Meaning and Modality: Influences of Context, Semantic Memory Organization, and Perceptual Predictability on Picture Processing

Kara D. Federmeier and Marta Kutas
University of California, San Diego

Using event-related potentials (ERPs), the authors investigated the influences of sentence context, semantic memory organization, and perceptual predictability on picture processing. Participants read pairs of highly or weakly constraining sentences that ended with (a) the expected item, (b) an unexpected item from the expected semantic category, or (c) an unexpected item from an unexpected category. Pictures were unfamiliar in Experiment 1 but preexposed in Experiment 2. ERPs to pictures reflected both contextual fit and memory organization, as do ERPs to words in the same contexts (K. D. Federmeier & M. Kutas, 1999). However, different response patterns were observed to pictures than to words. Some of these arose from perceptual predictability differences, whereas others seem to reflect true modality-based differences in semantic feature activation. Although words and pictures may share semantic memory, the authors' results show that semantic processing is not amodal.

Words (visual or auditory) and pictures are both physical objects that, through experience, have come to be associated with information not explicitly contained in the physical form of the word or picture itself. In this sense, both pictures and words can be thought of as symbols, or objects that "stand for" information that they do not directly represent. Of course, not all words have clear pictorial counterparts (e.g., function words), and in many cases it would take multiple words to convey the information in a single picture. However, pictures and words can often be used in similar ways and can lead to similar kinds of behavior. In a psychological experiment or in everyday life, for example, seeing either the written word "cat" or a picture of a cat can lead an individual to say /kat/ or to be reminded that cats meow, hate water, chase mice, and so on.

A long-standing question in psychology asks how that subset of knowledge that can be conveyed by either a word or a picture is stored and processed. Two general classes of models have been put forward. Multiple semantic system models hypothesize that pictures and words are processed in distinct, specialized semantic systems (e.g., Paivio, 1971, 1986, 1991; Shallice, 1988). In Paivio's "dual-code" model, for example, there is a "logogen" system for word processing and an "imagen" system for picture processing. These systems can communicate with one another but operate independently and have their own organization and pro-cessing parameters. In contrast, common semantics system, or "single-code," models hypothesize that words and pictures converge on a single, common semantic store (e.g., Caramazza, Hillis, Rapp, & Romani, 1990; W. R. Glaser, 1992; Potter & Faulconer, 1975; Pylyshyn, 1980; Riddoch, Humphreys, Coltheart, & Funnell, 1988; Snodgrass & McCullough, 1986). Different models allow different degrees of divergence between the processing of pictures and words before their convergence on the common store. However, these models all posit that semantic analysis takes place in a single, amodal system, and that information in that system is stored in a shared format that can no longer be traced back to the modality of input.

These two types of accounts make very different predictions about the pattern of results that should be seen when comparing the processing of words and pictures and when looking for interactions between them. Multiple-code accounts predict that information in different modalities will generally be processed independently, and that interactions across modalities, when they do occur, will be delayed and weak relative to within-modality interactions. This contrasts with the predictions of single-code accounts that there should be no differences in response time or size of facilitation or interference effects for within- and across-modality comparisons (e.g., Snodgrass, 1984). Both behavioral and electrophysiological studies aimed at testing these alternative accounts have uncovered a complex pattern of similarities and differences between the processing of words and pictures that often do not cleanly support either of the types of models.

Kara D. Federmeier, Department of Cognitive Science, University of California, San Diego; Marta Kutas, Departments of Cognitive Science and Neurosciences, University of California, San Diego.

Correspondence concerning this article should be addressed to Kara D. Federmeier, Department of Cognitive Science, University of California, San Diego, 9500 Gilman Drive, La Jolla, California 92093-0515. Electronic mail may be sent to kfederme@cogsci.ucsd.edu.

Behavioral Findings

The finding of cross-modal facilitation and interference with a time course similar to that seen for within-modality interactions has been taken as support for single-code models. For example, Potter, Kroll, Yachzel, Carpenter, and Sherman (1986) found that sentence-final pictures could be semantically integrated into a (verbal) sentence context as rapidly as could sentence-final words when the task was to make plausibility judgments. In addition,

many studies showed that semantic priming can be observed between pictures and words, at least when the task requires or encourages semantic analysis (Bajo, 1988; Sperber, McCauley, Ragain, & Weil, 1979; Theios & Amrhein, 1989; Vanderwart, 1984). Mixed-modality stimuli containing both pictures and words also yield Stroop-like interference effects (Dunbar & MacLeod, 1984; M. O. Glaser & Glaser, 1982; W. R. Glaser & Dungelhoff, 1984; W. R. Glaser & Glaser, 1989; La Heij, Dirkx, & Kramer, 1990). To the extent that these effects can be interpreted as "fast" or "automatic," they also tend to support common-code models (see review by W. R. Glaser, 1992).

However, single-code models generally predict equivalent response times for pictures and words in semantic tasks and symmetrical cross-modal interactions, and this is rarely the case. In fact, there are consistent, task-related differences associated with picture and word processing and with cross-modal interactions. Cattell (1886) was the first to note that reading a list of words aloud was faster than naming a corresponding set of pictures. This "picture-word naming difference" has been observed in numerous subsequent studies (Bajo, 1988; W. R. Glaser & Dungelhoff, 1984; W. R. Glaser & Glaser, 1989; Irwin & Lupker, 1983; Potter & Faulconer, 1975; Potter, So, von Eckardt, & Feldman, 1984; Smith & Magee, 1980; Theios & Amrhein, 1989). The naming difference holds across age and reading experience (Ligon, 1932; Lund, 1927), practice (Brown, 1915; Fraisse, 1969), and language type (e.g., Biederman & Tsao, 1979, studies in Chinese; Potter et al., 1984). However, although slower to name, pictures are categorized more rapidly than words, whether the task involves giving a superordinate label (e.g., W. R. Glaser & Dungelhoff, 1984; Irwin & Lupker, 1983; Smith & Magee, 1980) or deciding whether two items belong to the same semantic category (e.g., Rosch, 1975). Time to categorize mixed word-picture pairs seems to be intermediate (Pellegrino, Rosinski, Chiesi, & Siegel, 1977).

Related differences are seen in priming and interference tasks. When picture-word Stroop effects are examined in the context of a naming task, words interfere with picture naming but pictures do not interfere with word naming (i.e., reading; e.g., W. R. Glaser & Glaser, 1989), whereas the opposite pattern is obtained with categorization (W. R. Glaser & Dungelhoff, 1984; Smith & Magee, 1980). Another asymmetry emerges in priming tasks, wherein pictures are consistently both more effective primes and more susceptible to priming than are words (e.g., Bajo, 1988; Carr. McCauley, Sperber, & Parmelee, 1982; McCauley, Parmelee, Sperber, & Carr, 1980; Sperber et al., 1979). Furthermore, it has often been found that cross-modal facilitation is reduced relative to within-modality priming effects (however, see Bajo, 1988, for a case of equivalent within- and cross-modality priming effects). This difference can be explained within common-code models only by assuming that some portion of the within-modality priming is due to nonsemantic effects (e.g., visual similarity, lexical similarity).

In general, therefore, behavioral data consistently show that pictures and words interact (e.g., prime one another, interfere with one another), and that these interactions can be fast, automatic, and similar in time course to within-modality interactions. On the whole, then, these data argue against strict forms of multiple semantic systems accounts. In contrast, counter to the claims of the strongest forms of common semantic system accounts, these cross-modal interactions are generally weaker than within-modality interactions and show task-dependent asymmetries. In fact, it is this sensitivity to task that, in part, makes the pattern of behavioral data difficult to interpret, because it is not at all clear that the various tasks used (i.e., naming, categorization) are truly equivalent for words and pictures. Words may have an advantage in the naming task, for example, because of their greater lexical specificity and not because of any differences in semantic processing per se. Indeed, under those rare circumstances when the number of responses typically elicited by a picture is low (no more than two), the picture-word naming difference does seem to be reduced or eliminated (La Heij & Vermeij, 1987). Similarly, Snodgrass and McCullough (1986) argued that the picture advantage for categorization may be due to people's ability to match pictures on the basis of visual similarity alone. They found that, whereas classification of items into visually dissimilar categories was faster for pictures, classification into visually similar categories (e.g., fruits vs. vegetables) was actually slower for pictures than for words (although it is important to note that visual similarity cannot explain why word-picture pairs are categorized more rapidly than word-word pairs, because in neither case is there task-relevant visual similarity).

Electrophysiological Findings

Reaction time and error measures necessarily sample the entire processing stream, from stimulus encoding through the motor response. When functional asymmetries are observed between picture and word processing, therefore, it becomes difficult to know whether these reflect modality-specific differences in semantic processing or whether they arise from some other part of the processing stream. To circumvent this difficulty, several studies also examined the issue of single versus multiple semantic systems using event-related brain potentials (ERPs). ERPs have several advantages as a dependent measure, including millisecond-level temporal resolution, the ability to distinguish quantitative differences from qualitative differences, and a link to neurobiology. ERPs are especially amenable to examining modality effects on semantic processing because a component of the ERP-the N400, a negativity observed between 250 to 500 milliseconds after stimulus onset-has been specifically linked to semantic processing. The N400 seems to be the normal response to words, whether they are printed, spoken, or signed (e.g., Kutas, Neville, & Holcomb, 1987), and its amplitude varies inversely with an item's fit to the current semantic context, whether that context is a sentence or a single preceding word (e.g., Bentin, McCarthy, & Wood, 1985; Boddy & Weinberg, 1981; Kutas, 1993).

Kutas and Van Petten (1990) first reported that incongruous pictures in sentence contexts elicited N400-like activity-increased negativity 250 to 500 ms after stimulus onset-relative to congruous pictures. Barrett and Rugg (1990) also showed this in a semantic priming paradigm in which study participants made relatedness judgments. They found that the second of a sequentially presented pair of line drawings elicited a larger negativity at about 450 ms when unrelated than when related. This negativity thus behaved similarly to the N400 observed in related studies using words, although it was slightly later and had a different scalp distribution. Holcomb and McPherson (1994) likewise observed a negativity whose amplitude varied with the semantic relatedness of line drawings in an object decision task. Again, however, the

distribution of this negativity differed from the central-posterior maximum typically observed for words: It was significantly larger over frontal sites and almost absent over occipital sites.

Similar to work by Potter et al. (1986), pictures and words were compared more directly in two experiments by looking at sentence-final processing of words and pictures under the same task conditions. Nigam, Hoffman, and Simons (1992) recorded ERPs as individuals read sentences for comprehension. Half of the sentences ended with a word (matching or mismatching the context), whereas in the other half the final word was replaced by a black and white line drawing of the same object (again, matching or mismatching the context). Because both pictures and words elicited an N400 response identical in amplitude, latency, and distribution over the scalp, Nigam et al. took their results as strong support for a single conceptual system accessed similarly by both pictures and words. Ganis, Kutas, and Sereno (1996) also observed no differences in the amplitude or latency of N400 responses to sentence-final words and pictures. They did, however, find a significant distributional difference: The N400 response to pictures was larger over frontal sites and smaller over posterior sites than that to words. They concluded that the meaning of pictures and words is processed in functionally similar neural systems that are nonetheless at least partially nonoverlapping.

ERP studies seem generally to support common semantic systems models, because the semantic processing of both pictures and words is associated with a negativity whose amplitude varies with fit to context, with similar effect size and latency when tasks are held constant. That this holds true even when contexts are cross-modal (i.e., the impact of a verbal sentence context on the processing of a picture is similar to the impact of that context on the processing of a word) clearly goes against the predictions of strong forms of multiple semantic systems accounts. However, it remains unclear how to interpret the observed difference in the distribution of this negativity when elicited by pictures as opposed to words. On the one hand, the difference may imply a nonidentity of the neural generators responsible for this component in the two cases. Alternatively, McPherson and Holcomb (1999) suggested that the distributional difference may be due to the partial overlap of a truly modality-independent N400 with an earlier, frontally distributed negativity (N300) that is particular to the processing of pictures (see also Holcomb & McPherson, 1994). Either account, however, seems incompatible with the strong form of common-code models, which assert that beyond early perceptual processing the analysis of pictures and words should be identical.

As was also true for the behavioral data, then, the electrophysiological data seem inconsistent with the strongest form of either type of model. Processing within and across the two modalities is similar in many respects but is nevertheless not identical. Choosing between weaker forms of the two models, however, is not possible on the basis of the current data set alone, in part because neither model has ever been specified at a sufficient level of neurobiological detail to allow clear predictions at the level of ERP topography. For example, the topographical difference that has been observed could be taken to mean that a similar process is accessing semantic information that is stored independently in two different brain areas, as might be predicted by a version of the multiple semantic systems account. Alternatively, the difference could be interpreted as indicating that information is retrieved from a common, but neurally distributed, semantic system, and that the con-

cepts that have been tested as pictures have representations with a somewhat different distribution than the concepts that have been tested with words (none of the ERP studies finding a topographical difference used exactly the same items in both modalities). In the absence of a more detailed specification of which brain areas are presumed to be involved in semantic representation/processing and how those areas interact, it remains difficult to choose between these types of alternative explanations given just a pattern of topographical differences.

In contrast, even weaker versions of the two types of models make different predictions at a functional level (i.e., about how processing in different modalities would be affected by manipulations of input factors). In a common semantic system account, for instance, changing various stimulus parameters that influence semantic processing (e.g., the nature or strength of the semantic context) should always have a qualitatively similar effect on the semantic analysis of words and pictures and thus on the resulting pattern of N400 effects (topographical differences aside). In contrast, multiple semantic systems accounts clearly predict that one should be able to find some modality-dependent differences in the pattern of response to manipulations of semantic processing. It seems, then, that the next step in trying to build an understanding of how semantic information is represented and processed as a function of modality would be to gain a more detailed picture of how the brain responds to pictures and words, as a function of various semantic and contextual factors. This is the goal of the present study.

Experiment 1

The ERP studies mentioned previously have shown that the N400 (or N400-like) response to both pictures and words reveals something about an item's basic fit to a (single item or sentence) context (i.e., whether the item is unrelated or related). However, although we have a fairly detailed understanding of how the N400 response to words is affected by various semantic manipulations (see, e.g., Kutas & Van Petten, 1994, for review), we still know very little about how these factors influence the response to pictures. For instance, in previous work using words, we showed that N400-amplitude modulations reflect not only fit to context but also the context-independent organization of semantic knowledge in long-term memory (Federmeier & Kutas, 1999). Specifically, we observed that the N400 response to an unexpected sentence-final word was reduced in amplitude if that word shared a categorical relationship with the word most expected in the context. For example, given the sentence pair (where "palms" is the most expected ending):

> They wanted to make the hotel look more like a tropical resort. So along the driveway they planted rows of ...

we observed a smaller N400 response to the unexpected but categorically related ending "pines" than to the also unexpected but categorically unrelated ending "tulips." This difference between the two contextually unexpected endings could not be explained by lexical association, contextual plausibility, or the two in combination. Instead, it seems to vary with the degree of semantic feature overlap between the contextually expected item and the word that is actually presented.

In addition, we found that this influence of semantic memory organization on the N400 response was modulated by contextual strength. We compared the brain's response to the three types of endings (expected, unexpected but categorically related, and unexpected and not categorically related) in highly constraining contexts -- those that lead to a strong prediction or preference for the expected ending -- and in less constraining ones (with constraint defined by the cloze probability of the expected completion). An example of each, with the corresponding expected, unexpected but categorically related, and unexpected and not categorically related ending, is given next.

