sentence." To respond "yes" (the appropriate response), they would have had to overcome this tendency. The higher error rates for concrete neutral sentences support this interpretation, as according to the integration strategy concrete words signal a higher level of integration difficulty (because they have more semantic information to integrate), and therefore more strongly suggest that this is a meaningless sentence. Also consistent with this view, neutral concrete words produced a larger N400 than abstract neutral words.

To summarize, this experiment demonstrated that differences in N400 amplitude in response to anomalous concrete and abstract sentence final words were not due to differences in the degree of anomaly between the word types. Concreteness appears to influence final-word N400 amplitude when these words have low expectancy or cloze probability. In cases where the expectancy is high, concreteness effects are not manifested. This conclusion not only supports the hypothesis that the N400 is sensitive to word expectancy (Kutas et al., 1984), but that it can also be modulated by other variables such as concreteness value.

General Discussion

Continuing along the lines of our earlier study (Kounios & Holcomb, 1994), in the current experiments we examined ERPs during the processing of concrete and abstract words preceded by sentence contexts. The experimental design also incorporated the independent manipulation of the congruency of the final word with its preceding context. The N400 wave of the ERP was the particular focus of this investigation, as prior research has shown this component to reflect the influence of both contextual and structural semantic factors (e.g., Osterhout & Holcomb, 1995). It was shown that both concreteness and context influence neural processes during sentence processing, in particular those manifested by the N400. More specifically, when the sentences were anomalous, concrete final words elicited a larger N400 than abstract final words, with the effect of concreteness on the N400 exhibiting a scalp topography different from (i.e., more anterior than) the effect of congruency. Furthermore, a similar N400 concreteness effect was observed (in Experiment 2) even for neutral sentences that were congruous but did not provide any supportive context for the final word, indicating that the concreteness effect observed for anomalous sentences could not be attributed

