ELSEVIER

Cognitive Brain Research 16 (2003) 123–144

COGNITIVE BRAIN RESEARCH

www.elsevier.com/locate/cogbrainres

Research report

An electrophysiological study of scene effects on object identification

Giorgio Ganis[a,b,*], Marta Kutas[c]

[a]*Department of Psychology, Harvard University, Cambridge, MA 02138, USA*
[b]*Department of Radiology, Massachusetts General University, Charlestown, MA 02129, USA*
[c]*Department of Cognitive Science and Neurosciences, University of California San Diego, La Jolla, CA 92093-0515, USA*

Accepted 24 September 2002

Abstract

The meaning of a visual scene influences the identification of visual objects embedded in it. We investigated the nature and time course of scene effects on object identification by recording event-related brain potentials (ERPs) and response times (RTs). In three experiments, participants identified objects within a scene that were either semantically congruous (e.g., a pot in a kitchen) or incongruous (e.g., a desk in a river). As expected, RTs were faster for congruous than incongruous objects. The earliest sign of reliable scene congruity effects in the ERPs (greater positivity for congruous pictures between 300 and 500 ms) was around 300 ms. Both the morphology and time course of the N390 scene congruity effect are reminiscent of the N400 sentence congruity effect typically observed in sentence context paradigms, suggesting a functional similarity of the neural processes involved. Overall, these results support theories postulating that visual scenes do not appreciably affect object identification processes before associated semantic information is activated. We speculate that the N390 scene congruity effect reflects the action of visual scene schemata stored in the anterior temporal lobe.

Theme: Neural basis of behavior

Topic: Cognition

Keywords: Object identification; Event-related potentials; Visual scenes; Context effects; N400

1. Introduction

Most empirical work on visual object identification has focused on isolated objects (see for review Refs. [41,47,78]). Yet, in our everyday visual environment, objects are embedded in meaningful visual scenes. In this work, we focus on the processing effects of one specific high-level regularity in visual scenes, namely the long-term co-occurrence of object classes in certain scene contexts, on visual object identification. Specifically, we compare the identification of an object in a setting in which it is often seen (congruous context) versus another in which it is rarely seen (incongruous context) such as a personal computer in an office versus a bathroom scene. In these studies, we do not distinguish between 'associative' and 'semantic' regularities: objects that are semantically related not only tend to be seen in similar scenes but also typically tend to co-occur in the same scene.

Recent studies by Chun and collaborators (see for review Ref. [16]) using visual search paradigms have shown that the cognitive system can indeed acquire incidentally information about the co-occurrence of visual shapes. These studies have demonstrated that seeing a target shape in the context of an array of other shapes facilitates later search of the target shape when it appears in the same context array, relative to a novel one. Furthermore, the results of electrophysiological studies in non-human primates have provided direct evidence that neurons in the anterior temporal cortex can encode such co-occurrence patterns by means of associations between elaborate representations of visual stimuli [55,56]. Outside the laboratory, temporal associations of this type may occur routinely via systematic patterns of eye movements (e.g., Ref. [22]); after all, visual objects that tend to co-occur in a visual scene are likely to be fixated in close temporal proximity.

*Corresponding author. Tel.: +1-617-496-0429; fax: +1-617-496-3112.
E-mail address: ganis@wjh.harvard.edu (G. Ganis).

0926-6410/02/$ – see front matter
doi:10.1016/S0926-6410(02)00244-6

Cognitive psychological accounts have hypothesized the existence of specific memory representations for such co-occurrence patterns. Within a number of psychological accounts, cumulative interactions of the organism with the environment are presumed to lead to the formation of scene-specific schemata, or frames that represent the ''likelihoods, ranges and distributions of things and events'' (Ref. [25], p. 321). Some accounts also include memory representations that are built of a network of dynamic associations among neuron-like nodes and have the advantage of providing a potential link with the neurobiology (e.g., Ref. [64]).

Once activated, scene-specific memory representations are assumed to influence processing of incoming visual information, although both the mode and time course of this influence are debated. Different accounts of object identification in scenes postulate different loci for context effects. Among the processes into which object identification has been decomposed, perceptual processes are believed to analyze the visual input and to transform it into 'structural descriptions' (i.e., high-level visual representations of the shape and structure of visual objects [63]). Subsequent processes are presumed to match the structural descriptions of the object to be identified with those of object models stored in the structural description system [63]. If a good match is found, relevant semantic knowledge is activated. From a computational vision perspective, object identification terminates with a successful match; thus, the processes involved in the activation of semantic knowledge (and beyond) are typically referred to as 'post-identification'. In our view, the activation of semantic knowledge is instead an integral part of object identification because the purpose of object vision is precisely that of ascertaining semantic knowledge about visual stimuli.

The level of processing at which scene information can affect object analysis varies between accounts, among which there are three main subdivisions.

1.1. Early perception accounts

According to these accounts, an activated schema can affect the early perceptual analysis of objects within the scene. Schema activation is thought to occur rapidly on the basis of global, low-resolution contextual information [43,51]. Among the possible candidates for such low-resolution information are scene-emergent features, such as geon clusters typically associated with specific classes of scenes [9] or configurations of oriented 'blobs' [71]. An activated schema is assumed to facilitate the detection of perceptual features (color, texture, size, motion, etc.) that are associated with objects specified within the schema itself [4,25], by means of feature-selective attention. Objects consistent with a scene would thus be identified more quickly on the basis of such features than objects that are inconsistent with a scene.