> Highly constraining: "He journeyed to the African plains, hoping to get a photograph of the king of the beasts. Unfortunately, the whole time he was there he never saw a lion/tiger/panda."

> Less constraining: "By the end of the day, the hiker's feet were extremely cold and wet. It was the last time he would ever buy a cheap pair of boots/sandals/jeans."

Off-line rating data indicate that highly constraining contexts increase the plausibility of expected items and, correspondingly, decrease the plausibility of unexpected items. In the ERP data, constraint did not affect the response to either the expected items or the unexpected items that were not categorically related. However, constraint did affect the response to unexpected but categorically related items; these were associated with significantly smaller N400 amplitudes in highly constraining as opposed to weakly constraining contexts. That is, the N400 response to items like "tiger" (see prior example) was smaller (i.e., more similar to that for expected items) than the response to items like "sandals." Strikingly, then, in this case, N400 amplitudes went opposite to the items' rated plausibility in their sentence contexts. This suggests that semantic memory organization does not become relevant only when other cues are less available, but rather that its influence is an inherent consequence of the way that the brain processes linguistic input.

This kind of paradigm, in which context and semantic memory structure are manipulated in tandem, can be used to examine picture processing in a more fine-grained manner than has been done before. In turn, we then are able to compare the effects of context, contextual constraint, and semantic memory use on picture and word processing with more specificity and thereby test the predictions of the two classes of models. In this study, therefore, we used the same materials and procedure as in Federmeier and Kutas (1999), replacing the sentence-final word with a line drawing of the same concept. If, as assumed by common mode models, pictures and words access the same semantic knowledge base, then we would expect to see the same impact of its organization on the N400 response to each (even if the scalp distribution of the N400 responses themselves are modality dependent). Finding that categorical relatedness of the type tested here does not affect the ERP response to pictures as for words, in contrast, would clearly support some version of a multiple semantic systems account, because it would indicate that pictures access a conceptual store with a different organization than do words. Similarly, because common-code accounts assert that semantic information is amodal, factors like contextual constraint should have the same impact on picture as on word processing. Modality-related differences in the influence of contextual constraint, therefore, would imply that the semantic information accessed from pictures and words is differ-

ent, showing that semantic processing is not modality independent. Here, then, we examine the pattern of N400 responses to pictures as a function of fit to context, category membership, and contextual constraint. We also analyze early attention- and visual perception-related ERP components for modality-related differences as well as to see whether such differences, if observed, correlate with any later, semantic effects.

Method

Materials

Stimuli were derived from those used in Federmeier and Kutas (1999). They consisted of 132 pairs of sentences, ending with three types of targets: (a) *expected exemplars* (items with the highest cloze probability in the sentence contexts), (b) *within-category violations* (contextually unexpected items derived from the same taxonomic category as the expected exemplar), and (c) *between-category violations* (contextually unexpected items derived from a different category than the expected exemplar). Sentence-final words in the Federmeier and Kutas study were replaced in this experiment with black and white line drawings derived from the Snodgrass and Vanderwart (1980) set and from a commercial clip art package. Each line drawing was normed to ensure naming agreement with the word that it would replace. Nineteen individuals were asked to name each picture; pictures were used in the study only if the replaced word was the most commonly named response for that picture. On average, there was 88% agreement between the word used in the Federmeier and Kutas stimuli and the naming response to the corresponding line drawing used in this experiment.

The first sentence of each pair established the expectation for the target item and its category. In contrast, the second sentence, if presented alone, could be completed plausibly by any of the three possible targets. Targets were objects from 66 categories (two from each). Categories were chosen to be those at the lowest level of inclusion for which the average undergraduate student could be expected to differentiate several exemplars readily. For approximately half the categories used, this level was basic as determined by Rosch, Mervis, Gray, Johnson, and Boyes-Braem (1976) or by analogy. Other categories were based at the next highest level (a superordinate of the basic level) because it was unclear that the average participant could clearly and consistently differentiate below this level. To help control for the plausibility of the two violation types, between-category targets for each sentence pair were chosen from a related category that shared some key features (e.g., animacy, size, general function) with that from which the expected exemplar and within-category violation were derived.

The experimental sentences were divided into three lists of 132 sentences each; each participant viewed one list. Sentence contexts and line drawings were used only once per list; each list consisted of 44 of each type of target (expected exemplars, within-category violations, between-category violations) and the same set of 44 plausible filler sentence pairs. The ending-type condition of target items was rotated such that, across lists, each item appeared once as each type of ending. Thus, across the experiment, all conditions were perfectly controlled for both lexical and visual properties of the target, and context sentences in each ending-type condition also were perfectly controlled for length and grammatical complexity.

Cloze Procedure and Constraint

Cloze probabilities were obtained for the 132 sentence pair contexts (sentence pairs missing the final item of the second sentence). These were divided into two lists, such that the two sentence contexts presumed to be predictive of items coming from the same category did not both appear on the same list. Student volunteers were asked to complete each sentence pair

with "the first word that comes to mind." List 1 was completed by 56 students, and List 2 was completed by 59 different students. A subset of the original stimuli were rewritten and clozed separately by a third group of 55 students. Cloze probability for a given word in a given context was calculated as the proportion of individuals choosing to complete that particular context with that particular word. Expected exemplars were always the item with the highest cloze probability for a given context ($M = 0.74$). All violations had cloze probabilities of less than .05, yielding a mean cloze probability of .004 for the within-category violations and .001 for the between-category violations. To ensure that within- and between-category violations (which, with rare exceptions, were not generated in the cloze task) were, in fact, considered implausible completions for the sentence contexts, plausibility ratings of all items in their sentence contexts were also obtained from a different group of student volunteers. These ratings confirmed that, in addition to not being produced in the sentence contexts, both violation types were regarded as surprising/difficult to integrate when placed into them; see Federmeier and Kutas (1999) for details and analyses.

Although all expected exemplars were items with the highest cloze probability for their sentence contexts, the actual cloze probability of these items ranged from 0.17 to 1.0. In other words, the sentence contexts differed in their constraint, or the degree to which they led individuals to expect one particular item strongly versus a number of different items. To examine the effects of sentential constraint on the ERP response to target items, we divided the sentences into two groups-"high constraint" and "low constraint"-by a median split on the cloze probability of the expected exemplar. For the high-constraint sentences, the cloze probability of the expected exemplars had a range of .784 to 1.0 and an average value of .896 (median = .904). For the low-constraint sentences, the cloze probability of the expected exemplars had a range of .17 to .784 and an average value of .588 (median = .608). High-constraint sentences are thus those in which there is a single, highly preferred ending, whereas low-constraint sentences are those that are compatible with a larger range of ending types and in which the expected exemplar has at least one, and generally several, close competitors.

Participants

Eighteen University of California, San Diego (UCSD) undergraduate volunteers (9 women and 9 men; age range = 18-28 years; mean age = 21 years) participated in the experiment (none of these volunteers took part in any of the norming procedures) for course credit or for cash. All were right-handed (as assessed by the Edinburgh Inventory; Oldfield, 1971) monolingual English speakers with normal vision and no history of reading difficulties or neurological/psychiatric disorders. Six participants were randomly assigned to each of the three stimulus lists.

Experimental Procedure

Volunteers were tested in a single experimental session conducted in a soundproof, electrically shielded chamber. They were seated in a comfortable chair 45 inches in front of a monitor and instructed to read the stimulus sentences for comprehension. They also were informed that they would be given a recognition memory test over the stimuli at the conclusion of recording. The session began with a short practice run.

Each trial began with the first sentence of a pair appearing in full on a cathode-ray tube screen. Volunteers read this sentence at their own pace and pushed a button to view the second sentence. Presentation of the second sentence was preceded by a series of crosses to orient the volunteer toward the center of the screen. The second sentence was then presented one word at a time for a duration of 200 ms with a stimulus onset asynchrony of 500 ms. Sentence-final pictures subtended 7.5 degrees of horizontal visual angle and 6.5 degrees of vertical visual angle and were presented for a duration of 500 ms. Volunteers were asked not to blink or

move their eyes during the second sentence. The final, target picture was followed by a blank screen for 3000 ms, after which the next sentence appeared automatically. Volunteers were given a short break after every 17 pairs of sentences.

At the conclusion of the recording session, participants were given a recognition memory test consisting of 50 sets of sentence pairs-10 new, 20 unchanged experimental pairs (of which 10 ended with expected exemplars, 5 ended with within-category violations, and 5 ended with between-category violations)-and 20 modified sentence pairs in which the final item had been changed from that originally viewed by the volunteer (10 in which violations had been changed to expected exemplars and 10 in which expected exemplars had been changed to violations). Pictures were not used in the recognition memory test; instead, each sentence-final item was the word corresponding to the picture's label. Volunteers were instructed to classify the sentences as new, old, or similar (changed).

Electroencephalographic Recording Parameters

The electroencephalogram (EEG) was recorded from 26 tin electrodes arranged geodesically in an Electro-cap, referenced to the left mastoid. These sites included midline prefrontal (MiPf), left and right medial prefrontal (LMPf and RMPf), left and right lateral prefrontal (LLPf and RLPf), left and right medial frontal (LMFr and RMFr), left and right mediolateral frontal (LDFr and RDFr), left and right lateral frontal (LLFr and RLFr), midline central (MiCe), left and right medial central (LMCe and RMCe), left and right mediolateral central (LDCe and RDCe), midline parietal (MiPa), left and right mediolateral parietal (LDPa and RDPa), left and right lateral temporal (LLTe and RLTe), midline occipital (MiOc), left and right medial occipital (LMOc and RMOc), and left and right lateral occipital (LLOc and RLOc). Blinks and eye movements were monitored with electrodes placed on the outer canthus (left electrode serving as reference) and infraorbital ridge of each eye (referenced to the left mastoid). Electrode impedances were kept below 5 k-ohms. EEG was processed through Grass amplifiers set at a band-pass of 0.01 to 100 Hz. EEG was continuously digitized at 250 Hz and stored on hard disk for later analysis.

Data Analysis

Data was re-referenced off line to the algebraic sum of the left and right mastoids. Trials contaminated by eye movements, blinks, excessive muscle activity, or amplifier blocking were rejected off-line before averaging; on average 11 % of trials were lost because of such artifacts. ERPs were computed for epochs extending from 100 ms before stimulus onset to 920 ms after stimulus onset. Averages of artifact-free ERP trials were calculated for each type of target picture (expected exemplars, within-category violations, between-category violations) after subtraction of the 100-ms prestimulus baseline.

Results

Behavior

Volunteers correctly classified an average of 93% (range = 82%-100%) of the items on the recognition memory test, indicating that they were attending to the experimental sentences during the recording session. The two most common types of errors (67%) were a misclassification of "similar" sentences (those in which the final word did not match the picture seen in that sentence context) as "old" or a misclassification of "old" sentences (those in which the final word did match the picture seen in that context) as "similar." The remainder of the errors primarily consisted of volunteers classifying "old" or "similar" sentences as "new." Only

two errors in which "new" sentences were classified as "old" or "similar" were observed across the 18 participants.

ERPs

Grand average ERPs ($N = 18$) to sentence-final pictures from all recording sites in high- and low-constraint sentences are shown in Figure 1. Early components in all conditions include, at posterior sites, a positivity peaking at about 80 ms (P1), a negativity peaking at about 150 ms (N 1), a positivity peaking at about 200 ms (P2), and, at frontal sites, a negativity peaking at about 100 ms (NI) and a positivity peaking at about 175 ms (P2). Early components are followed by broadly distributed negativities peaking at about 300 ms (early N400/N300) and 425 ms (late N400). At all sites and in both constraint conditions, these negativities were smallest in response to the expected exemplars, largest to the between-category violations, and intermediate in amplitude to the within-category violations. These negativities are followed by an extended late positivity largest over the central and posterior sites and similar in size for all conditions.

Overall mean amplitude analyses. Mean voltage measures were taken in five time windows encompassing major components of the ERP: 50-150 ms (frontal Nl), 150-250 ms (P2), 250-350 ms (early N400/N300), 350-500 ms (late N400), and 500-700 ms (late positivity). These measures were subjected to an omnibus analysis of variance (ANOVA). Repeated measures included two levels of constraint (high vs. low), three levels of ending type (expected exemplar vs. within-category violation vs. between-category violation), and 26 levels of electrode. All p values in this and all subsequent analyses are reported after epsilon correction (Huynh-Felt) for repeated measures with greater than 1 degree of freedom. Table 1 shows the ANOVA results.

Constraint affected the ERP response in the time windows encompassing the N1 (main effect modulated by a Constraint X Ending Type interaction) and the P2 (trend toward a main effect modulated by a significant Constraint X Ending Type interaction), with an additional trend toward a Constraint X Ending Type interaction in the early N400/N300 time window (along with a significant Constraint X Electrode interaction). Ending type affected the ERP response in the time windows encompassing the P2, the early N400/N300, and the late N400; in all cases, there was also a significant interaction with electrode, suggesting differences in the distribution of the response elicited by the three ending
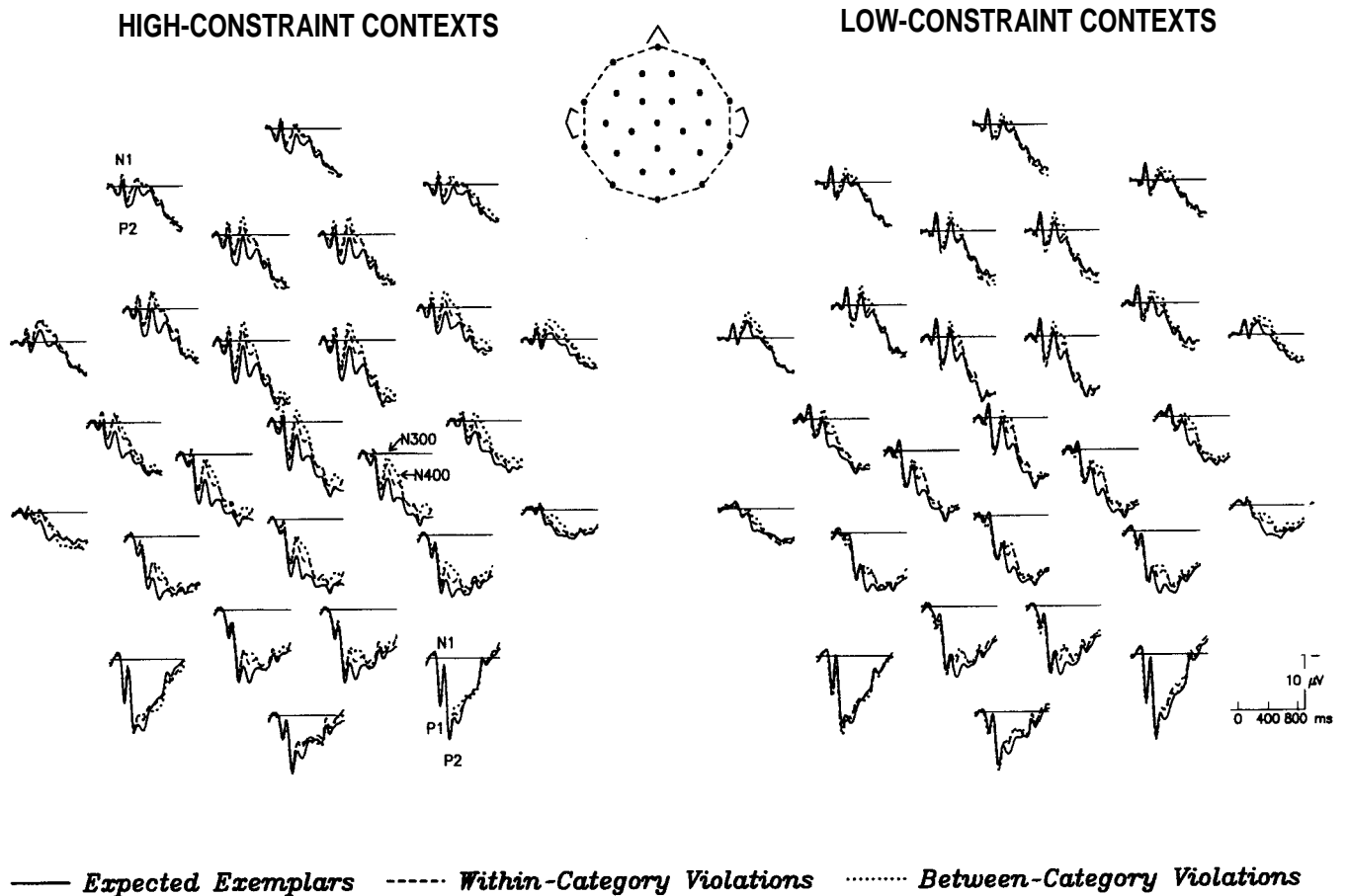
HIGH-CONSTRAINT CONTEXTS   LOW-CONSTRAINT CONTEXTS

——— *Expected Exemplars*   ------ *Within-Category Violations*   ········· *Between-Category Violations*

Figure 1. Event-related potential (ERP) responses to expected exemplars (solid line), within-category violations (dashed line), and between-category violations (dotted line) in high-constraint (left) and low-constraint (right) contexts. Negative is plotted up. The 26 electrode site locations are laid out to approximate the arrangement over the head as seen from the top (illustrated by head icon). Major ERP components are labeled on the left side.