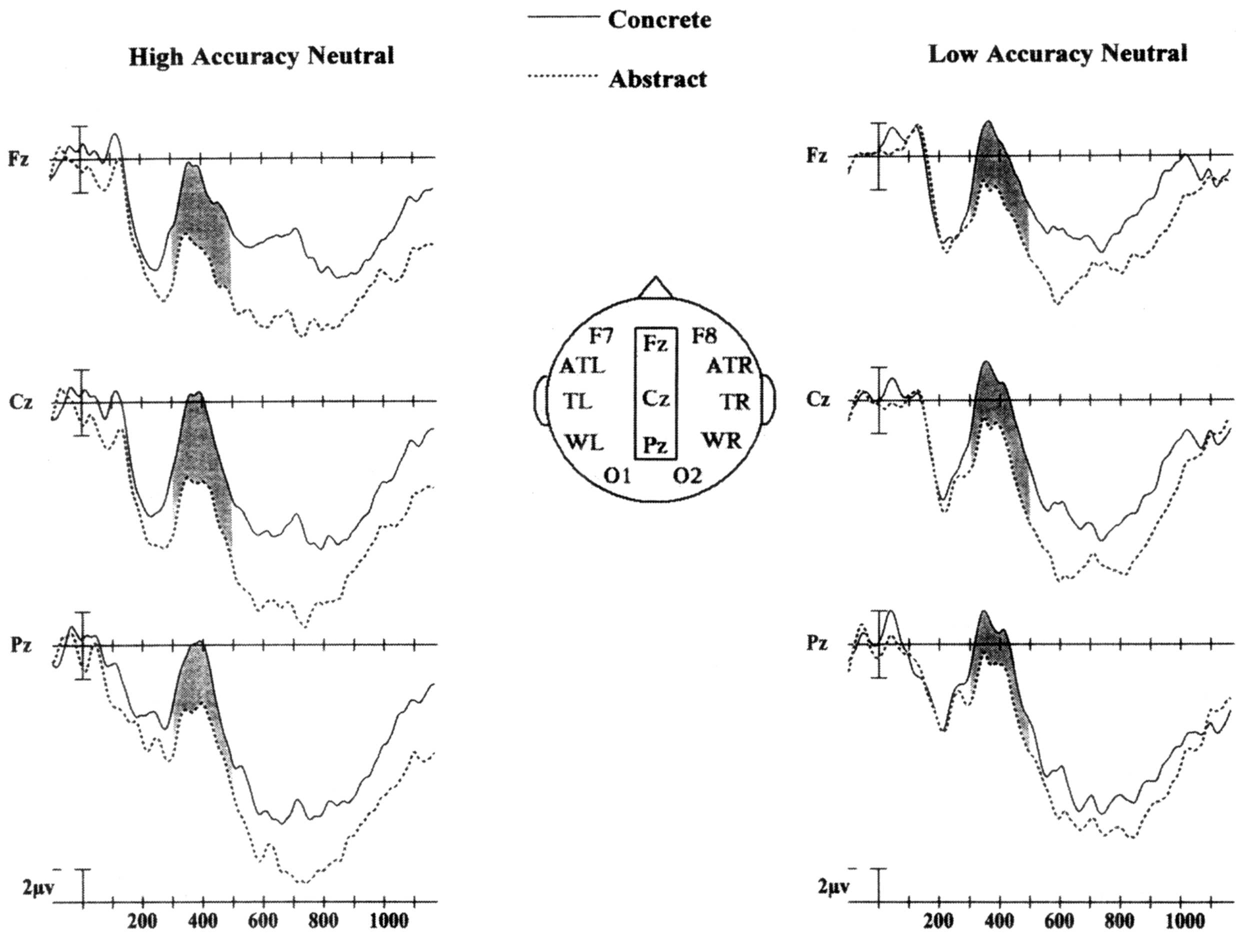

Figure 9. Grand mean event-related potentials at midline sites for neutral concrete and neutral abstract final words for the 12 most accurate participants (left) and the 12 least accurate participants (right) in Experiment 2.

to differences between concrete and abstract final words in their degree of anomaly within the provided contexts. These findings are consistent with the context extensions proposed for dual-coding theory in the introduction, but are difficult to reconcile with single-code theories such as context availability.

Although the different scalp distributions of the N400 for concrete and abstract words appear to be most consistent with a dual-coding type of account of semantic memory representation, the current data do not conclusively specify the nature of the differential processes or representations for the two word types. The findings are consistent with at least three different cognitive–neural architectures: multiple stores and a single semantic integration process, multiple stores and multiple integration processes, or a single store with multiple integration processes. Note that the latter possibility is a variant of single-code theory, which if correct would weaken the claims made here about the implausibility of such a representational architecture. However, we think the likelihood that this type of architecture is the "correct one" is lower because of difficulties inherent in specifying how separate processes would differentially access a common memory system.[5]

Our working hypothesis is that there is a family of N400s

[5] In our view, the dual process, single representation system possibility is not worthy of the same level of consideration as the alternative dual representation architectures. This is because for two separate processes to access a common memory system the two processes would have to either extract qualitatively different information from memory or extract the same information but deal with it differently. Either of these is possible, although careful consideration of the former reveals that it boils down to a variant of dual representation. This is because concrete concepts can be distinguished from abstract concepts on the basis of their qualitatively different representations, which under the spatial distinctiveness principle would be expected to produce different scalp distribution of ERPs. The latter possibility, that qualitatively similar information is processed differently by two distinct processing systems, also has problems in that there would have to be some representational criteria that allow the two processing systems to make a distinction between representations. These differences would also be expected to reveal different ERP scalp distributions.

that reflect activity in multiple semantic memory processes and multiple information stores. One tentative possibility for the pattern of N400s in this study is that there is a posterior "linguistically sensitive N400," which is activated by both word types and a more frontal "imagistically sensitive N400," which is activated relatively more by concrete words.[6] Evidence for or against this hypothesis will have to await the outcome of future studies, although it is interesting that work with pictorial stimuli has suggested that ERP negativities sensitive to contextual manipulations also tend to have a more anterior focus (e.g., Barrett & Rugg, 1990; Ganis et al., 1996; Holcomb & McPherson, 1994; McPherson & Holcomb, in press). McPherson and Holcomb have suggested that part of the more anterior semantic effect with pictures is due to a separate ERP component (the N300), which, for pictures, begins about 100 ms earlier than the more posterior N400. Although there was no evidence here of the anterior concreteness effect starting earlier than the posterior linguistic effect, this could be because processing in the anterior system must await input via referential connections from the linguistic system in the case of concrete words (pictures would presumably directly activate the imagistic system). In future studies it will be interesting to test the N300/N400 hypothesis of the anterior concreteness/picture priming effects by manipulating the modality of the contextual and target stimuli.