1.2. Matching process accounts

According to a second class of accounts, scene schemata are assumed to have their effect on the processes involved in matching the structural description with those of object models stored in memory, beyond early perceptual analysis. If we conceptualize this matching process as a selection process wherein the visual system must scan multitudes of representations in the structural description system to find a good match, scene constraints could serve to reduce the size of the search space by priming/biasing classes of the most likely object model representations (e.g., Ref. [78]).

1.3. Post-identification accounts

This third class of accounts relegates scene effects to even later processes, such as during semantic knowledge activation or later (e.g., Refs. [21,33,37,47]). According to these accounts, bottom-up visual analysis is sufficient to discriminate between entry-level object categories, after which context may have its effects.

1.4. Rationale and predictions

It has proven difficult to infer the time course of context effects from behavioral measures alone because they reflect the 'downstream' effects of an experimental manipulation (from the earliest perceptual stages to the motor response). We, therefore, chose to examine the time course of scene effects on object identification via a brain measure with greater temporal precision, namely, recordings of event-related brain potentials (ERPs). Scalp ERPs enable continuous monitoring of the modulations of synchronous neural activity elicited by experimental manipulations in a relatively direct manner. As ERPs have provided crucial information regarding the time course of semantic context effects in language processing [42,80], we aimed to use a similar approach to investigate the timing of context effects on nonlinguistic, visual processing.

To assess the time course of scene effects on object identification, we compared ERPs elicited by objects appearing in congruous versus incongruous scene contexts. In such an analysis, the timing of ERP differences between these two conditions provides an estimate of the time by when neural representations of scene information must have begun to interact with identification processes for the target object. Furthermore, the spatial distribution of these ERP congruity effects across the head can provide some clues about the nature of the processes involved. To our knowledge, this is the first ERP study that directly addresses scene effects on object identification.

Experiment 1 is a behavioral study to demonstrate the scene congruity effect with our stimulus set. Experiment 2 is an ERP study to determine the time course and spatial signature of the scene congruity effect. Experiment 3 is a

replication of Experiment 2 with a modified paradigm not requiring explicit congruity judgments in order to allow a more direct comparison of the congruity effects obtained using scenes with those found in earlier studies with sentential contexts.

Note that throughout the paper we will use the traditional electrophysiological nomenclature to refer to most ERP components (i.e., voltage deflections): the first letter indicates whether the component is negative ('N') or positive ('P') relative to the chosen reference, whereas the subsequent number(s) indicates either the average latency of the component in ms (e.g., N200) or the ordinal position of the component (e.g., 'P1' refers to the first positive ERP component in the visual ERP).

Based on numerous behavioral studies that have reported scene effects on object identification (e.g., Refs. [10,14]), we expected to obtain a reaction time advantage for congruous relative to incongruous objects. The timing of ERP congruity effects is thus the main focus of the present study. We also used ERP congruity effects to evaluate the three main accounts of scene effects on object cognition. On the one hand, we reasoned that if scene schemata affect the early perceptual analysis of objects, then ERP components indexing early perceptual processes should be modulated by congruity. Short-latency ERPs (onsetting during the first 200 ms poststimulus) are believed to index early perceptual processes because (a) they are modulated by variations in the perceptual attributes of a stimulus, such as spatial location (e.g., Ref. [17]), luminance [46], color [15,46], spatial frequency [40] coherent motion [59], and shape [68,77], and (b) they are modulated by task manipulations that affect the perceptual encoding of a stimulus: the P1 and the N1 components (which onset between 80 and 130 ms poststimulus) are modulated by spatial attention (e.g., Ref. [32]); furthermore, the selection negativity (SN, which onsets between 140 and 180 ms poststimulus) is elicited by selective attention to nonspatial features of a visual stimulus such as color [3], shape [72], direction of motion [2], spatial frequency [31], and orientation [39,62]. The time course and estimated neuroanatomical location of these ERPs (early extrastriate cortex) suggest that they occur prior to stimulus identification.

On the other hand, if scene schemata act only later to affect the activation of semantic knowledge, then congruity should modulate only later components, such as the N400 [42]. N400 amplitude is typically reduced when the eliciting stimulus is preceded by an associatively or semantically related one (e.g., Refs. [5–8,12,36,50,69,70]). Indeed, one view of the N400 is that it reflects neural processes involved in the activation of semantic knowledge [61,80] probably stored in the anterior temporal lobes.

Predictions for the ERP correlates of scene schemata effects on structural description matching processes can be estimated from prior ERP investigations of the time course of object identification. A class of negativities ('N300/N350', 'Ncl') peaking around 350 ms with a frontal scalp distribution (with mastoid reference site) has been proposed to reflect activation of structural description matching processes [24,26,36,50,67]; we refer to these as the structural description negativity, Nsd. By 200 to 300 [230/250] ms, the Nsd is greater for unidentified real objects and non-objects (i.e., images of objects that do not exist) than for identified or real objects. These effects can last for several hundred milliseconds, consistent with the idea that the Nsd reflects repeated but ultimately failed attempts to find a good match for unidentified objects and non-objects; in contrast, Nsd is rapidly reduced within 300 ms when structural description matching succeeds with identified real objects. Thus, if scene schemata affect object identification at the level of structural description matching processes, we might expect to find an ERP congruity effect that onsets around 200 ms with an anterior scalp distribution similar to Nsd effects.