Table 1

Overall Results for Experiment 1

Time window (ms)	Constraint $F(1, 17)$	Constraint X Electrode $F(25, 425)$	Ending type $F(2, 34)$	Ending Type X Electrode $F(50, 850)$	Constraint X Ending Type $F(2, 34)$	Constraint X Ending Type X Electrode $F(50, 850)$
50-150	**9.03****	1.83	0.08	1.59	**6.66****	1.68
150-250	2.90, $p = .11$	0.68	**5.03***	**3.28****	**4.34***	1.18
250-350	0.04	**2.46***	**15.21****	**3.08****	2.31, $p = .11$	1.34
350-500	1.33	0.54	**18.06****	**3.37****	0.15	1.55
500-700	0.02	0.54	1.79	1.72	0.24	0.83

Note: Significant effects are identified in boldface.
* $p < .05$. ** $p < .01$.

types. No significant effects were observed on the late positivity, so that time window was dropped from follow-up analyses. To characterize the observed effects better, follow-up analyses were done for each constraint condition as a function of ending type and then for each ending type as a function of constraint.

Mean amplitude analyses for high-constraint sentences. Figure 2, Panel A shows the ERPs to the three ending types in high-constraint sentences at eight representative electrode sites. In all time windows showing effects of ending type, the response to expected exemplars was most positive and that to between-category violations most negative; the response to within-category violations was generally intermediate in amplitude. To test these differences, mean voltage measures were taken in the same four time windows in which significant effects of experimental condition (constraint or ending type) had been observed in the overall analysis: 50-150 ms (frontal N1), 150-250 ms (P2), 250-350 ms (early N400/N300), and 350-500 ms (late N400). Expected exemplars were compared with within-category violations and within-category violations were compared with between-category violations in two separate ANOVAs. Repeated measures for each included two levels of ending type and 26 levels of electrode. Table 2 shows the results.

Expected exemplars significantly differed from within-category violations in all four time windows. This effect interacted with electrode in the early N400/N300 time window, with a trend toward an interaction in both the N1 time window and the late N400 time window. At all four time windows, therefore, expected exemplars were significantly more positive than violations of either type (because the response to between-category violations was always at least as large, if not larger, than the response to within-category violations).

To follow-up on the observed Ending Type X Electrode interactions, distributional analyses were conducted. Mean amplitude measures were normalized according to the procedure described in McCarthy and Wood (1985) and then subjected to an ANOVA on four repeated measures: two levels of ending type, two levels of hemisphere (left vs. right), two levels of laterality (lateral vs. medial), and four levels of anterior/posterior (prefrontal vs. frontal vs. parietal vs. occipital). This and subsequent distributional analyses thus used 16 electrode sites, divided into left lateral sites (from front to back: LLPf, LLFr, LLTe, LLOc), left medial sites (LMPf, LMFr, LMCe, LMOc), right medial sites (RMPf, RMCe, RMOc), and right lateral sites (RLPf, RLFr, RLTc, RLOc).

After normalization, no statistically significant distributional effects were observed for the comparison between expected exemplars and within-category violations in any time window, indicating that the ERP response to these ending types was very similar over the scalp once amplitude differences were accounted for.

Within-category violations and between-category violations differ significantly in a relatively narrow time window from 300 to 400 ms, $F(1, 17) = 5.39, p < .05$, in which time there is also a marginal Ending Type X Electrode interaction, $F(25, 425) = 2.13$, p = .09. In this time window, between-category violations are more negative than within-category violations. Distributional analysis (as discussed previously) suggest that the N400 response to between-category violations is more frontally distributed than that to within-category violations (Ending Type X Anteriority interaction, $F(3, 51) = 3.97, p < .05$).

In summary, in high-constraint sentences, the effect of context (difference between expected items and any violations) begins earlier and lasts longer than the effect of category (difference between within- and between-category violations as a function of semantic similarity to the expected item). Context begins to have its effects very early in the ERP, on components such as the N1, and this influence continues into the early and late N400 time windows. Category, in contrast, begins to influence the ERP only later and in a narrower time window encompassing the latter part of the early N400/N300 and the beginning part of the late N400.

Mean amplitude analyses for low-constraint sentences. Figure 2, Panel B shows the ERPs to the three ending types in low-constraint sentences. Similar to effects seen in high-constraint sentences, in time windows showing an ending type effect, the response to expected exemplars was most positive and that to between-category violations was most negative; responses to within-category violations were generally of intermediate amplitude. To test these differences, mean voltage measures were analyzed in the same four time windows and for the same comparisons as for the high-constraint condition (Table 3).

Expected exemplars significantly differed from within-category violations only in the late N400 time window (with a trend observed in the N1 time window, reflecting a tendency for larger amplitudes to expected exemplars than within-category violations), and this effect interacted with electrode. For low-constraint sentences, then, expected exemplars are significantly more positive than violations starting at about 350 ms; the negativity to within-category violations was more right-lateralized over poste-
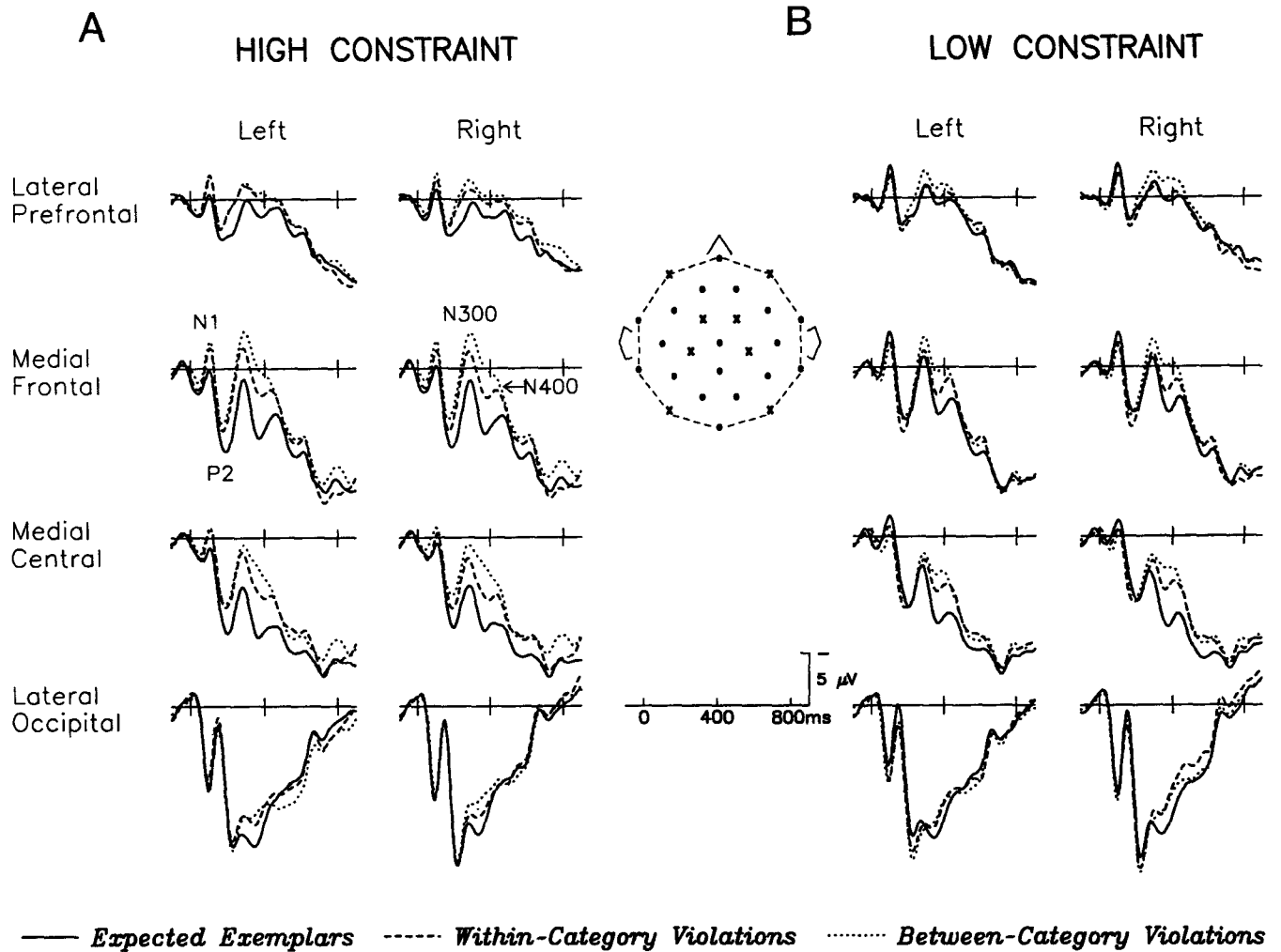
A HIGH CONSTRAINT   **B** LOW CONSTRAINT

——— *Expected Exemplars*   ------ *Within-Category Violations*   ········· *Between-Category Violations*

Figure 2. Event-related potential responses to the three ending types in high-constraint (A) and low-constraint (B) contexts at eight representative electrode sites (indicated with **X**s on the head icon). Negative is plotted up. For both sentence types, expected exemplars (solid line) elicited the most positive responses in the 250- to 500-ms window, and between-category violations (dotted line) elicited greater negativity than within-category violations (dashed line).

rior sites and more prominent over anterolateral electrode sites than the response to expected exemplars (Ending Type X Hemisphere X Laterality X Anteriority interaction, $F(3, 51) = 3.30$, $p < .05$).

Within-category violations and between-category violations differ significantly in a narrow time window (starting slightly later than the difference observed for high-constraint sentences) from 350 to 400 ms, $F(1, 17) = 4.34$, $p < .05$, with no significant

Table 2
Results for High-Constraint Comparisons

Time window (ms)	Expected vs. within $F(1, 17)$	Expected/Within X Electrode $F(25, 425)$	Within vs. between $F(1, 17)$	Within/Between X Electrode $F(25, 425)$
50-150	**5.99***	2.06, $p = .08$	0.31	1.41
150-250	**14.72***	1.96	0.57	1.86
250-350	**15.48***	**2.63***	3.12, $p = .09$ [a]	1.56
350-500	**14.84***	2.19, $p = .06$	2.09[a]	1.82

Note. Significant effects are identified in boldface.
[a] Marginal or nonsignificant results that are found to be significant for analyses in part of the time window or at a subset of all electrode sites.
* $p < .05$. ** $p < .01$.

Table 3
Results for Low-Constraint Comparisons

Time window (ms)	Expected/within $F(1, 17)$	Expected/Within X Electrode $F(25, 425)$	Within/between $F(1, 17)$	Within/Between X Electrode $F(25, 425)$
50-150	3.25, $p = .09$	0.78	0.18	0.57
150-250	1.23	0.86	1.71	1.07
250-350	0.80	1.27	2.80, $p = .11$	1.76
350-500	**7.57** *	**3.44** *	1.73°	1.36

Note: Significant effects are identified in boldface.
[a] Marginal or nonsignificant results that are found to be significant for analyses in part of the time window or at a subset of all electrode sites.
* $p < .05$.

Ending Type X Electrode interaction, $F(25, 425) = 1.77$, p = us. In this time window, between-category violations are more negative than within-category violations over most electrode sites.

In summary, in low-constraint sentences, the effect of context (difference between expected items and violations) and the effect of category (difference between within- and between-category violations as a function of semantic similarity to the expected item) both begin in the N400 time window (350 ms); the effect of context lasted slightly longer. These differences occur later than those noted for either variable in high-constraint sentences.

Mean amplitude analyses of ending type across constraint. Figure 3 shows the ERPs to the three ending types as a function of constraint. Effects of constraint on mean voltages were analyzed for each of the ending types at the same four time windows (50-150 ms, 150-250 ms, 250-350 ms, 350-500 ms) in three ANOVAs. Repeated measures for each included two levels of constraint (high vs. low) and 26 levels of electrode. Table 4 shows the results.

The effect of constraint on the response to expected exemplars begins in the N1 time window, with a reduced N1 to expected exemplars in high- as opposed to low-constraint sentences. The response in low-constraint sentences is largest over medial frontal sites, whereas the response in high-constraint sentences is less frontally skewed and largest over right lateral sites (Ending Type X Hemisphere X Laterality X Anteriority interaction, $F(3, 51) = 5.08$, $p < .05$). Significant effects are also seen in the P2 and early N400/N300 time window (along with a marginal Constraint X Electrode interaction in the latter time window), and a Constraint X Electrode interaction is observed in the late N400 time window. In all cases, the response to expected exemplars is more positive in high-constraint than low-constraint sentences. Significant differences between expected items in high- and low-constraint sentences are also observed in the 350- to 450-ms time window if analyses are restricted to the 11 prefrontal and frontal sites, $F(1, 17) = 6.74$, $p < .05$.

No effects of constraint on the response to within-category violations are observed in any time window. When analyses are restricted to the eight most posterior sites (MiPa, LDPa, RDPa, LMOc, RMOc, MiOc, LLOc, and RLOc), the response to between-category violations is significantly more negative in high- than in low-constraint sentences between 250 and 300 ms, $F(1, 17) = 4.78$, $p < .05$.

In summary, the response to expected exemplars was more positive in high- than in low-constraint sentences in all time windows measured, whereas the response to between-category violations was more negative (over medioposterior electrodes) in high- than in low-constraint sentences from 250 to 300 ms. Constraint did not influence the response to within-category violations in any time window.