The data from these experiments also imply that the presence of a supportive context can short-circuit or override the N400 concreteness effect as the congruous sentences in both experiments showed no significant effect of the concreteness manipulation in any of the time windows examined (ranging from 150 to 800 ms after final-word onset). So how does context achieve this effect? One possibility is that the contextual manipulation did not actually prevent an effect of concreteness on the ERP. It could be argued that context and concreteness both influenced the same processes in this study, but that the effect of context was so much larger than that of concreteness that the latter effect could not be discerned. In other words, the failure to find a concreteness effect for the congruent sentences might have been due to a poor signal (concreteness) to noise (context) ratio. However, this explanation is relatively implausible because no concreteness effects were observed for congruous sentences during any epoch examined. Such an explanation rests on the restrictive assumption that concreteness and context can influence only a common pool of processes (at least from the subset of processes reflected by the ERP technique); otherwise, concreteness would have had an effect on at least one ERP component for congruent sentences.

A different sort of explanation provides a more plausible account of the absence of concreteness effects for congruent sentences. The time courses of the concreteness and context effects overlapped to a substantial degree, but were not identical. The concreteness effect started later in Experiment 1 (300–500 ms window) and lasted longer in both experiments (500–800 ms window) than the context effect, which started in the 150–300 ms window and was over at all but the right anterior sites by the end of the 300–500 ms window in both experiments. Although far from clear, on the basis of the data from these two experiments, the time course differences suggest the possibility that a supportive context overcomes concreteness effects in part by starting earlier. But how could this work?

The most likely possibility is that a supportive context works by laying down a contextual representation prior to the onset of the last word (e.g., Gernsbacher, 1990; Hess, Foss, & Carrol, 1995). In this scheme, when a contextually appropriate final word occurs, its semantic properties are effortlessly integrated into the framework. When an inappropriate word occurs, its semantic properties conflict with those of the sentence and the reader either gives up, or, more likely, attempts to determine if there is a way to fit the semantic properties of the inappropriate word into the meaning of the sentence (e.g., perhaps an inference is required for this word to make sense). We think the latter is the more likely explanation for the N400 in cases involving a sentence context, because words in natural language are rarely as predictable as the high cloze words in this study. We also think this is the reason why concrete words generate a larger and more anterior N400 effect. This is because when initial integration fails, all of the semantic information associated with the word potentially becomes relevant. In this situation, readers might be attempting to integrate much or all of the semantic properties of the item in an attempt to make sense of the sentence. This could include both verbal–linguistic and imagistic information. This explanation also works for the neutral condition of Experiment 2, which also produced large N400s. These sentences, like the anomalies, did not strongly suggest which semantic property of the target word is most appropriate for a correct reading of the sentence and therefore multiple representations might become candidates for integration.

In conclusion, this study has provided data in support of two main points. First, concreteness effects in language comprehension are not reducible to differences in supportive context (Kounios & Holcomb, 1994; Paivio, 1986, 1991), indicating that both contextual and structural variables must be considered to play a role in language comprehension (Kounios, 1996). Second, the different ERP results for concrete and abstract words (in the absence of a supportive context) argue against the view that the semantic system is unitary and amodal (e.g., Caramazza et al., 1990); instead, these results are indicative of multiple semantics of an (as yet) undetermined level of complexity (Saffran & Schwartz, 1994).