2. Materials and methods

2.1. Participants

All 42 participants were UCSD students and native speakers of English. They received course credit, or were paid $5.00/h for participating. Ten participants took part in Experiment 1 (four men, six women between 18 and 25 years of age, mean 20; nine right-handed). Seventeen participants, different from those employed in Experiment 1, took part in Experiment 2 (nine men, eight women between 18 and 25 years of age, mean 20.3; 14 right-handed); data from two participants were discarded due to excessive movement, and a third (left-handed) could not perform the task as the stimulus rate was 'too fast'. Fifteen participants, different from those in Experiments 1 and 2, took part in Experiment 3 (seven males, eight females, 18 to 24 years of age, mean 20.7). The data from one participant was discarded because the recordings were contaminated with large eye movements and muscle artifacts. All analyses were performed on the remaining participants. All the experiments were undertaken with the understanding and written consent of each participant.

2.2. Materials and apparatus

Ninety-six color photographs of common indoor and outdoor environments were used. These were selected such that (a) they contained at least one predictable target object, (b) the target object was clearly visible, and (c) the target object had a companion object of similar size in a control set of objects, that would be incongruent within the same context. For each scene, a congruous and an incongruous version were created, for a total of 192 stimuli. These stimuli were divided into two sets such that congruous scenes in the first set were incongruous in the second,

and vice versa, thereby counterbalancing the scenes across participants.

Each scene subtended an area of 150 000 square pixels (85 square degrees). The average luminance and contrast of the scenes in the congruous and incongruous sets were matched. The distribution of target object locations was roughly uniform across the four quadrants. The average size and luminance of the target objects in the two conditions was matched (about 2.5×2.5° of visual angle and average pixel value equal to 170, after accounting for relative RGB weighting).

For each of the 96 intact scenes, a scrambled scene was created using the algorithm employed in [44], adapted to color images. These scrambled scenes have the same amplitude spectrum as the original images but are unrecognizable. Target objects, different from the ones used in the intact scenes, were placed in half of the scrambled scenes at the same fixation location as in the original scenes (scrambled scene condition). The remaining scrambled scenes (fillers) were presented without any object in them. Thus, each participant saw 48 stimuli in each of the four conditions (congruous, incongruous, scrambled, and fillers).

The stimulus presentation apparatus consisted of an IBM-compatible 486 microcomputer with a 20-in color monitor (120 Hz refresh rate) centrally positioned at eye level, 150 cm from the observer's eyes. The experiment was driven by custom-made software. RTs were sampled at 250 Hz.

2.3. Procedure

2.3.1. Experiment 1

Each participant was tested in one experimental session lasting about 90 min. Participants sat on a comfortable chair in front of the computer screen. At the beginning of each trial, a small cross (8×8 pixels in size) appeared on the screen for 800±125 ms. Participants were instructed to fixate the cross, and were told that the target object would appear at the fixation location. The location of the fixation cross and the target object varied from trial to trial. The fixation cross was followed immediately by a scene. Three hundred ms into the presentation of the scene, the target object appeared at the cued location and was shown together with the scene for 300 ms. Subjectively, the target object seemed to 'pop' into the scene. The screen was then cleared and the cross reappeared at the cued location to help the participants maintain fixation.

Participants were instructed to press a hand-held button (dominant hand) as soon as they identified the target object. About 2 s after the screen cleared, participants were further asked to indicate (a) their confidence in their identification of the target object, (b) their confidence in their identification of the scene context, and (c) the degree of congruity between the target object and the scene. These responses were made with the nondominant hand using a custom-made, five-key, response board. Finally, participants were asked to name the target object, or to say ''don't know'' if they could not identify it (Fig. 1). Participants experienced a practice run that introduced them to the stimuli and nature of the responses required.

2.3.2. Experiment 2

The only procedural change from Experiment 1 was for the ERP recording. The experimental session lasted about 2 h (55 min for EEG set-up). Before the practice session, participants were instructed on the importance of reducing the number of eye movements and blinks, and the amount of muscle tension. At the end of the experiment, horizontal eye movement distances were calibrated during a session lasting about 5 min. A fixation cross was flashed at various eccentricities along the horizontal axis, and participants were asked to follow its movement with their eyes.

2.3.3. Experiment 3

We used the same procedures from Experiment 2, except for the task requirements, which were limited to target object identification. To verify that object identification had occurred, participants were asked to name the object at the end of each trial, approximately 3000 ms after target onset.

2.4. Electrophysiology

The electroencephalogram (EEG) was sampled at 250 Hz. The EEG was recorded from tin electrodes at 26 locations (sites) arranged in a geodesic array, with mean interelectrode distance of approximately 4 cm (Fig. 2). All electrode recordings, including one from over the right mastoid process, were referred to the left mastoid. Data were re-referenced off-line to the average of the activity at left and right mastoids. Blinks and vertical eye movements were monitored via an electrode placed on the lower orbital ridge, also referred to the left mastoid. Lateral eye movements were monitored by a bipolar pair of electrodes placed lateral to each eye (right versus left canthi). Electrode impedance was below 2 kΩ for the scalp channels, and below 10 kΩ for the eye channels. The EEG was amplified by Grass amplifiers with a 0.01–100-Hz half-amplitude cutoff bandpass. ERPs were averaged off-line for an epoch of 2048 ms. Most plots show a stretch encompassing 1700 ms (−450 to 1250 ms) of this epoch.