Discussion

At a general level, the results indicate that there is a functional similarity in how pictures and words are used to access semantic information from long-term memory and how they are integrated into a sentence context. Previous work indicated that the integration of word and picture information into a sentence context can unfold with a similar time course (Potter et al., 1986), and that fit to a sentence context is associated with a qualitatively similar ERP effect (reduction in negative amplitude 250-500 ms after stimulus onset) whether the target is a word or a picture (Ganis et al., 1996; Nigam et al., 1992). We replicated these results. As was observed for word processing using the same stimuli (Federmeier & Kutas, 1999), contextually unexpected pictures in both highly and more weakly constraining contexts are associated with enhanced negativity (N400) 250 to 500 ms after stimulus onset relative to contextually expected pictures. We also extend these findings by showing that the processing of both words and pictures is influenced by semantic similarity, here defined along the lines of taxonomic category structure. When we compare the response to unexpected pictures from the expected category with that to unexpected pictures from a different category, we find, as we did for words, less N400 activity in response to a contextual violation that shares more semantic features with the item most expected in a context than to one that shares significantly less features in common with the expected completion.

For pictures as for words, therefore, we find that both sentence contextual information and the context-independent structure of information in semantic memory affect the ERP between 300 and 500 ms in a similar manner. This does not rule out the possibility that semantic information for pictures and words is stored independently. However, it does suggest that semantic knowledge for objects, whether accessed by a word or a picture, has a similar structure in long-term memory, namely one based on the kind of perceptual and functional similarity underlying taxonomic categories. This is consistent with the claims of single-code models and reinforces our hypothesis that the organization of long-term mem-
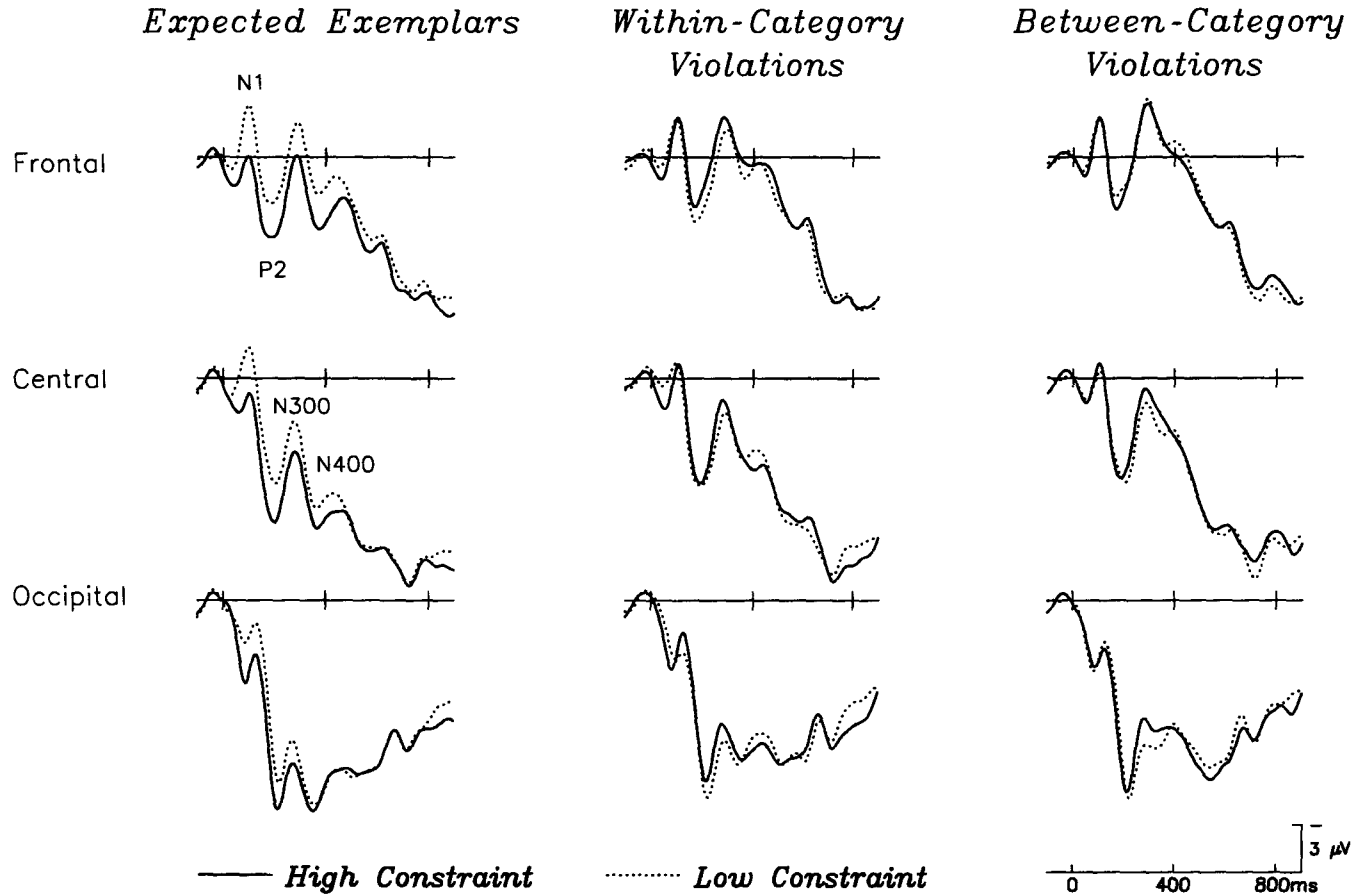
Figure 3.   Responses to expected exemplars, within-category violations, and between-category violations in high-constraint (solid line) and low-constraint (dotted line) contexts at three representative electrode sites. Negative is plotted up. Constraint influenced the response to expected exemplars in the 50- to 150-ms (frontal NI), 150- to 250-ms (P2), 250- to 350-ms (early N400/N300), and 350- to 500-ms (late N400) time windows. No effects *of* constraint were observed on the response to within-category violations. Effects of constraint on the response to between-category violations could be observed in the early N400/N300 time window at posterior electrode sites (bottom right), with more negative responses to these items in high- than in low-constraint contexts.

ory has an inherent impact on sentence processing, independent of modality.

Differences between word and picture processing emerge, however, when the data are examined at a more detailed level. First,

there seems to be a shift in the time course with which sentence context information and semantic category structure have their effects on processing. For words we find that, regardless of contextual constraint, effects of both context and category are limited

Table 4

Results for Each Ending Type

Time window (ms)	Expected (high vs. low constraint) $F(1, 17)$	Constraint X Electrode $F(25, 425)$	Within (high vs. low constraint) $F(1, 17)$	Constraint X Electrode $F(25, 425)$	Between (high vs. low constraint) $F(1, 17)$	Constraint X Electrode $F(25, 425)$
50-150	**31.88****	**3.51***	0.65	0.59	0.01	0.44
150-250	**10.90****	1.91	0.48	0.53	0.00	0.56
250-350	**4.45***	2.19, $p = .08$	0.38	0.65	1.11[a]	2.40, $p = .06$
350-500	1.67[a]	**2.36***	0.22	0.98	0.13	0.50

Note.   Significant effects are identified in boldface.
[a] Marginal or nonsignificant results that are found to be significant for analyses in part *of* the time window or at a subset *of* all electrode sites.
* $p < .05$.   ** $p < .01$.

to the N400 time window. This suggests that for word processing both variables primarily influence the case with which semantic information is activated and integrated with context. The same pattern is observed for pictures when these are in low-constraint contexts; the first and only effects of both context and category are seen in the N400 time window. We find a markedly different pattern, however, for pictures in highly constraining contexts. In this case, in addition to effects of context and category membership in the N400 time windows, we see effects of context (but not category) on much earlier components, namely the frontal N1 (50-150 ms) and the P2 (150-250 ms). The frontal N1 has been linked to allocation of visuospatial attention and has been hypothesized to reflect the output of a capacity-limited attentional system (e.g., Clark & Hillyard, 1996; Mangun, Hillyard, & Luck, 1993). Enhanced N 1 amplitudes are observed, for example, in response to target stimuli presented at attended relative to unattended locations. P2 amplitudes have been linked to processes of visual feature (color, orientation, size) detection; increased amplitudes have been observed in response to stimuli containing target features (e.g., Hillyard & Muente, 1984; Luck & Hillyard, 1994).

Effects on early components in semantic tasks have been reported before-Boddy and Weinberg (1981), for example, observed P2 enhancements associated with the detection of semantic features-though have sometimes proven difficult to replicate. We are the first to report such effects for the processing of pictures, likely because we are the first to examine the specific influences of constraint on picture processing. Here we see reliable reductions in the amplitude of the N1 and increases in the amplitude of the P2 for expected items in highly constraining contexts compared with less constraining contexts; these are apparent on a subject by subject basis (only 4 of the 18 participants do not seem to show this effect). We hypothesize that the effects are indicative of reduced attentional load and enhanced visual-semantic feature extraction for these items. Thus, for pictures (although not for words, at least those that are not visually degraded), strong contextual constraint not only eases the semantic processing of expected items but also seems to provide top-down information that can facilitate visual processing and attentional allocation.

If we examine the influence of constraint on the response to the three types of endings directly, additional differences between pictures and words are revealed. For words, the response to expected exemplars was unaffected by contextual constraint, suggesting that contextually compatible items were equally easy to process in very predictive and less predictive contexts. In contrast, for picture processing, the response to expected exemplars showed effects of constraint in several time windows. First, as previously noted, there are reduced N1 and enhanced P2 amplitudes to expected items in highly constraining contexts; this increased positivity in response to congruent items in highly constraining contexts continues into the early N400/N300 and late N400 time windows. In this case, then, increased ease of perceptual processing and semantic analysis go hand in hand, perhaps because the reduced visual processing load in highly constraining contexts frees attentional resources that would normally be required for perceptual processing to be shifted to conceptual integration. This finding is intriguing given that the same concepts were equally easy to integrate into high- and low-constraint sentences when they appeared as words. In short, for picture processing, we observe an apparent link between perceptual and semantic processing that we

did not observe for word processing with the same sentence contexts and the same concepts; this is explored in more detail later.

When the target items were words, constraint also had no effect on the response to between-category violations, affecting only the response to within-category violations. Increased contextual constraint, therefore, led to increased facilitation for unexpected but categorically related words and in a direction opposite from their rated plausibility in the sentence context. We thus had concluded that the language-processing system predicts the features of upcoming words (not the actual words themselves), such that unexpected words containing predicted features are facilitated in proportion to the consistency of the prediction allowed by the context (Federmeier & Kutas, 1999). For pictures, however, we find no influence of contextual constraint on the response to within-category violations in any time window. Instead, we find that constraint influences the N400 to between-category violations, with larger responses to these violations in high- than in low-constraint sentences. For both pictures and words, then, high contextual constraint magnified the N400 difference between unexpected items that are and are not semantically related to the expected sentence completion. However, for words this difference was driven by an increase in the case with which unexpected but semantically related items were processed, whereas for pictures the difference was driven by a decrease in the ease with which unexpected and semantically unrelated items were processed.

There are thus striking differences in how words and pictures representing the same semantic concept are integrated into identical sentence contexts. The question that remains is whether these differences are truly modality specific or are instead driven by a more general, mediating factor that correlates with modality. Because we have used the same sentence contexts, the same target concepts, and the same task to examine picture and word processing, most general differences have been controlled for. However, one potentially important, and frequently overlooked, factor to consider when comparing pictures with words is their differential perceptual predictability. Words are generally perceptually more predictable than pictures. If, as we have argued, contextual information allows the language-processing system to make predictions about the semantic features of items likely to appear next (Federmeier & Kutas, 1999), then in the case of words it also may facilitate predictions about the actual lexical item, including its physical appearance (e.g., font, size, color, spacing). The mapping between concept and physical form for pictures, however, is more variable, thereby making it difficult to predict in detail the actual physical characteristics of the picture that is likely to appear next. For example, even if a reader expects to see a line drawing of a cat, there are literally an infinite number of possible depictions of a cat, differing in viewpoint and the presence and form of various features (e.g., whether you will see a tail and, if so, its size, color, markings, shape). An open question, therefore, is whether such differences in perceptual predictability might underlie at least some of the ERP differences we observed between word and picture processing. If, for example, our hypothesis about the nature of the early perceptual/attentional effects we observe is correct, we would expect to see reductions in N1 and P2 amplitudes when the features of the pictures are easier to perceive and more predictable.

Experiment 2

To address the possibility that differences in the perceptual predictability of words and pictures could underlie some of the modality differences we have observed, we conducted a second experiment that was identical to the first in all but one respect: The perceptual predictability of the line drawings was more like that for words. This was accomplished by familiarizing participants with the full set of line drawings before the experimental session and informing them that only familiar line drawings would be used in the sentence contexts. This manipulation eases the perceptual load associated with parsing an unfamiliar picture and, more importantly, ensures that (as is also generally true for words) predictions about semantic features can also give rise to fairly accurate predictions about the nature of the upcoming physical stimulus. This manipulation thus allows us to determine which, if any, of the differences we observed between words and pictures are due to perceptual predictability and which reflect modality-specific differences that hold despite the familiarity or predictability of the physical stimulus itself. In all other respects this experiment was identical to Experiment 1.

Method

Materials and procedures for Experiment 2 were identical to those used in Experiment 1 except that participants viewed all the line drawings before the recording session. Eighteen right-handed (Oldfield, 1971) UCSD undergraduate volunteers (9 women; age range = *18-28* years of age; mean age = 21 years), none of whom had previously participated in Experiment 1, took part in this experiment for course credit or cash. Before the recording session, line drawings were presented on a computer monitor in the same size and format as used in the experimental trials; participants pushed a button to move through the set of pictures, which were presented in random order. Participants were told that these pictures would be used during the recording session and were instructed to take some time to look at each. To ensure that participants would spend time looking at each picture, they were instructed to write down what each picture represented and to rate the "quality" of the line drawing as a representation of that object on a 7-point scale, ranging from 1 *("l have difficulty determining what this line drawing is supposed to represent")* to 7 *("This line drawing is one of the best possible representations 1 can imagine for this item")*. The average rating across items was 5.7 (range = 2.7-6.6), indicating that participants found the items to be generally good representations of what they believed to be the target concept.

Results

Behavior

Volunteers correctly classified an average of 93% (range = 80%-98%) of the items on the recognition memory test, indicating that they were attending to the experimental sentences during the recording session. The two most common types of errors (78%) were a misclassification of "similar" sentences (those in which the final word did not match the picture seen in that sentence context) as "old" followed by a misclassification of "old" sentences (those in which the final word did match the picture seen in that context) as "similar." The remainder of the errors primarily consisted of volunteers classifying "old" or "similar" sentences as "new." Only one error in which a "new" sentence was classified as "similar" (and none in which a "new" sentence was classified as "old") was observed across the 18 participants.

ERPs

Grand average ERPs (N = 18) to sentence final preexposed pictures from all recording sites in high- and low-constraint contexts are shown in Figure 4. As was true for Experiment 1, early components in all conditions include, at posterior sites, a positivity peaking at about 80 ms (P1), a negativity peaking at about 150 ms (N1), and a positivity peaking at about 200 ms (P2), and, at frontal sites, a negativity peaking at about 100 ms (N1) and a positivity peaking at about 175 ms (P2). Early components are followed by a broadly distributed negativity at 250 to 500 ms, peaking at about 350 ms (N400), which is smallest in response to the expected exemplars, largest to the between-category violations, and intermediate in amplitude to the within-category violations. The N400 is followed by a late positivity from 500 to 900 ms, which is smaller in amplitude to expected exemplars than to either within- or between-category violations.