Finally, though the present study has provided some clarity to a difficult issue, many important questions remain. First, the notion of multiple systems (e.g., Paivio, 1986) suggests the existence both of processes that operate between these systems (e.g., translation) and processes that operate on representations only within a particular system (e.g., mental rotation for pictorial representations), as well as

[6] Alternatively, there may be but one N400 process with a similar distribution for both word types. What may differentiate concrete and abstract words are one or more other components that temporally overlap with and therefore (at the scalp) sum with the N400 appearing to change its distribution.

the possibility of processes that can operate on any kind of representation (e.g., similarity judgments). More attention should be given to isolating the relevant types of processes and determining their architecture.

A second issue is that it is yet to be determined why the concreteness effects reported here and in Kounios and Holcomb (1994) have a relatively bilateral frontal distribution. If this activity is specific to the concrete or possibly the imaginal properties of words, it is unclear (to us) what neural system would be mediating such a pattern. Part of the problem is that the ERP technique has a limited ability to localize effects within the brain (i.e., it has relatively poor spatial resolution), therefore making it difficult to pinpoint the structures generating the effects of interest. Although a number of functional imaging studies, which have greater spatial resolution, have implicated specific frontal areas in semantic processing, most of these studies have found this activity to be lateralized primarily to the left hemisphere and have not differentiated between different semantic categories (e.g., Binder et al., 1997; Demb et al., 1995; Peterson, Fox, Posner, Mintum, & Racichle, 1988). Our effects are clearly bilateral and emerge as a dissociation between word types. Moreover, many of the functional imaging studies mentioned above suffer from serious design flaws (e.g., the use of inappropriate subtraction techniques or in some cases generation rather than comprehension tasks) as well as limitations in temporal resolution (the best functional magnetic resonance imaging [FMRI] studies sum activity over several seconds of cognitive processing) rendering their findings of dubious value for understanding the microstructure of cognitive operations such as word processing. This and other issues discussed above may be overcome in future studies if better ways of combining techniques with good temporal but poor spatial resolution (e.g., ERPs) and better spatial, but poorer temporal characteristics (e.g., FMRI), can be found.

References

Barrett, S. E., & Rugg, M. D. (1990). Event-related potentials and the semantic matching of pictures. *Brain & Cognition, 14,* 201–212.

Binder, J. R., Frost, J. A., Hammeke, T. A., Cox, R. W., Rao, S. M., & Prieto, T. (1997). Human brain language areas identified by functional magnetic resonance imaging. *Journal of Neuroscience, 17,* 353–362.

Boddy, J. (1986). Event-related potentials in chronometric analysis of primed word recognition with different stimulus onset asynchronies. *Psychophysiology, 23,* 232–245.

Bransford, J. D., & McCarrell, N. S. (1974). A sketch of a cognitive approach to comprehension: Some thoughts on what it means to comprehend. In W. Weimer & D. Palermo (Eds.), *Cognition and the symbolic processes* (pp. 189–230). Hillsdale, NJ: Erlbaum.

Brown, C., & Hagoort, P. (1993). The processing nature of the N400: Evidence from masked priming. *Journal of Cognitive Neuroscience, 5,* 34–44.

Caramazza, A., Hillis, A. E., Rapp, B. C., & Romani, C. (1990). The multiple semantics hypothesis: Multiple confusions? *Cognitive Neuropsychology, 7*(3), 161–189.

Coles, M. G. H., & Rugg, M. D. (1995). Event-related potentials: An introduction. In M. D. Rugg & M. G. H. Coles (Eds.), *Electrophysiology of mind* (pp. 1–26). New York: Oxford University Press.

Demb, J. B., Desmond, J. E., Wanger, A. D., Vaidya, C. J., Glover, G. H., & Gabrieli, J. D. E. (1995). Semantic encoding and retrieval in the left inferior prefrontal cortex: A functional MRI study of task difficulty and process specificity. *Journal of Neuroscience, 15,* 5870–5878.

Donchin, E. (1981). "Surprise! . . . Surprise?." *Psychophysiology 18,* 493–513.

Francis, W. N., & Kučera, H. (1982). *Frequency analysis of English usage.* Boston: Houghton Mifflin.