2.5. Data analyses

2.5.1. Behavior

ANOVAs were conducted on the RTs, identification rates, and identification ratings. RT analyses were based on median RTs to avoid problems associated with outlier removal [52]. Analyses included only those trials on which (a) both the object and the scene were identified correctly

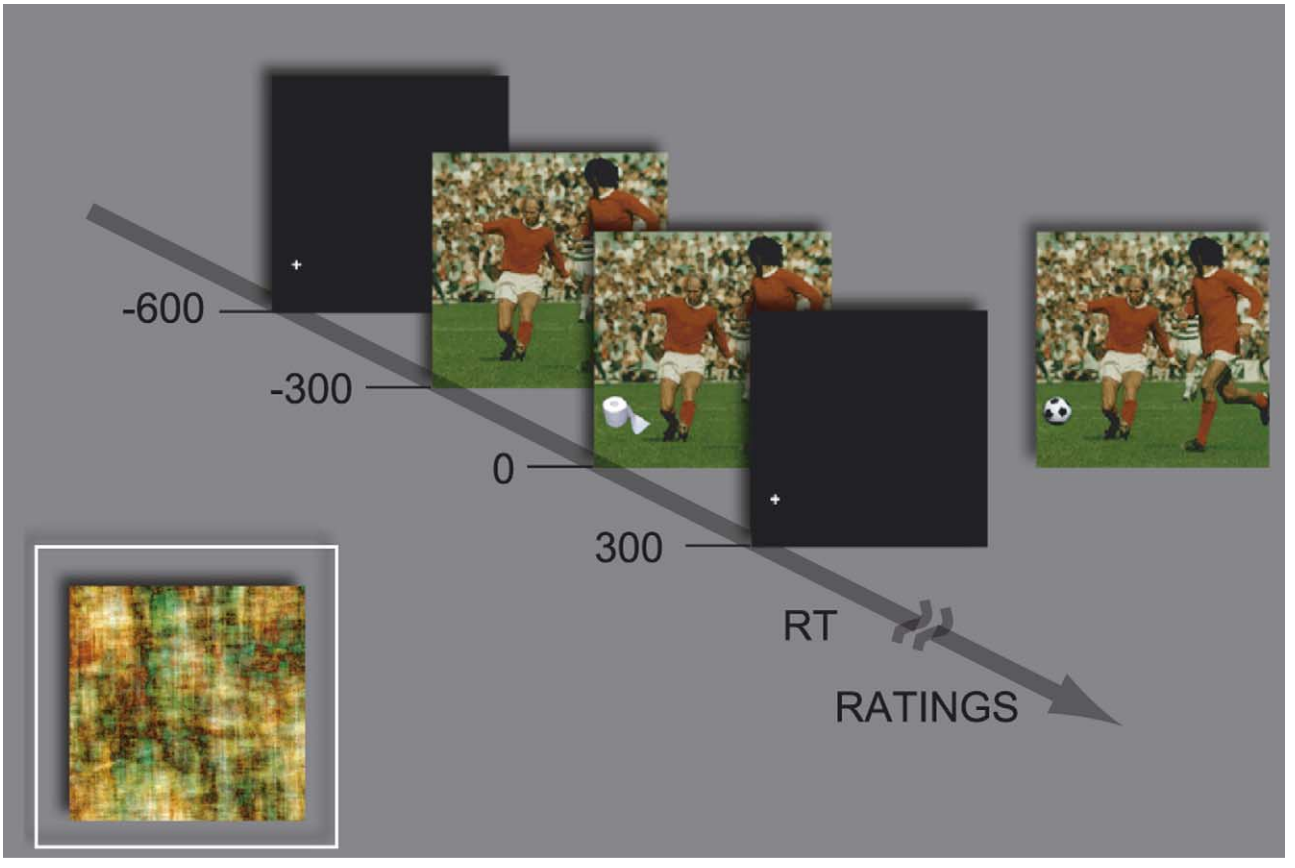

Fig. 1. Schematic diagram of the stimulus sequences used in Experiments 1 and 2. Time goes from left to right. The figure shows an example of a congruous stimulus (right), an incongruous stimulus (center), and a scrambled stimulus (bottom left). The task required an immediate response. At the end of each trial participants provided ratings of object confidence, scene confidence, and congruity; finally, they named the target object. This stimulus sequence was also used in Experiment 3, with the only difference that no ratings were required at the end of each trial.

with high confidence (‘4’), and (b) the congruity rating was either very high or very low (‘4’ or ‘0’).

2.5.2. ERPs

Trials contaminated by eye movements or amplifier blocking were rejected off-line. These included trials with horizontal eye movements exceeding about 1.5° (60 pixels) during the 600 ms when the scene was on the screen, as determined by the calibration procedure. Of the remaining trials, only those wherein (a) both the object and the scene were identified correctly with high confidence (‘4’), and (b) the congruity rating was very high or very low (‘4’ or ‘0’) were included in the analyses of Experiment 2. In Experiment 3, only correct object identification trials were included in the analyses (since no additional ratings were required by the task).