Overall mean amplitude analyses. Mean voltage measures were taken in five time windows encompassing major components of the ERP: 50-150 ms (frontal N1), 150-250 ms (P2), 250-350 ms (early N400/N300), 350-500 ms (late N400), and 500-900 ms (late positivity). These measures were subjected to an omnibus ANOVA. Repeated measures included two levels of constraint (high vs. low), three levels of ending type (expected exemplar vs. within-category violation vs. between-category violation), and 26 levels of electrode. All p values in this and all subsequent analyses are reported after epsilon correction (Huynh-Felt) for repeated measures with more than 1 degree of freedom. Table 5 shows the ANOVA results.

Constraint influenced the ERP only in the form of a marginally significant Constraint x Ending Type interaction in the early N400/N300 time window. Ending type affected the ERP in the time windows encompassing the early N400/N300, the late N400, and the late positivity and showed a marginally significant effect in the time window of the P2; in all cases there was also a significant interaction with electrode, suggesting differences in the distribution of the response elicited by the three ending types. Because no significant effects were observed during the frontal N I window, it was dropped from follow-up analyses. To characterize the observed effects better, follow-up analyses were done for each constraint condition as a function of ending type and then for each ending type as a function of constraint.

Mean amplitude analyses for high-constraint sentences. Figure 5, Panel A shows the ERPs to the three ending types in high-constraint sentences at eight representative electrode sites. As was true in the overall analyses, in all time windows showing effects of ending type, the response to expected exemplars was most positive and that to between-category violations most negative; the response to within-category violations was generally intermediate in amplitude. To test these differences, mean voltage measures were taken in the same four time windows in which significant effects of experimental condition (constraint or ending type) had been observed in the overall analysis: 150-250 ms (P2), 250-350 ms (early N400/N300), 350-500 ms (late N400), and 500-900 ms (late positivity). Expected exemplars were compared with within-category violations, and within-category violations were compared with between-category violations in two ANOVAs. Repeated measures for each included two levels of ending type and 26 levels of electrode. Table 6 shows the results.

HIGH CONSTRAINT CONTEXTS   LOW CONSTRAINT CONTEXTS
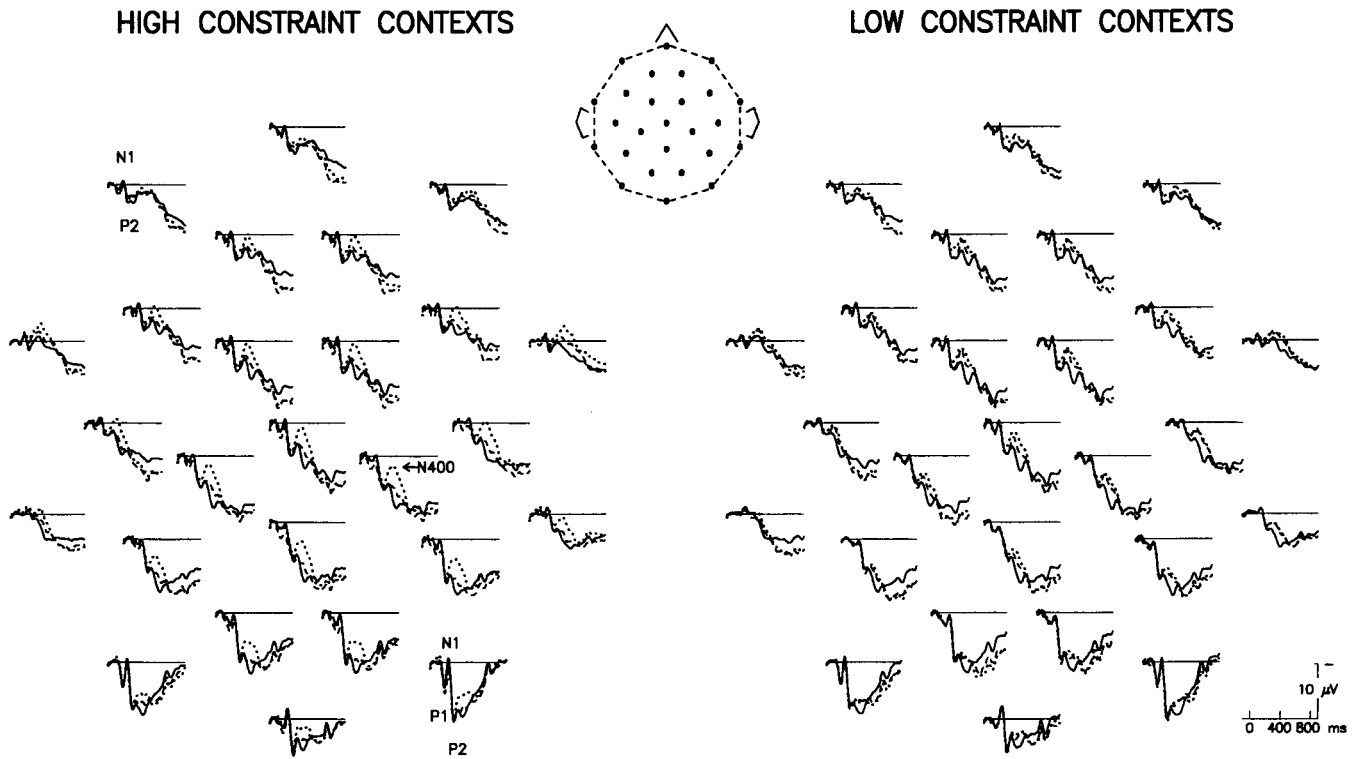
———— *Expected Exemplars*   ------ *Within-Category Violations*   ·········· *Between-Category Violations*

Figure 4.   Event-related potential (ERP) responses to preexposed expected exemplars (solid line), within-category violations (dashed line), and between-category violations (dotted line) in high-constraint (left) and low-constraint (right) contexts. Negative is plotted up. The 26 electrode site locations are laid out to approximate the arrangement over the head as seen from the top (illustrated by head icon). Major ERP components are labeled on left side.

Expected exemplars significantly differed from within-category violations only in the time window of the late positivity, although a marginally significant difference was observed for the early N400/N300 time window. In fact, further analyses show that expected exemplars are significantly more positive (i.e., show less N400 activity) than within-category violations (and also between-category violations, which are always at least as negative as within-category violations) between 300 and 450 ms, $F(1, 17) = 5.07$, $p < .05$; no interaction with electrode is observed. This effect then reverses in the time window of the late positivity, in which the response to violations of either type is significantly more positive than that to expected exemplars.

Table 5
Overall Results for Experiment 2

Time window (ms)	Constraint $F(1, 17)$	Constraint X Electrode $F(25, 425)$	Ending type $F(2, 34)$	Ending Type X Electrode $F(50, 850)$	Constraint X Ending Type $F(2, 34)$	Constraint X Ending Type X Electrode $F(50, 850)$
50-150	0.50	1.13	1.70	0.77	1.95	0.86
150-250	0.32	1.11	2.76, $p = .08$	**2.58****	0.22	0.59
250-350	0.00	1.17	**14.53****	**3.27****	2.47, $p = .10$	0.82
350-500	0.72	1.13	**8.15****	**2.78****	1.04	1.13
500-900	1.90	0.71	**6.36****	**2.98****	0.09	1.12

Note.   Significant effects are identified in boldface.
** $p < .01$.

A HIGH CONSTRAINT

B LOW CONSTRAINT

――――― Expected Exemplars   ------ Within-Category Violations   ········ Between-Category Violations
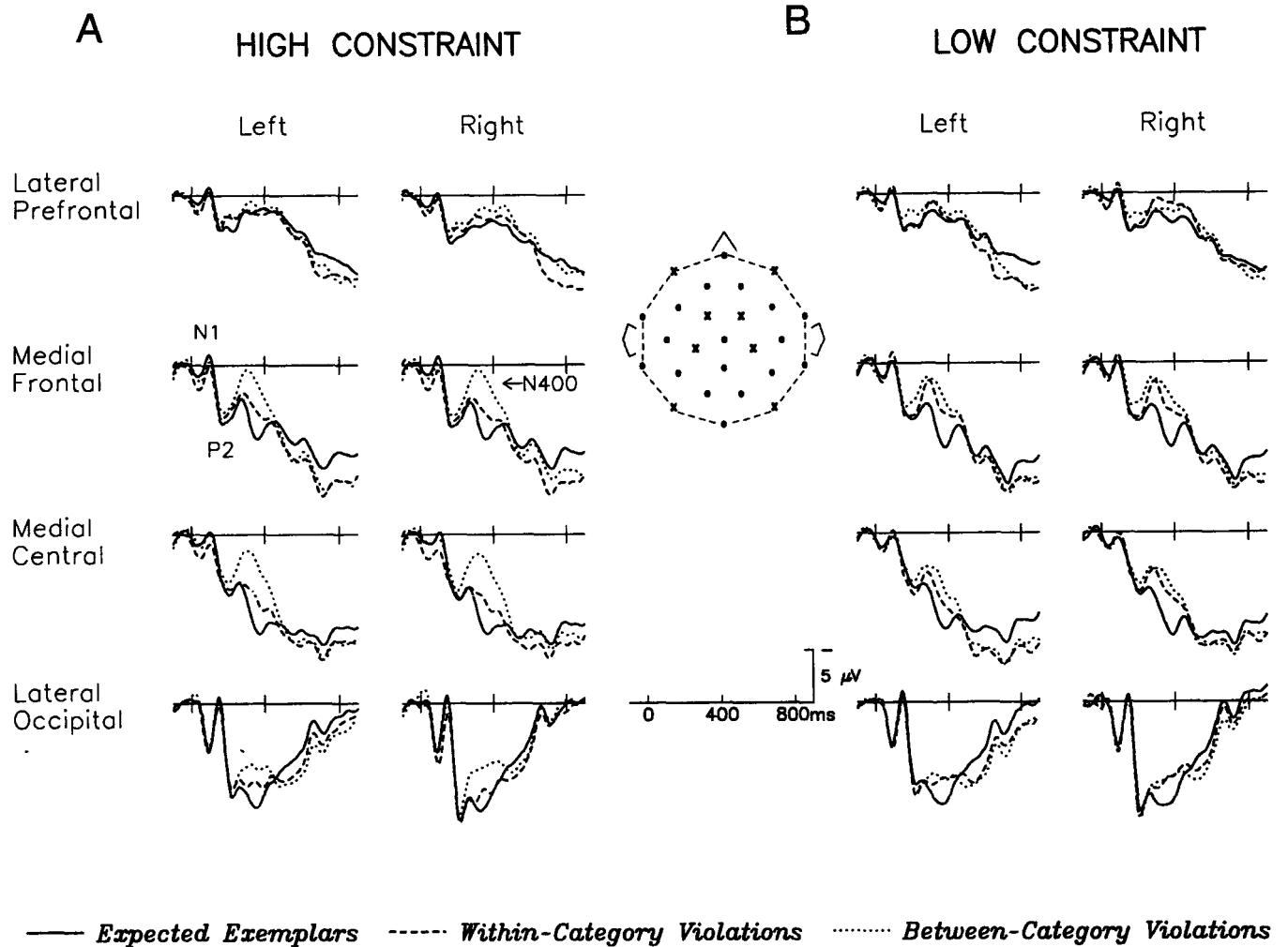
Figure 5.   Event-related potential responses to the three ending types in high-constraint (A) and low-constraint (B) contexts at eight representative electrode sites (indicated with **X**s on the head icon). Negative is plotted up. For both sentence types, expected exemplars (solid line) elicited the most positive responses in the 250- to 500-ms window, and between-category violations (dotted line) elicited greater negativity than within-category violations (dashed line).

Within-category violations and between-category violations differ significantly in both the early N400/N300 time window and the late N400 time windows (in both there is also a significant interaction with electrode) and show a marginally significant difference in the P2 time window. In these time windows, between-category violations are more negative than within-category violations. Also in both time windows, the response to within-category violations has a slightly greater medial to lateral slope than does the response to between-category violations; in addition, although the overall magnitude of response to the two violation types is similar over the most anterior and posterior sites, responses to between-category violations are more negative over more central sites-250-350: Ending Type X Laterality interaction, $F(1, 17) = 4.84$, $p < .05$, and Ending Type X Anteriority interaction, $F(3, 51) = 4.67$, $p < .01$; 350-500 ms: Ending Type X Laterality interaction, $F(1, 17) = 5.25$, $p < .05$, and marginal Ending Type X Anteriority interaction, $F(3, 51) = 2.83$, $p = .07$. In addition, in the late N400

time window, there is slightly more negativity over the right than the left for between-category violations, whereas the opposite pattern is observed for within-category violations (Ending Type X Hemisphere interaction, $F(1, 17) = 4.55$, $p < .05$).

In summary, in high-constraint sentences, the effect of category (difference between within- and between-category violations as a function of semantic similarity to the expected item) and the effect of context (difference between expected items and violations) are both first observed in the N400 time window (with the category effect beginning slightly earlier). Expected items and violations also continue to be distinguished in the late positivity from 500 to 900 ms.

Mean amplitude analyses for low-constraint sentences.   Figure 5, Panel B shows the ERPs to the three ending types in low-constraint sentences. Similar to effects seen in high-constraint sentences, in time windows showing an ending type effect the response to expected exemplars was most positive and that to between-category violations was most negative, with responses of

Table 6

Results for High-Constraint Comparisons

Time window (ms)	Expected vs. within $F(1, 17)$	Expected/Within X Electrode $F(25, 425)$	Within vs. between $F(1, 17)$	Within/Between X Electrode $F(25, 425)$
150-250	0.19	0.86	3.18, $p = .09$	0.31
250-350	3.39, $p = .08^a$	0.45	**18.00****	**3.74****
350-500	1.06^a	1.42	**6.35***	**2.47***
500-900	**6.73***	1.41	1.50	1.72

Note.   Significant effects are identified in boldface.
^{a} Marginal or nonsignificant results that are found to be significant for analyses in part of the time window or at a subset of all electrode sites.
* $p < .05$. ** $p < .01$.

generally intermediate amplitude to within-category violations. To test these differences, mean voltage measures were taken in the same four time windows and analyzed in the same manner as for the high-constraint condition (Table 7).

Expected exemplars significantly differed from within-category violations between 250 and 450 ms, $F(l, 17) = 5.39$, p < .05, and in the window encompassing the late positivity; no interactions with electrode are observed. In the earlier time window, expected exemplars were more positive (showed less N400 activity) than violations of either type (because between-category violations were always at least as negative as within-category violations), whereas this effect reversed in the later time window. When analyses are restricted to the eight medial-central electrodes (LMFr, RMFr, MiCe, LMCe, RMCe, MiPa, LDPa, and RDPa), in which N400 effects are typically largest, between-category violations are also found to be significantly more negative than within-category violations between 300 and 450 ms, $F(1, 17) = 4.91$, $p < .05$.

In summary, in low-constraint sentences, as in high, the effect of category (difference between within- and between-category violations as a function of semantic similarity to the expected item) and the effect of context (difference between expected items and violations) is first observed in the N400 time window (in this case with the effect of context beginning slightly earlier). Context continues to affect the ERP in the late positivity from 500 ms to beyond 900 ms.