Ganis, G., Kutas, M., & Sereno, M. I. (1996). The search for "common sense": An electrophysiological study of the comprehension of words and pictures in reading. *Journal of Cognitive Neuroscience, 8,* 89–106.

Geisser, S., & Greenhouse, S. (1959). On methods in the analysis of profile data. *Psychometrika, 24,* 95–112.

Gernsbacher, M. A. (1985). Surface information loss in comprehension. *Cognitive Psychology, 17,* 324–363.

Gernsbacher, M. A. (1990). *Language comprehension as structure building.* Hillsdale, NJ: Erlbaum.

Hess, D. J., Foss, D. J., & Carrol, P. (1995). Effects of global and local context on lexical processing during language comprehension. *Journal of Experimental Psychology: General, 124,* 62–82.

Holcomb, P. J. (1993). Semantic priming and stimulus degradation: Implications for the role of the N400 in language processing. *Psychophysiology, 30,* 47–61.

Holcomb, P. J., & McPherson, W. B. (1994). Event-related brain potentials reflect semantic priming in an object decision task. *Brain and Cognition, 24,* 259–274.

Johnson, R., Jr. (1986). A triarchic model of P300 amplitude. *Psychophysiology, 23,* 367–384.

Kieras, D. (1978). Beyond pictures and words: Alternative information processing models for imagery effects in verbal memory. *Psychological Bulletin, 85,* 532–554.

Kolb, B., & Whishaw, I. Q. (1996). *Fundamentals of human neuropsychology.* New York: Freeman.

Kounios, J. (1996). On the continuity of thought and the representation of knowledge: Electrophysiological and behavioral time-course measures reveal levels of structure in semantic memory. *Psychonomic Bulletin & Review, 3,* 265–286.

Kounios, J., & Holcomb, P. J. (1992). Structure and process in semantic memory: Evidence from event-related brain potentials and reaction times. *Journal of Experimental Psychology: General, 121,* 459–479.

Kounios, J., & Holcomb, P. J. (1994). Concreteness effects in semantic processing: ERP evidence supporting dual-coding theory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 20,* 804–823.

Kroll, J. F., & Potter, M. C. (1984). Recognizing words, pictures, and concepts: A comparison of lexical, object, and reality decisions. *Journal of Verbal Learning and Verbal Behavior, 23,* 39–66.

Kutas, M. (1987). Event-related brain potentials elicited during rapid serial visual presentation of congruous and incongruous sentences. In R. Johnson, J. W. Rohrbaugh, & R. Parasuraman (Eds.), *Current trends in event-related potential research* (pp. 406–411). Amsterdam: Elsevier Science.

Kutas, M., & Hillyard, S. A. (1980, Jan. 11). Reading senseless sentences: Brain potentials reflect semantic incongruity. *Science, 207,* 203–205.

Kutas, M., & Hillyard, S. A. (1983). Event-related brain potentials to grammatical errors and semantic anomalies. *Memory & Cognition, 11,* 539–550.

Kutas, M., & Hillyard, S. A. (1984, Jan. 12). Brain potentials during reading reflect word expectancy and semantic association. *Nature, 307,* 161–163.

Kutas, M., Lindamood, T., & Hillyard, S. A. (1984). Word expectancy and event-related brain potentials during sentence processing. In S. Kornblum & J. Requin (Eds.), *Preparatory states and processes* (pp. 217–238). Hillsdale, NJ: Erlbaum.

Kutas, M., & Van Petten, C. (1988). Event-related brain potential studies of language. In P. K. Ackles, J. R. Jennings, & M. G. H. Coles (Eds.), *Advances in psychophysiology* (pp. 139–187). Greenwich, CT: JAI Press.

McCarthy, G., & Wood, C. C. (1985). Scalp distributions of event-related potentials: An ambiguity associated with analysis of variance models. *Electroencephalography and Clinical Neurophysiology, 88,* 203–208.

McClelland, J. L. (1979). On the time relations of mental processes: An examination of systems of processes in cascade. *Psychological Review, 86,* 287–330.