ERP analyses were conducted with ANOVA performed on the mean amplitude of the average ERPs within specified time-windows. The baseline was the 150-ms window before the onset of each scene. Greenhouse–Geisser adjustments to degrees of freedom were applied to correct for violation of the assumption of sphericity, $\alpha = 0.05$. Congruity effects were assessed using an ANOVA with two within-participant factors: congruity (congruous versus incongruous) and recording site (26 locations). Hemispheric asymmetries were assessed with an ANOVA on the 11 pairs of sites located symmetrically relative to the mid-sagittal plane with three within-participant factors (congruity, hemisphere, and site pair). Time windows that revealed a reliable effect of congruity where subjected to additional ANOVAs for lateralized effects.

Additional ANOVAs on subsets of sites were performed to test for the presence of congruity effects on specific ERP components, as detailed in the Results section of each experiment.

ANOVAs comparing the scalp distribution of scene congruity effects with those of the sentence context experiment reported in Ganis et al. [27] were conducted in Experiment 3 using the method described by McCarthy and Wood [49]. In these analyses, there was a between-group factor (study), and one within-group factor (recording site). To address the possibility that some effects might not be detected with our average mastoid reference, all analyses that included the congruity factor were performed also on the data referenced off-line to the vertex electrode. Although rereferencing the data to the vertex changed the

Fig. 2. Schematic of the electrode montage used for the ERP recordings. The montage is composed of 26 electrodes arranged in a quasi-geodesic configuration.

shape of the ERPs, the pattern of the effects was essentially identical to that obtained with the average mastoid reference; therefore, these analyses will not be discussed further.

Finally, to address the possibility that early effects might not be detected with our montage we used ANOVA (a) to assess the effect of hemifield of scene presentation on the P1 peak (mean from 100 to 125 ms after scene onset for congruous and incongruous trials collapsed) at two lateral occipital sites (i.e., sites 5 and 7); trials were divided into two sets, depending on whether the fixation cross was on the left (scene mostly in the right hemifield) or on the right (scene mostly in the left hemifield) relative to the center of the scene, and (b) to assess early processing differences between intact and scrambled scenes; this analysis was conducted on the ERPs in consecutive 25-ms windows from −250 to 0 ms before target onset. The two factors used were scrambling (intact versus scrambled scenes) and recording site (26 locations).

3. Results

3.1. Behavioral data

In Experiment 1 (Table 1) participants correctly identified 93% of the target objects. Congruity affected neither the object identification rates ($F(2,18)=1.14$, $P>0.1$) nor the object confidence ratings ($F(2,18)=0.02$, $P>0.1$). In contrast, mean confidence ratings for the scenes were higher in the congruous (3.93) than incongruous (3.86) condition ($F(1,9)=5.62$, $P<0.05$). Finally, congruous stimuli were rated significantly higher in congruity than incongruous ones ($F(1,9)=3156.84$, $P<0.00001$). Median RTs for incongruous targets (637 ms) were slower than those for congruous targets (599 ms, $F(1,9)=6.83$, $P<0.05$). RTs for objects identified in isolation (606 ms) were slightly slower than RTs for congruous objects and slightly faster than those for incongruous ones, although they were not significantly different from either ($F(1,9)<1.7$, $P>0.1$).

In Experiment 2 (Table 1), participants correctly identified 90% of the target objects. Congruity affected neither identification rates ($F(1,13)=0.3$, $P>0.1$) nor object confidence ratings ($F(1,13)=1.32$, $P>0.1$). In contrast, confidence ratings for congruous scene contexts were higher than those for the incongruous ones ($F(1,13)=9.39$, $P<0.01$). Finally, congruity ratings for congruous stimuli were

Table 1
Summary of behavioral results

	Object ID rate (%)	Object confidence	Scene confidence	Congruity	RTs (ms)
Experiment 1:					
Congruous	93.7 (0.9)	3.87 (0.03)	3.93 (0.04)	3.67 (0.07)	599 (39)
Incongruous	93.8 (0.8)	3.87 (0.03)	3.86 (0.06)	0.12 (0.03)	637 (44)
Scrambled	91.2 (0.7)	3.86 (0.03)	NA	NA	606 (38)
Experiment 2:					
Congruous	90.5 (0.7)	3.88 (0.03)	3.85 (0.04)	3.74 (0.05)	687 (33)
Incongruous	90.1 (0.8)	3.87 (0.03)	3.70 (0.06)	0.27 (0.04)	724 (31)
Scrambled	89.7 (0.5)	3.87 (0.03)	NA	NA	671 (27)
Experiment 3:					
Congruous	87.4 (.7)	NA	NA	NA	NA
Incongruous	87.8 (.8)	NA	NA	NA	NA
Scrambled	87.1 (.7)	NA	NA	NA	NA

The standard error of the mean is indicated in parenthesis.

higher than those for incongruous ones $F(1,13)=2512.54$, $P<0.00001$. RTs for incongruous targets (724 ms) were slower than those for congruous ones (687 ms, $F(1,13)=6.55$ $P<0.05$). Reaction times for the identification of objects in scrambled scenes (671 ms) were only numerically faster than for congruous ($F(1,13)=1.95$, $P>0.1$) but reliably faster than for incongruous objects ($F(1,13)=31.64$, $P<0.001$).