Mean amplitude analyses of ending type across constraint. Figure 6 shows the ERPs to the three ending types as a function of constraint. Effects of constraint on mean voltages were analyzed

for each of the ending types at the same four time windows (150-250 ms, 250-350 ms, 350-500 ms, 500-900 ms) in three ANOVAs. Repeated measures for each included two levels of constraint (high vs. low) and 26 levels of electrode (Table 8).

No effects of constraint are observed on the response to expected exemplars or to within category in any of the time windows used here. For between-category violations, constraint has a significant effect between 300 and 450 ms, $F(1, 17) = 4.64$, $p < .05$; during this time period, the response is more negative in high- than in low-constraint contexts. This effect does not interact with electrode.

Direct comparisons between data from Experiments 1 and 2. Figure 7 shows the ERP to target pictures (collapsed across ending type and constraint) in Experiment 1 overlapped with that to the same pictures when preexposed in Experiment 2. Preexposure seemed to affect the amplitude of several early components and of the early N400/N300 over frontal sites. To assess these effects, we measured mean amplitudes in the 50- to 110-ms (P1) and 150- to 250-ms (P2) time windows at the 5 most posterior sites (LMOc, RMOc, MiOc, LLOc, RLOc) and in the 50- to 150-ms (NI) and 275- to 325-ms (early N400/N300) time windows at the 11 most anterior sites (MiPf, LLPf, RLPf, LMPf, RMPf, LLFr, LDFr, LMFr, RMFr, RDFr, RLFr). Preexposure (two levels) was a between-participants variable, whereas repeated measures included two levels of constraint (high vs. low), three levels of ending type (expected exemplar vs. within-category violation vs. between-category violation) and 5 or 11 levels of electrode.

Preexposure significantly affected the amplitude of both the posterior P1, $F(1, 34) = 3.97$, $p = .05$, and the posterior P2, $F(1,$

Table 7

Results for Low-Constraint Comparisons

Time window (ms)	Expected/within $F(1, 17)$	Expected/Within X Electrode $F(25, 425)$	Within between $F(1, 17)$	Within/Between X Electrode $F(25, 425)$
150-250	0.09	1.03	1.82	1.48
250-350	**5.19***	1.04	0.78^a	0.79
350-500	2.21'	0.84	0.99'	1.17
500-900	**4.46***	1.52	0.20	1.18

^{a} Marginal or nonsignificant results that are found to be significant for analyses in part of the time window or at a subset of all electrode sites.
* p < .05.

Expected Exemplars   Within-Category Violations   Between-Category Violations

Frontal

Central

Occipital

N400

—— High Constraint   ·········· Low Constraint
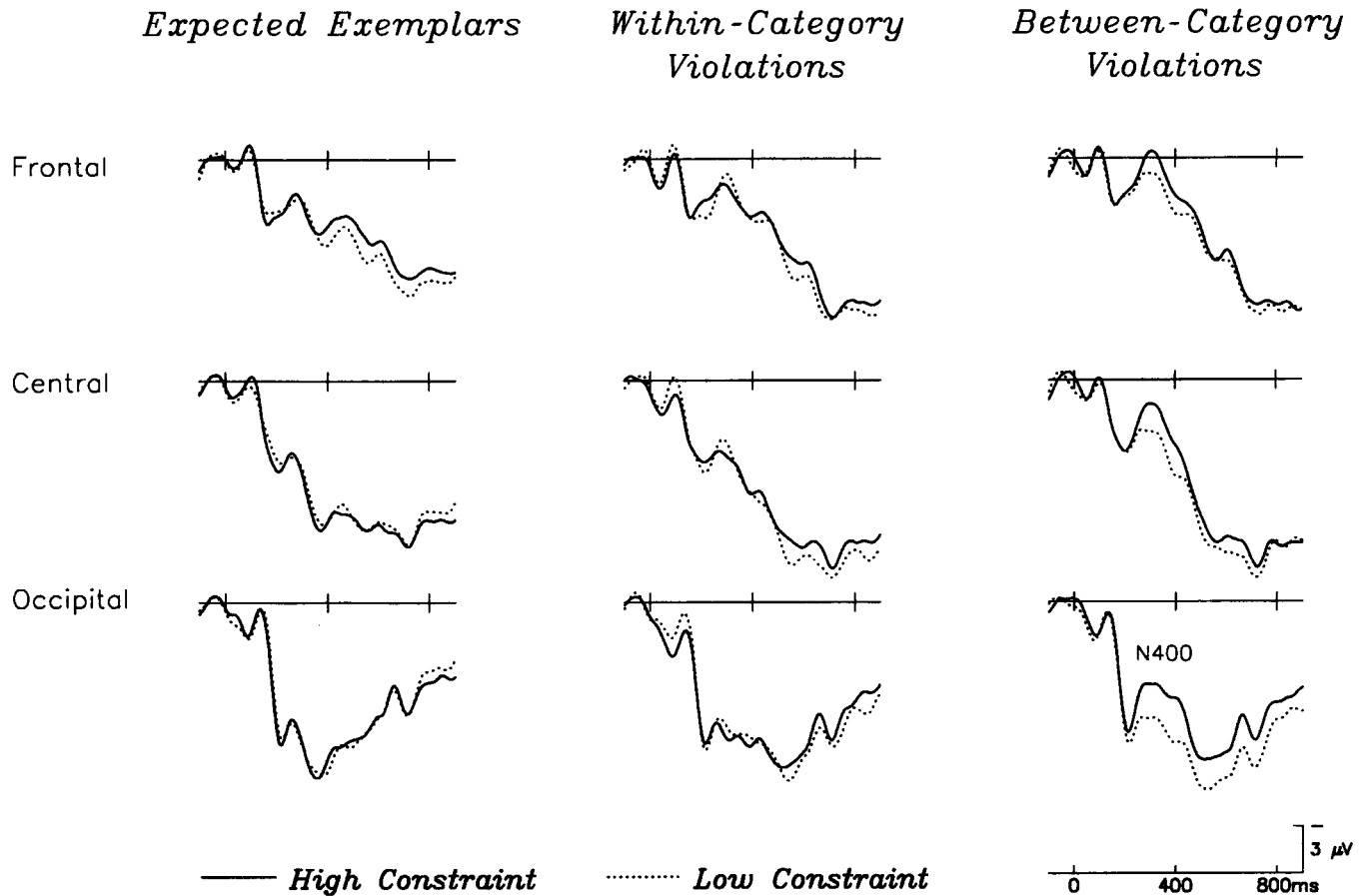
0   400   800ms

3 μV

Figure 6. Responses to preexposed expected exemplars, within-category violations, and between-category violations in high-constraint (solid line) and low-constraint (dotted) contexts at three representative electrode sites. Negative is plotted up. Constraint did not influence the response to either expected exemplars or within-category violations. Between-category violations elicited more negative responses 300 to 450 ms after stimulus onset in high- than in low-constraint contexts.

34) = 5.53, $p < .05$, both being smaller for preexposed (than not) pictures. Preexposure also affected the frontal N1 response, in the form of a significant Preexposure X Constraint X Ending Type interaction, $F(2, 68) = 5.14, p < .01$. For most conditions, NIs were of smaller amplitude in Experiment 2 (with preexposure) than in Experiment 1 (without preexposure). However, for ex-

pected exemplars in high-constraint contexts, N1s were of similar amplitude regardless of preexposure. Finally, there was a trend toward smaller (frontal) N400 responses in Experiment 2 versus Experiment 1, $F(1, 34) = 2.88, p = .10$.

Effects across experiments. To facilitate comparisons across the two experiments described here and that of Federmeier and

Table 8

Results for Each Ending Type

Time window (ms)	Expected (high vs. low constraint) $F(1, 17)$	Constraint X Electrode $F(25, 425)$	Within (high vs. low constraint) $F(1, 17)$	Constraint X Electrode $F(25, 425)$	Between (high vs. low constraint) $F(1, 17)$	Constraint X Electrode $F(25, 425)$
150-250	0.59	0.30	0.01	0.93	0.03	0.92
250-350	0.09	0.48	1.77	0.61	2.53[a]	1.78
350-500	0.17	0.84	0.30	1.03	3.41, $p = .08$[a]	1.64
500-900	0.23	1.18	0.20	0.84	1.32	0.95

[a] Marginal or nonsignificant results that are found to be significant for analyses in part of the time window or at a subset of all electrode sites.
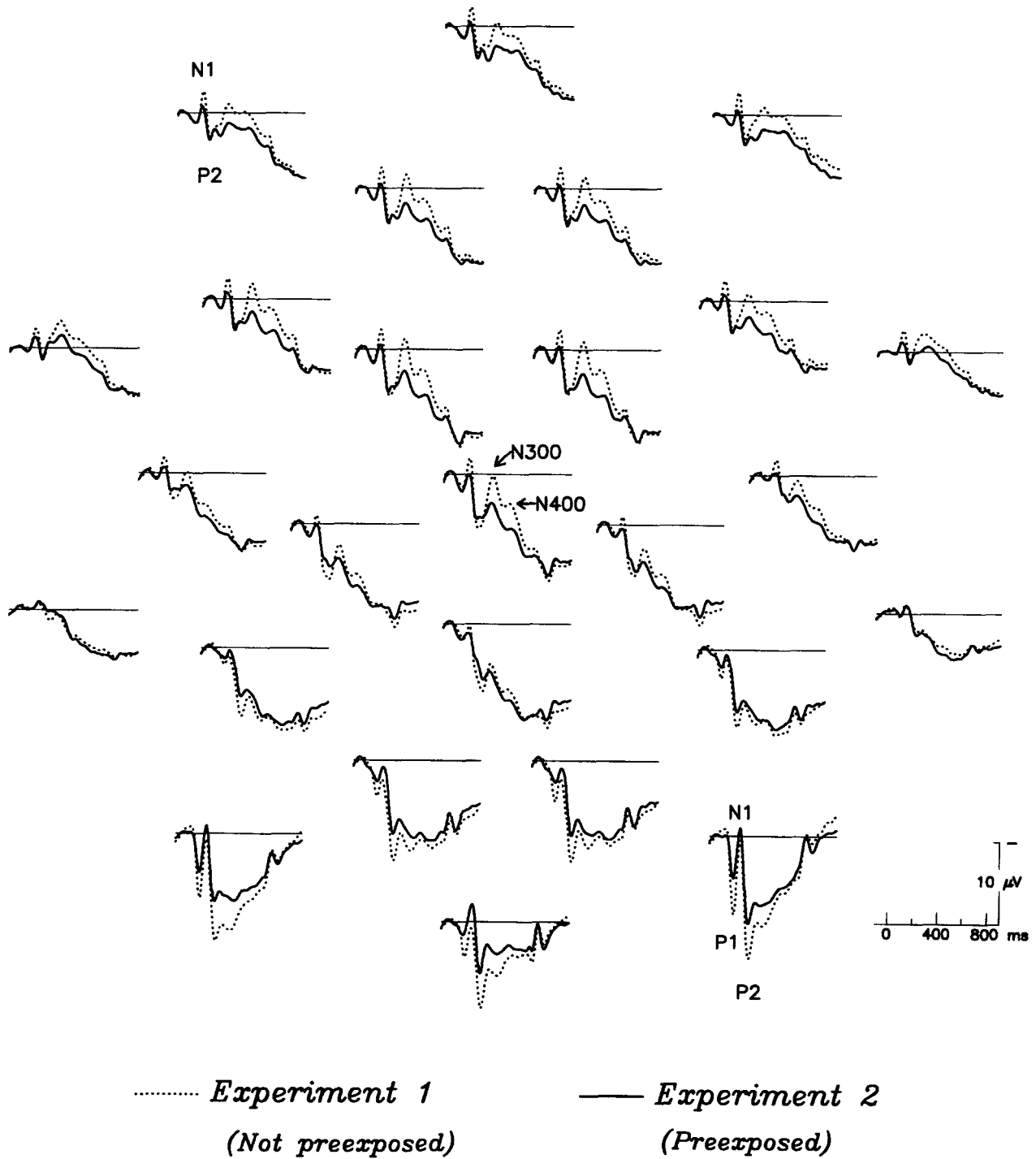
Figure 7. Responses to pictures (collapsed across ending type) in Experiment 1 (dotted line) compared with Experiment 2 (solid line), in which the pictures were preexposed. All 26 electrode sites are shown, laid out to approximate the arrangement over the head going from front (top) to back (bottom). Negative is plotted up. Preexposure reduced the amplitude of the frontal N1 and the posterior PI and P2 and tended to reduce the amplitude of the N300/N400 over frontal sites.

Kutas (1999), Table 9 lists the major effects and indicates whether each held for words (data from Federmeier and Kutas, 1999), for pictures [Experiment 1], and for preexposed pictures [Experiment 2]. Because this article focuses on modality-related differences in the pattern of response to input variables and not on topography, topographical effects are not listed. In Table 9, where no "X" appears, the comparison conditions did not differ from one another. Also note that Table 9 does not reflect main effects across experiments, such as the previously described general reduction in the amplitude of the N1, P1, and P2 as a result of preexposure. In addition, Figure 8 compares the waveforms (at a representative channel) elicited in each of the three experiments as a function of ending type and constraint.

Discussion

Relative to the ERP response observed in Experiment 1, the responses in Experiment 2 to the same line drawings when these were preexposed (i.e., familiar) were characterized by amplitude reductions of early visual processing and attention-related components, including the posterior P1 and P2 and the anterior N1. As we hypothesized, prior experience with a picture seems to reduce visual processing load and allow more efficient allocation of attentional resources. The reductions in the amplitude of the posterior P1 and P2 are simple main effects of preexposure, unaffected by either the relation of the target to the context or the strength of the context. These components thus reflect "stimulus-driven" processes influencing the ease of perception. The anterior N1 component is also reduced in amplitude by preexposure and is thus generally smaller for familiar than for unfamiliar pictures with one important exception. Unfamiliar pictures that serve as congruent endings in highly constraining sentence contexts elicit anterior N1s, which are just as small as those observed for familiar pictures. At least for the processes reflected in the anterior N1, then, perceptual difficulty/novelty seemingly can be compensated for by a congruent and constraining context. The data thus indicate that by about 100 ms bottom-up factors (e.g., perceptual familiarity) and top-down factors (e.g., expectancy for an item based on prior context) come together to affect visual processing.

Despite a tendency toward a reduction in N300/N400 amplitudes for preexposed (vs. unfamiliar) pictures, the general influ-

Table 9
Patterns of Effects

Effect	Words	Pictures	Preexposed pictures
General			
N400: within > expected	X	X	X
N400: between > within	X	X	X
Constraint			
N1, expected exemplar: low > high		X	
P2, expected exemplar: high > low		X	
N400, expected exemplar: low > high		X	
N400, within-category violation:			
low > high	X		
N400, between-category violation:			
high > low		X	X

Note: X = Comparison conditions differed from one another.

ence of both context and category membership on the ERP between 250 and 500 ms is functionally the same regardless of picture familiarity. In both experiments, the ERP to expected completions is less negative than that to unexpected completions at both levels of constraint. In addition, unexpected completions derived from the same category as (and, therefore, sharing many semantic features with) expected items generate less N400 activity than unexpected completions from a different semantic category. We again find that sentence context information and the organization of semantic memory influence the ERP in a qualitatively similar fashion. The organization of long-term memory thus seems to affect processing in a manner that is relatively independent of both modality and perceptual familiarity.