McPherson, W. B., & Holcomb, P. J. (in press). An electrophysiological investigation of semantic priming with pictures of real objects. *Psychophysiology.*

Neely, J. H. (1991). Semantic priming effects in visual word recognition: A selective review of current findings and theories. In D. Besner & G. Humphreys (Eds.), *Basic processes in reading: Visual word recognition* (pp. 264–336). Hillsdale, NJ: Erlbaum.

Neville, H. J., Kutas, M., Chesney, G., & Schmidt, A. L. (1987). Event-related brain potentials during encoding and recognition memory of congruous and incongruous words. *Journal of Memory and Language, 25,* 75–92.

Nobre, A. C., & McCarthy, G. (1994). Language-related ERPs: Scalp distributions and modulation by word type and semantic priming. *Journal of Cognitive Neuroscience, 6,* 233–255.

Osterhout, L., & Holcomb, P. J. (1995). Event-related potentials and language comprehension. In M. Rugg & M. G. H. Coles (Eds.), *Electrophysiology of mind* (pp. 171–215). Oxford, England: Oxford University Press.

Paivio, A. (1971). *Imagery and verbal processes.* New York: Holt, Rinehart & Winston.

Paivio, A. (1986). *Mental representations: A dual coding approach.* New York: Oxford University Press.

Paivio, A. (1991). Dual coding theory: Retrospect and current status. *Canadian Journal of Psychology, 45,* 255–287.

Peterson, S. E., Fox, P. T., Posner, M. I., Mintum, M., & Racichle, M. E. (1988, Feb. 18). Positron emission tomographic studies of the cortical anatomy of single word processing. *Nature, 331,* 585–589.

Pylyshyn, Z. W. (1984). *Computation and cognition: Toward a foundation for cognitive science.* Cambridge, MA: MIT Press.

Rugg, M. D., & Coles, M. G. H. (1995). The ERP and cognitive psychology: Conceptual issues. In M. D. Rugg & M. G. H. Coles (Eds.), *Electrophysiology of mind* (pp. 27–39). Oxford, England: Oxford University Press.

Rummelhart, D., & McClelland, J. (1982). An interactive activation model of context effects in letter perception: Part 2. The contextual enhancement effect and some tests and extensions of the model. *Psychological Review, 89,* 60–94.

Saffran, E. M., & Schwartz, M. F. (1994). Of cabbages and things: Semantic memory from a neuropsychological perspective—A tutorial review. In C. Umilta & M. Moscovitch (Eds.), *Attention and performance XV: Conscious and nonconscious information processing* (pp. 508–536). Cambridge, MA: MIT Press.

Schwanenflugel, P. (1991). Why are abstract concepts hard to understand? In P. J. Schwanenflugel (Ed.), *The psychology of word meanings* (pp. 223–250). Hillsdale, NJ: Erlbaum.

Schwanenflugel, P., Harnishfeger, K. K., & Stowe, R. W. (1988). Context availability and lexical decisions for abstract and concrete words. *Journal of Memory and Language, 27,* 499–520.

Schwanenflugel, P., & Shoben, E. (1983). Differential context effects in the comprehension of abstract and concrete verbal materials. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 9,* 82–102.

Schwanenflugel, P., & Stowe, R. W. (1989). Context availability and the processing of abstract and concrete words in sentences. *Reading Research Quarterly, 24,* 114–126.

Shallice, T. (1988). Specialization within the semantic system. *Cognitive Neuropsychology, 5,* 133–142.

Shallice, T. (1993). Multiple semantics: Whose confusions? *Cognitive Neuropsychology, 10,* 251–261.

Stanovich, K. E., & West, R. F. (1983). On priming by a sentence context. *Journal of Experimental Psychology: General, 112,* 1–136.

Sternberg, S. (1969). The discovery of processing stages: Extensions of Donders' method. In W. G. Koster (Ed.), *Attention and performance II* (pp. 276–315). Amsterdam: North-Holland.

West, W. C., & Holcomb, P. J. (1998). Imaginal, semantic and surface-level processing of concrete and abstract words: An electrophysiological investigation. Manuscript submitted for publication.

Received April 7, 1997
Revision received November 30, 1998
Accepted November 30, 1998 ■