In Experiment 3 (Table 1) participants correctly identified 87% of the target objects. Congruity did not affect the identification rates ($F(1,13)=0.2$, $P>0.1$). Note that no other behavioral measures were obtained in this experiment.

3.2. Electrophysiological data

Electrophysiological data were obtained for Experiments 2 and 3.

3.2.1. Experiment 2

Overall, there was no difference in the number of trials rejected due to eye movements across conditions (approximately 15%). Almost all the trials that were discarded based on the ratings also contained eye movement artifacts. On average, ERP analyses were based on 40 trials (out of 48) in each condition per participant.

3.2.1.1. Overview of ERPs. Scenes and target objects elicited stereotypical early visual evoked potentials (Fig. 3). Note that scene onset time is −300 ms because the time scale is relative to the onset of the target object. The first potential was a C1 component, peaking at −225 ms, and maximal at occipito-parietal sites. Subsequent visual evoked potentials elicited by scenes comprised a P1–N1–P2 sequence, with an average peak latency at −194, −154, and −94 ms, respectively; these were maximal at occipital, occipito-parietal, and occipital sites, respectively. Target object presentation elicited a similar, albeit less visible, P1–N1–P2 sequence, overlapping the later potentials elicited by the prior scene. The P1, N1, and P2 peaked at 110, 162, and 258 ms, respectively. Stimulus offset potentials were also visible at occipital sites, starting at 400 ms (100 ms after stimulus offset): a P1 peak at 400 ms, an N1 peak at 450 ms, and a P2 peak at 474 ms.

Figs. 4 and 5A show the ERPs elicited by congruous and incongruous objects during the entire epoch (−450 to 1250 ms). Following the P2, scenes elicited a broad negativity peaking at −50 ms (N250), maximal at centro-frontal sites. Later components, presumably a combination of neural activity elicited by the scene and the target object, included a centro-frontal negativity peaking at 142 ms, followed by another broad negativity peaking at 390 ms (N390). This latter component had a centro-parietal maximum and was larger for incongruous than congruous target-scene combinations. The N390 was followed by broad, slow positive potentials between 600 and 1200 ms, which also appeared to be modulated by congruity.

3.2.1.2. ERP congruity effects. One of the anticipated outcomes of this experiment was an enhanced negativity in the 'N400' time window for incongruous relative to congruous target objects (see Fig. 5B for difference ERPs). An ANOVA conducted on the lateral sites in the 300–500-ms time window showed this to be the case (5.49 vs. 6.98 μV; main effect of congruity $F(1,13)=10.94$, $P<0.01$). The size of the congruity effect varied with site (congruity by site interaction, $F(1,13)=3.40$, $P<0.05$), but *not* hemisphere (congruity by hemisphere interaction, $F(1,13)=0.45$, $P>0.1$). Overall, the ERPs were more positive over the left than the right hemisphere (main effect of hemisphere $F(1,13)=9.14$, $P<0.01$). The ANOVA of the midline sites, likewise, showed that between 300 and 500 ms the ERPs were more negative for incongruous than congruous objects (7.47 vs. 5.91 μV; main effect of congruity $F(1,13)=10.94$, $P<0.01$). The size of the congruity effect was largest at central sites and smallest over frontal sites (congruity by site interaction, $F(1,13)=5.31$, $P<0.05$).

A more detailed analysis of the scalp distribution and hemispheric asymmetries was performed with ANOVAs on each pair of 11 laterally symmetric sites. The congruity effect was not reliable at frontal (2–10) pair ($F(1,13)=5.09$, $P>0.05$; $F(1,13)=0.33$, $P>0.1$) and occipital (5–7) pair ($F(1,13)=5.0$, $P>0.05$; $F(1,13)=0.44$, $P>0.1$), but was reliable at all other pairs. The only sign of hemispheric asymmetry was at temporal pair 4–8 (main effect of congruity: $F(1,13)=10.91$, $P<0.01$; congruity by hemisphere interaction: $F(1,13)=5.08$, $P<0.05$, not corrected for multiple comparisons), where the congruity effect was larger over the right than the left (2.8 vs. 1.8 μV).

The detailed time course of congruity effects in the N400 time window was assessed within adjacent 25-ms time segments, beginning at 300 ms and ending at 500 ms. There was a main effect of congruity within every time window (Table 2). Follow-up analyses revealed no reliable hemispheric asymmetries at any time (all F values <1.50; all P values >0.2). Since this congruity effect overlaps with a large negative-going ERP component peaking at 390 ms (Figs. 4 and 5B), it will be referred to as the N390 congruity effect.