At the same time, however, we do observe an ERP difference between familiar and unfamiliar pictures after the initial effect of congruity on the N400. Unlike the pattern for unfamiliar pictures, for familiar pictures the amplitude of a late positive component, between 500 ms after stimulus onset and the end of the epoch and beyond, varies with congruity. In this time window, expected items generate less positive responses than violations of either type (which do not differ), regardless of degree of contextual constraint. This effect is similar to that seen in sentence repetition paradigms, in which reduced late positive components have been observed to congruous relative to incongruous sentence-final words in repeated sentences (Besson, Kutas, & Van Petten, 1992). Van Petten, Kutas, Kluender, Mitchiner, and McIsaac (1991) suggested that this reduction in the late positive component to congruent items reflects the recollection of the repeated item, aided by the repeated sentence context. Note that in the present study context information was not repeated. However, a congruent sentence context may nevertheless have led to the retrieval of the expected picture and thus a reduction in the late positivity when that picture was, in fact, presented. As might be expected, therefore, preexposure influenced both perception of and memory for the experimental pictures. These effects could be differentiated in the ERP response; perceptual predictability influenced early components linked to visual processing, and attentional allocation and memorability influenced later components that have been linked to episodic retrieval.

The primary aim of Experiment 2, however, was to determine whether perceptual familiarity-predictability (preexposure) would alter any of the processing differences between pictures and words noted in Experiment 1, so that we could determine which differences were modality specific and which were not. The results clearly show that some of the observed picture-word differences were due to perceptual predictability rather than modality. For example, in Experiment 1 for unfamiliar pictures in high-constraint sentences the influence of context preceded that of category by about 250 ms, whereas in low-constraint sentences these two variables first impacted the ERP at about the same time in the N300/N400 time window. We had not observed such an effect of contextual constraint on the response to expected words. Nor did we see such an effect with pictures once they were preexposed (as in Experiment 2); for these familiar pictures, as for words, both context and category first influence the ERP in the N300/N400 time window regardless of contextual constraint. It seems, therefore, that when targets are relatively easy to perceive, as in the case of words and familiar line drawings, context has its primary impact on processes related to semantic integration. In contrast, when

Expected Exemplars

Within-Category Violations
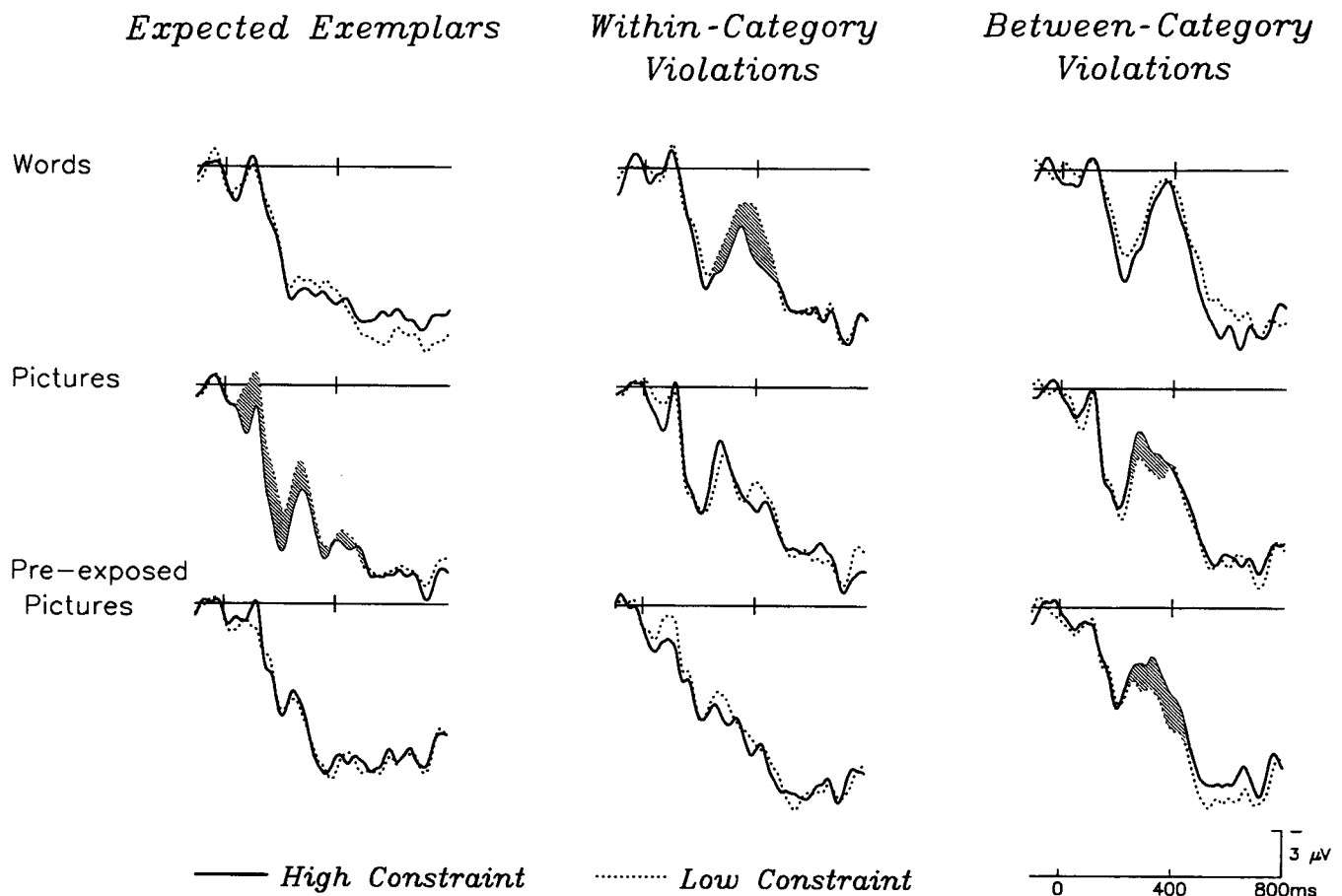
Between-Category Violations

Figure 8.   Responses to words (Federmeier & Kutas, 1999), pictures, and preexposed pictures as a function of ending type and contextual constraint. Responses are shown at a representative channel (midline parietal) in which N400 effects are typically prominent. Negative is plotted up. Note that the differences in the effects of constraint on expected exemplars observed between words and pictures are likely to be a function of perceptual predictability, because the response to preexposed pictures is similar to that for words for these items. In contrast, the difference in the pattern of response to violations as a function of constraint likely reflects a true modality-related difference in semantic processing, because it holds for both unfamiliar and preexposed pictures.

perception is more difficult (as for pictures seen for the first time), strong contextual information like that available in highly constraining contexts affects earlier processing stages related to attentional allocation and perceptual analysis as well as later ones related to semantic integration.

Other consequences of stimulus predictability are also evident when we examine the influence of contextual constraint on the three ending types directly. Although the response to unfamiliar expected exemplars (Experiment 1) varied in amplitude as a function of constraint in every time window measured, that to familiar expected exemplars (Experiment 2) did not vary in any time window. Here again, the results for preexposed pictures correspond to the results originally observed for words and not to those observed for unfamiliar pictures. Perceptual, and not just semantic, predictability, therefore, seems to influence the response to those items for which predictions can actually be made (i.e., expected endings). In short, when the features of a semantically predictable item are also perceptually predictable, then that item, be it a word or a picture, can be integrated as well with a congruent context that

is strongly constraining as with one that is less constraining. The different pattern of responses observed for the same pictures in exactly the same sentence contexts when they have not been seen previously-and thus when their perceptual features are not predictable-indicates that strongly predictive contexts lead to expectations about perceptual and not just semantic features of upcoming items. The ERPs to unfamiliar pictures show a sensitivity to contextual constraint on early components that is not observed for either preexposed pictures or words. Moreover, preexposure leads to differences in semantic processing. Whereas the same unfamiliar pictures that elicit reduced N1 and increased P2 components (and that are thus presumably processed more easily at a perceptual level) in highly constraining contexts also elicit reduced N400 responses, no differences in semantic integration are observed for these pictures if they were preexposed. Apparently, the semantic processing of the same stimuli in the same contexts is altered when conditions relating to the perceptual processing of those stimuli are changed. In sum, we are led to conclude that semantic processing is not isolated from perceptual factors, as strong forms of single

code models often imply. Instead, there seems to be an important link, perhaps mediated through attentional resource allocation, between the ease with which a stimulus can be perceptually decoded and the nature of the semantic information subsequently derived from that stimulus and available for integration into a sentence context.

Visual feature predictability is, of course, much less important for within- and between-category violations, because these items are not predicted in the sentence contexts, and, in fact, preexposure of the pictures did not seem to change the pattern of response to these items. Recall that for words the only influence of contextual constraint to reduce the N400 to within-category violations in high relative to low-constraint contexts; the response to between-category violations was unaffected by contextual constraint. In Experiment 1, however, we found no influence of contextual constraint on the response to within-category picture violations, together with increased N400 amplitudes to between-category picture violations in high- as opposed to low-constraint sentences. In Experiment 2, we see the same pattern, with, if anything, an even larger effect on the between-category (familiarized) picture violations. Therefore, the difference, first noted in Experiment 1, in the pattern of response to the two violation types when these are presented as words versus pictures seems to be a true modality-based difference. This effect is not altered by making the perceptual processing of words and pictures more similar. For words, the processing of unexpected, categorically related items is facilitated by an increase in contextual strength, an increase in the predictability of the semantic features of the expected category exemplar. In contrast, for pictures it is the processing of unexpected and unrelated items that is altered by contextual strength. This effect goes in the direction of the rated plausibility of these items in their sentence contexts, suggesting that strong contexts may provide a larger barrier to the integration of these unexpected items than weaker contexts. Because the contextual information and the task were exactly the same for word as for picture processing, the observed modality difference suggests that the semantic information derived from words is not the same as that derived from pictures of the same concepts.

General Discussion

Prior electrophysiological work, like some behavioral work before it, has suggested that words and pictures are affected similarly by manipulations of fit to a semantic context. Integrating either type of stimulus with a related prime (e.g., Barrett & Rugg, 1990; Bentin et al., 1985; Holcomb & McPherson, 1994) or a congruent sentence context (e.g., Ganis et al., 1996; Kutas & Hillyard, 1980, 1984; Nigam et al., 1992) reduces the negativity between 250 and 500 ms after stimulus onset (N400). Furthermore, the N400 reduction is similar for words and pictures in both effect size and time course when visual aspects of the stimuli (such as size and spatial frequency) are controlled for (Ganis et al., 1996). This functional similarity has been taken to support models postulating a single semantic store shared by pictures and words (Nigam et al., 1992). However, even under carefully controlled conditions, the N400 effect to words and pictures differs in scalp distribution, implicating nonidentical neural generators (Ganis et al., 1996; Holcomb & McPherson, 1994) and, in turn, the possibility of processing differences between the two in response to

manipulations of input factors such as semantic context or relatedness, for example. The goal of this study, therefore, was to use ERPs to examine picture processing such that it could be compared qualitatively with word processing in much more detail than has been done heretofore. To that end, we examined the response to three types of sentence-final targets (contextually expected items, unexpected items from the same semantic category as the contextually expected item, and unexpected items from a different semantic category) in strongly constraining and more weakly constraining sentence contexts. In Experiment 1, targets were line drawings that were unfamiliar to the participants before their presentation as experimental stimuli, whereas in Experiment 2 participants were familiarized with all the line drawings before the recording session in an attempt to make the perceptual predictability of the words and pictures more similar.

Like prior ERP studies, we observe a general similarity between the response to words (Federmeier & Kutas, 1999) and the response to pictures. There did seem to be some distributional differences as a function of modality; the N400 elicited by pictures was more prominent over frontal sites than that typically described for words. However, because our word and picture data were collected from different participants, strong statements about modality-based distributional differences are not possible here. Moreover, for the reasons already discussed in the introduction, the focus of this study was not on topographical differences but rather on functional differences (i.e., differences in how the processing of pictures and words is affected by factors such as semantic similarity and contextual strength). We observed, as have prior studies, that for both words and pictures the ERP response between 250 and 500 ms reflected an item's fit to the sentence context. Previously, we had shown that the N400 to a word is sensitive not only to contextual fit but also to the organization of information in long-term semantic memory (Federmeier & Kutas, 1999). Here we show that this is also true for pictures, whether familiar or unfamiliar. In both experiments with pictures, unexpected items that shared significant numbers of semantic features in common with the item predicted by the sentence context were integrated with the sentence context more easily than those not sharing as many features in common. From this we conclude that, as for words, the semantic processing of pictures is influenced simultaneously both by fit to context and the context-independent organization of semantic memory. In turn, it seems that the kind of perceptual and functional similarity underlying semantic categories must form at least part of the basis for this memory organization, whether the input is a word or a picture. Thus, in line with the predictions of common code models, the organization of the semantic knowledge store that is accessed by pictures and words seems to be basically similar.

Although we find similarity between picture and word processing at a general level, the more finely structured design of our experiment uncovered patterns of modality-related differences that have not been reported previously. First, we find that semantic analysis is affected by perceptual factors. The semantic integration of perceptually predictable items (words or pictures) into a congruent context is unaffected by contextual strength. Thus, when perceptual as well as semantic features can be anticipated, highly constraining contexts do not seem to provide an advantage for the semantic integration of congruent items over what is provided by less predictive context. The perceptual predictability is critical,

however, because a different pattern emerges when participants must integrate into context pictures that they are seeing for the first time. In fact, picture processing seems to be generally more difficult when pictures are unfamiliar, as evidenced by increased amplitudes of early ERP components associated with visual processing and the allocation of visuospatial attention. Under these more perceptually taxing circumstances, contextual strength seems to play more of a role not only in perceptual but in semantic processing as well. Specifically, the ERPs indicate that attentional load is reduced-and semantic integration is correspondingly facilitated-in strong as opposed to weaker contexts. Because the contexts and the stimuli are exactly the same in the two picture experiments, this difference between them can only be due to a difference in the participants' perceptual experience with the stimuli.

The assumption of common code models is that pictures and words, once they have been perceptually processed and identified, impinge on a central semantic store that uses an abstract and amodal internal code (see, e.g., W. R. Glaser, 1992, for review). In these models, then, there is a clear demarcation between perceptual and semantic processing. Differences in perceptual aspects of stimuli can make visual processing more difficult or time consuming, but once the stimulus has been identified and has entered the semantic system, such differences, which are clearly modality specific, are presumed to vanish. Our results do not square with this description, because we find that perceptual factors interact with stimulus congruency and contextual strength; semantic processing of the same stimulus in the same context is altered by perceptual experience. This is the case even though our stimuli were never perceptually ambiguous nor particularly visually taxing or difficult to identify. In fact, presenting line drawings without preexposure-our "perceptually difficult" condition-is the norm for essentially all prior behavioral and ERP work investigating picture processing. We find that the semantic integration of congruent picture and word information is identical only when the predictability of the two is fairly similar (although it is likely that the predictability of pictures after only a single exposure is still somewhat less than the predictability of word forms), and we find that semantic feature analysis and contextual integration differ when the item is perceptually predictable in that context than when it is not. We believe that this link between perceptual and semantic processing is difficult to reconcile with common code accounts, as they are typically formulated.