Early congruity effects first were assessed with ANOVAs covering adjacent 25-ms time segments, beginning at 50 ms and ending at 300 ms. A lower limit of 50 ms was chosen because this is the earliest known time when neural signals elicited by visual stimuli can reach human striate cortex [73]. Neither main effects of congruity nor interactions of congruity with site were found before 300 ms (Table 2). Early congruity effects were also assessed with ANOVAs performed on the amplitude of specific ERP components. The analysis of the P1 (between 100 and 125 ms) and the P2 (between 225 and 275 ms)

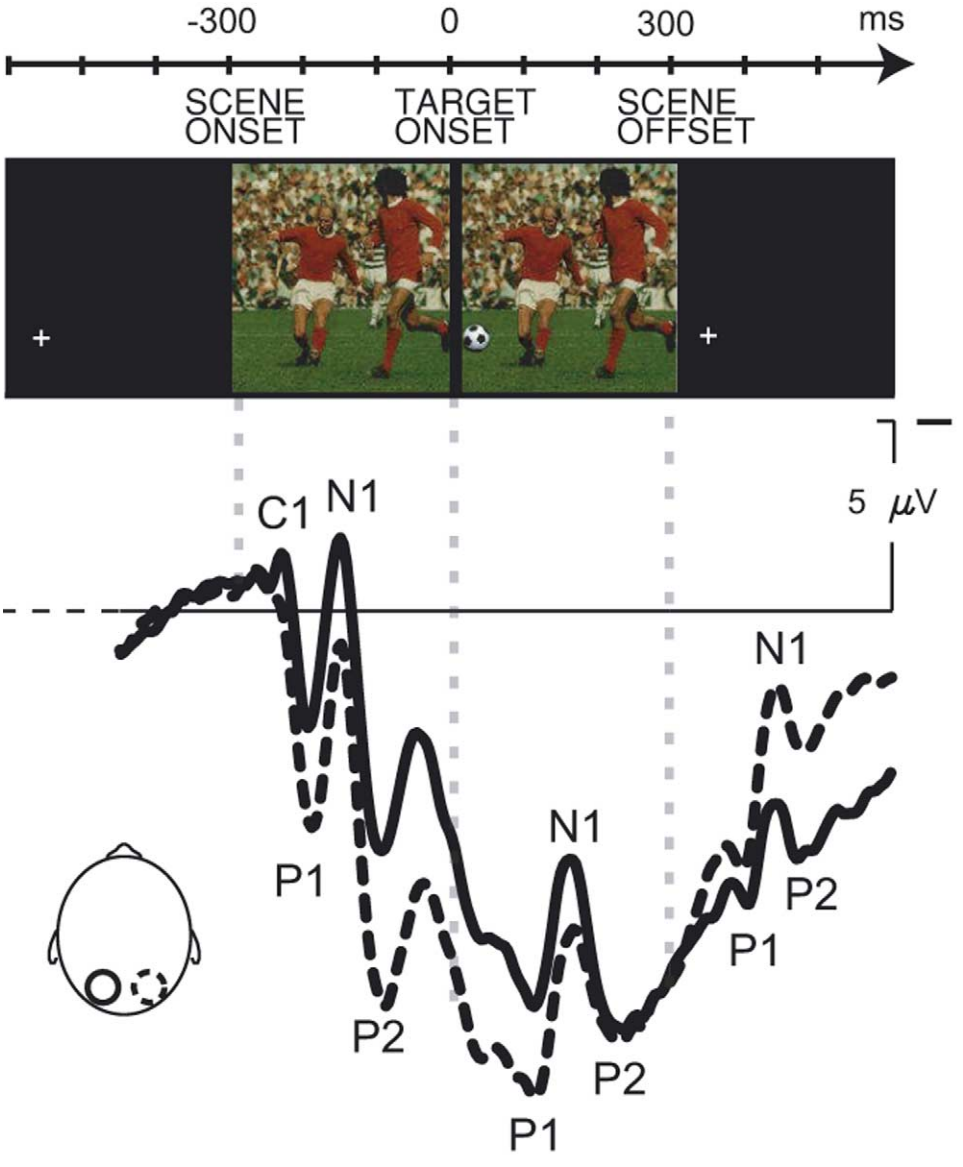

Fig. 3. Early visual evoked potentials elicited by the stimulus sequence at left (solid line) and right (dashed line) occipital electrodes 5 and 7 (Experiment 2). The origin of the time scale is at the onset of the target object (the soccer ball in this example). In this scale the scene context onsets at −300 ms. By convention, negative is plotted upward. All waveforms were digitally low-pass filtered (half-amplitude cutoff was approximately 30 Hz).

were performed on site pair 5–7. The analysis of the N1 (between 150 and 200 ms) was performed across all sites because the N1 complex has a broad scalp distribution [17]. None of these tests revealed an effect of congruity (all P values >0.1). Congruity effects in the time window corresponding to the early phase of the typical posterior SN, between 150 and 250 ms [3], were performed over the five most posterior sites. Since the SN is often accompanied by a frontal selection positivity, [3] a similar analysis was performed over the five most anterior sites. Finally, an analysis on the five most anterior sites between 200 and 275 ms was performed to assess modulations of the Nsd. None of the analyses revealed an effect of congruity (all P values >0.2).

To assess the offset of the N390 congruity effect, ANOVAs were conducted on the mean ERPs in 25 ms windows from 500 to 600 ms. Neither main effects nor interactions involving the congruity factor were significant in any of the time windows. In summary, the N390 congruity effect begins around 300 ms and ends around 500-ms post-target onset.