Analysis of sententially incongruent endings also revealed differences in the semantic feature information extracted from pictures and words that are independent of perceptual predictability and familiarity. Increased contextual strength facilitated the processing of incongruent words that were semantically related to the expected, congruent sentence completion (Federmeier & Kutas, (1999) and had no effect on the processing of incongruent words that did not bear such a close semantic relationship with the expected completion. In contrast, in both experiments, increased contextual strength seemed to make the processing of incongruent and unrelated pictures more taxing but did not influence the processing of those incongruent pictures that were related. Because the contexts and the target concepts were the same in all cases, the fact that different patterns were observed as a function of modality ¡mplies that different semantic information must be derived from a word than from a corresponding picture.

Further research is needed to determine exactly what kind of semantic information is activated by words and by pictures. However, we believe that the pattern of results is consistent with the hypothesis that pictures may activate certain more specific semantic feature information than words. The words "pterodactyl" and "triceratops," for example, both refer to large, familiar prehistoric animals. At first glance, therefore, a sentence pair from this experiment, such as "I'm always amused by the Flintstones' version of an airplane. Everyone piles onto a triceratops" may seem felicitous, because many features of the final word cohere at a general level with those predicted by the context (cf. the semantic "illusion" effect; Erickson & Mattson, 1981; Kamas, Reder, & Ayers, 1996; Van Oostendorp & de Mul, 1990). Even (or perhaps especially) in a highly constraining context like this, the critical difference-that pterodactyls have wings whereas triceratops do not-may be swamped by the more general similarities in the semantic features associated with these words, allowing facilitation for the semantically related but inappropriate item. However, the crucial difference between a pterodactyl and a triceratops is, in this case, explicitly represented in the line drawings of these two animals, thereby making the facilitation of the incorrect semantically related item much less likely for pictures than for words. The activation of more specific semantic feature information may also explain why pictures (both unfamiliar and preexposed), but not words, show an increased barrier to integration for inappropriate, unrelated targets in more constraining contexts. Of course, this hypothesis implies that the degree to which word and picture processing are observed to diverge in any given case will be a function of many variables, including the nature of the items and the sentence contexts involved. Modality differences could be deemphasized by constructing contexts in which critical differences between items are based on fairly abstract features that would not be explicitly represented in pictures. Similarly, such differences would presumably be accentuated in contexts that emphasized visual feature information. However, the precise nature of the difference in semantic feature information activated by pictures and words-and its degree of contextual dependency-notwithstanding, we believe that the very existence of a reliable difference (in any context) runs counter to the predictions of all single code models.

In the end, then, what do the results of these experiments-in combination with previous behavioral and electrophysiological findings-imply for the single versus multiple semantic systems debate? On the one hand, (a) there do not seem to be significant differences in the time course with which pictures and words are integrated into a sentence context, (b) picture and word processing in sentence contexts do not elicit radically different brain responses, and (c) the organization of the semantic information gleaned from pictures and words does not seem to be strikingly different. These findings make it seem unlikely that pictures and words are processed in completely independent systems. Alternatively, (a) pictures and words do behave differently in a number of tasks and do seem to activate different semantic feature information even in the same contexts; furthermore, (b) perceptual factors do seem to impact semantic processing. These findings make it seem unlikely that semantic processing can truly be considered amodal. Part of the difficulty is that these two accounts differ along multiple lines: shared versus independent, amodal/abstract versus modality specific. We believe that it is possible for aspects of both

models to be correct (e.g., for semantic information to exist in a shared, distributed system but one whose coding scheme is not strictly amodal). In fact, increasing evidence suggests that semantic information may be distributed over multiple cortical areas that each preferentially process information from a particular modality. Although all areas may be involved in the representation of semantic information in general, the relative contribution of the various areas will differ for different types of information (see, e.g., Small, Hart, Nguyen, & Gordon, 1995, for an example of a model of semantic memory that uses this structure). The representation of action verbs, for example, may preferentially involve frontal areas related to motor processing, whereas the representation of objects may rely more heavily on posterior, visual processing areas (e.g., Damasio & Tranel, 1993). Similarly, the representation of concrete, imageable words may rely more heavily on posterior vision-related areas than the representation of more abstract words (e.g., Holcomb, Kounios, Anderson, & West, 1999). In this view, one would expect to find general functional similarity between the semantic processing of pictures and words, because the semantic information derived from both exists in a shared system; this is, in fact, what we observe. However, the representation derived from a word and a picture of the "same" concept (or even that derived from the same picture under different perceptual conditions) may often be partially nonoverlapping, resulting in slightly different ERP scalp distributions (e.g., Ganis et al., 1996) and, more importantly, different patterns of facilitation in the same sentence contexts, as we reported here.

References

Bajo, M.-T. (1988). Semantic facilitation with pictures and words. *Journal of Experimental Psychology: Learning, Memory, and Cognition,* 14(4), 579-589.

Barrett, S. E., & Rugg, M. D. (1990). Event-related potentials and the semantic matching of pictures. *Brain & Cognition,* 14(2), 201-212.

Bentin, S., McCarthy, G., & Wood, C. C. (1985). Event-related potentials associated with semantic priming. *Electroencephalography and Clinical Neurophysiology,* 60, 343-355.

Besson, M., Kutas, M., & Van Petten, C. (1992). An event-related potential (ERP) analysis of semantic congruity and repetition effects in sentences. *Journal of Cognitive Neuroscience, 4(2),* 132-149.

Biederman, I., & Tsao, Y.-C. (1979). On processing Chinese ideographs and English words: Some implications from Stroop-test results. *Cognitive Psychology,* 11(2), 125-132.

Boddy, J., & Weinberg, H. (1981). Brain potentials, perceptual mechanisms and semantic categorisation. *Biological Psychology,* 12(1), 43-61.

Brown, W. (1915). Practice in associating color-names with colors. *Psychological Review,* 22(1), 45-55.

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

Carr, T. H., McCauley, C., Sperber, R. D., & Parmelee, C. M. (1982). Words, pictures, and priming: On semantic activation, conscious identification, and the automaticity of information processing. *Journal of Experimental Psychology: Human Perception and Performance, 8(6),* 757-777.

Cattell, J. M. (1886). The time to see and name objects. *Mind, 11,* 63-65.

Clark, V. P., & Hillyard, S. A. (1996). Spatial selective attention affects early extrastriate but not striate components of the visual evoked potential. *Journal of Cognitive Neuroscience, 8(5),* 387-402.

Damasio, A. R., & Tranel, D. (1993). Nouns and verbs are retrieved with

differently distributed neural systems. *Proceedings of the National Academy of Sciences, USA,* 90, 4957-4960.

Dunbar, K., & MacLeod, C. M. (1984). A horse race of a different color: Stroop interference patterns with transformed words. *Journal of Experimental Psychology: Human Perception and Performance,* 10(5), 622-639.

Erickson, T. D., & Mattson, M. E. (1981). From words to meaning: A semantic illusion. *Journal of Verbal Learning & Verbal Behavior,* 20(5), 540-551.

Federmeier, K. D., & Kutas, M. (1999). A rose by any other name: Long-term memory structure and sentence processing. *Journal of Memory and Language,* 41(4), 469-495.

Fraisse, P. (1969). Why is naming longer than reading? *Acta Psychologica,* 30, 96-103.

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.

Glaser, M. O., & Glaser, W. R. (1982). Time course analysis of the Stroop phenomenon. *Journal of Experimental Psychology: Human Perception and Performance, 8(6),* 875-894.

Glaser, W. R. (1992). Picture naming. Special issue: Lexical access in speech production. *Cognition,* 42(1-3), 61-105.

Glaser, W. R., & Dungelhoff, F.-J. (1984). The time course of picture-word interference. *Journal of Experimental Psychology: Human Perception and Performance,* 10(5), 640-654.

Glaser, W. R., & Glaser, M. O. (1989). Context effects in Stroop-like word and picture processing. *Journal of Experimental Psychology: General,* 118(1), 13-42.

Hillyard, S. A., & Muente, T. F. (1984). Selective attention to color and location: An analysis with event-related brain potentials. *Perception & Psychophysics,* 36(2), 185-198.

Holcomb, P. J., Kounios, J., Anderson, J. E., & West, W. C. (1999). Dual-coding, context-availability, and concreteness effects in sentence comprehension: An electrophysiological investigation. *Journal of Experimental Psychology: Learning, Memory, and Cognition,* 25(3), 721-742.

Holcomb, P. J., & McPherson, W. B. (1994). Event-related brain potentials reflect semantic priming in an object decision task. *Brain & Cognition,* 24(2), 259-276.

Irwin, D. I., & Lupker, S. J. (1983). Semantic priming of pictures and words: A levels of processing approach. *Journal of Verbal Learning & Verbal Behavior,* 22(1), 45-60.

Kamas, E. N., Reder, L. M., & Ayers, M. S. (1996). Partial matching in the Moses illusion: Response bias not sensitivity. *Memory & Cognition,* 24(6), 687-699.

Kutas, M. (1993). In the company of other words: Electrophysiological evidence for single-word and sentence context effects. Special issue: Event-related brain potentials in the study of language. *Language & Cognitive Processes, 8(4),* 533-572.

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

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

Kutas, M., Neville, H. J., & Holcomb, P. J. (1987). A preliminary comparison of the N400 response to semantic anomalies during reading, listening, and signing. *Electroencephalography and Clinical Neurophysiology Supplement,* 39, 325-330.

Kutas, M., & Van Petten, C. (1990). Electrophysiological perspectives on comprehending written language. In P. M. Rossini & F. Maugiere (Eds.), *New trends and advanced techniques in clinical neurophysiology* (pp. 155-167). Amsterdam: Elsevier Science.

Kutas, M., & Van Petten, C. K. (1994). Psycholinguistics electrified: Event-related brain potential investigations. In M. A. Gemsbacher (Ed.),

Handbook of psycholinguistics (pp. 83-143). San Diego: Academic Press.

La Heij, W., Dirkx, J., & Kramer, P. (1990). Categorical interference and associative priming in picture naming. *British Journal of Psychology,* 81(4), 511-525.

La Heij, W., & Vermeij, M. (1987). Reading versus naming: The effect of target set size on contextual interference and facilitation. *Perception & Psychophysics,* 41(4), 355-366.

Ligon, E. M. (1932). A genetic study of color naming and word reading. *American Journal of Psychology,* 44, 103-122.

Luck, S. J., & Hillyard, S. A. (1994). Electrophysiological correlates of feature analysis during visual search. *Psychophysiology,* 31(3), 291-308.

Lund, F. H. (1927). The role of practice in speed of association. *Journal of Experimental Psychology, 10,* 424-434.

Mangun, G. R., Hillyard, S. A., & Luck, S. J. (1993). Electrocortical substrates of visual selective attention. In D. E. Meyer & S. Kornblum (Eds.), *Attention and performance* 14: *Synergies in experimental psychology, artificial intelligence, and cognitive neuroscience (pp.* 219-243). Cambridge, MA: MIT 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,* 62(3), 203-208.

McCauley, C., Parmelee, C. M., Sperber, R. D., & Carr, T. H. (1980). Early extraction of meaning from pictures and its relation to conscious identification. *Journal of Experimental Psychology: Human Perception and Performance, 6(2),* 265-276.

McPherson, W. B., & Holcomb, P. J. (1999). An electrophysiological investigation of semantic priming with pictures of real objects. *Psychophysiology,* 36(1), 53-65.

Nigam, A., Hoffman, J. E., & Simons, R. F. (1992). N400 to semantically anomalous pictures and words. *Journal of Cognitive Neuroscience, 4(1),* 15-22.

Oldfield, R. C. (1971). The assessment and analysis of handedness: The Edinburgh Inventory. *Neuropsychologia, 9,* 97-113.

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(3), 255-287.

Pellegrino, J. W., Rosinski, R. R., Chiesi, H. L., & Siegel, A. (1977). Picture-word differences in decision latency: An analysis of single and dual memory models. *Memory & Cognition, 5(4),* 383-396.

Potter, M. C., & Faulconer, B. A. (1975). Time to understand pictures and words. *Nature,* 253, 437-438.

Potter, M. C., Kroll, J. F., Yachzel, B., Carpenter, E., & Sherman, J. (1986). Pictures in sentences: Understanding without words. *Journal of Experimental Psychology: General,* 115(3), 281-294.

Potter, M. C., So, K.-F., von Eckardt, B., & Feldman, L. B. (1984). Lexical and conceptual representation in beginning and proficient bilinguals. *Journal of Verbal Learning & Verbal Behavior,* 23(1), 23-38.

Pylyshyn, Z. W. (1980). Computation and cognition: Issues in the foundations of cognitive science. *Behavioral and Brain Sciences,* 3, 111-132.

Riddoch, M. J., Humphreys, G. W., Coltheart, M., & Funnell, E. (1988). Semantic systems or system? Neuropsychological evidence re-examined. Special issue: The cognitive neuropsychology of visual and semantic processing of concepts. *Cognitive Neuropsychology, 5(1),* 3-25.

Rosch, E. (1975). Cognitive representations of semantic categories. *Journal of Experimental Psychology: General,* 104(3), 192-233.

Rosch, E., Mervis, C. B., Gray, W. D., Johnson, D. M., & Boyes-Braem, P. (1976). Basic objects in natural categories. *Cognitive Psychology, 8(3),* 382-439.

Shallice, T. (1988). Specialisation within the semantic system. Special issue: The cognitive neuropsychology of visual and semantic processing of concepts. *Cognitive Neuropsychology, 5(1),* 133-142.

Small, S. L., Hart, J., Nguyen, T., & Gordon, B. (1995). Distributed representations of semantic knowledge in the brain. *Brain,* 118(12, Pt. 2), 441-453.

Smith, M. C., & Magee, L. E. (1980). Tracing the time course of picture-word processing. *Journal of Experimental Psychology: General,* 109(4), 373-392.

Snodgrass, J. G. (1984). Concepts and their surface representations. *Journal of Verbal Learning & Verbal Behavior,* 23(1), 3-22.

Snodgrass, J. G., & McCullough, B. (1986). The role of visual similarity in picture categorization. *Journal of Experimental Psychology: Learning, Memory, and Cognition,* 12(1), 147-154.

Snodgrass, J. G., & Vanderwart, M. (1980). A standardized set of 260 pictures: Norms for name agreement, image agreement, familiarity, and visual complexity. *Journal of Experimental Psychology: Human Learning and Memory, 6(2),* 174-215.

Sperber, R. D., McCauley, C., Ragain, R. D., & Weil, C. M. (1979). Semantic priming effects on picture and word processing. *Memory & Cognition, 7(5),* 339-345.

Theios, J., & Amrhein, P. C. (1989). Theoretical analysis of the cognitive processing of lexical and pictorial stimuli: Reading, naming, and visual and conceptual comparisons. *Psychological Review,* 96(1), 5-24.

Van Oostendorp, H., & de Mul, S. (1990). Moses beats Adam: A semantic relatedness effect on a semantic illusion. *Acta Psychologica, 74(1),* 35-46.

Van Petten, C., Kutas, M., Kluender, R., Mitchiner, M., & Mclsaac, H. (1991). Fractionating the word repetition effect with event-related potentials. *Journal of Cognitive Neuroscience, 3(2),* 131-150.

Vanderwart, M. (1984). Priming by pictures in lexical decision. *Journal of Verbal Learning & Verbal Behavior,* 23(1), 67-83.