Congruity effects after the N400 window were assessed within three adjacent 250-ms time segments, beginning at 500 ms and ending at 1250 ms. No main effects were found in any of the time windows (Table 2). The absence of main effects reflects the finding that congruous stimuli are more positive than incongruous ones at anterior sites, while the opposite occurs at posterior sites. Consistent with this, condition by site interactions were significant between 750 and 1250 ms. Follow-up ANOVAs revealed no

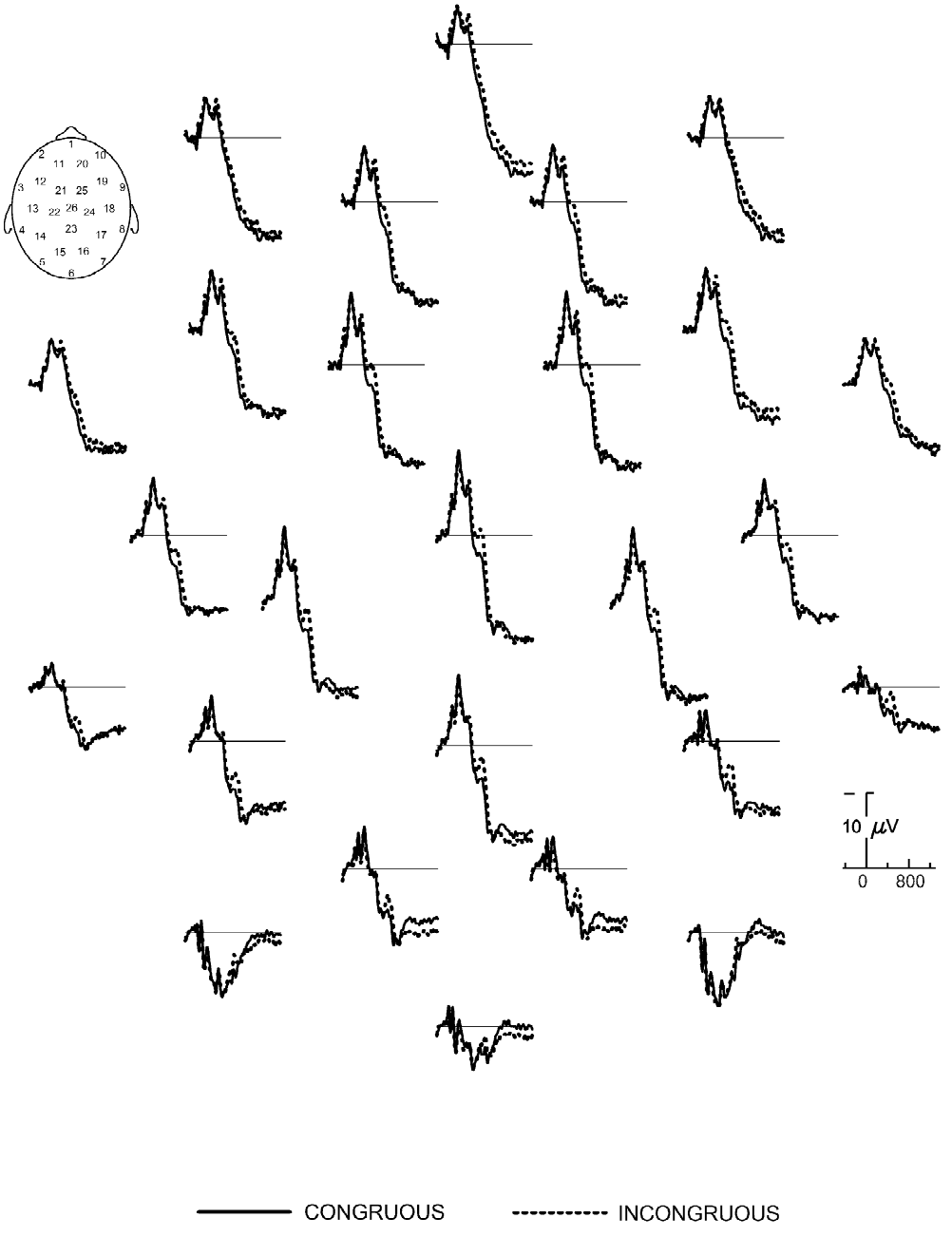

Fig. 4. Overview of ERPs elicited by the congruous (solid line) and incongruous (dotted line) sequences (Experiment 2).

congruity by hemisphere interactions (all F values <1; all P values >0.2).

3.2.1.3. Early ERPs to intact and scrambled scenes. Since scenes were shown for 300 ms before the target object appeared, it was possible to investigate the early time course of scene processing by comparing the ERPs to intact and scrambled scenes *before* the visual evoked potentials elicited by the target object (Fig. 6A,B). The first reliable difference between intact and scrambled scenes was between 100 and 125 ms after scene onset (Table 3). An ANOVA conducted on the P1 at occipital site pair 5–7 showed that intact scenes elicited a larger P1 than scrambled ones ($F(1,13)=15.13$, $P<0.001$), with the difference greater on the right than on the left (1.5 vs. 0.7 μV: $F(1,13)=7.82$, $P<0.01$). A larger effect of scene scrambling was observed on the N250, from 200 to 300 ms after scene onset (Table 3). The N250 was larger for intact than scrambled scenes, as for the P1. This effect had a centro-frontal distribution but was clearly visible at occipital sites as well.

3.2.1.4. Effects of hemifield of presentation on early visual evoked potentials. In approximately half of the scenes the

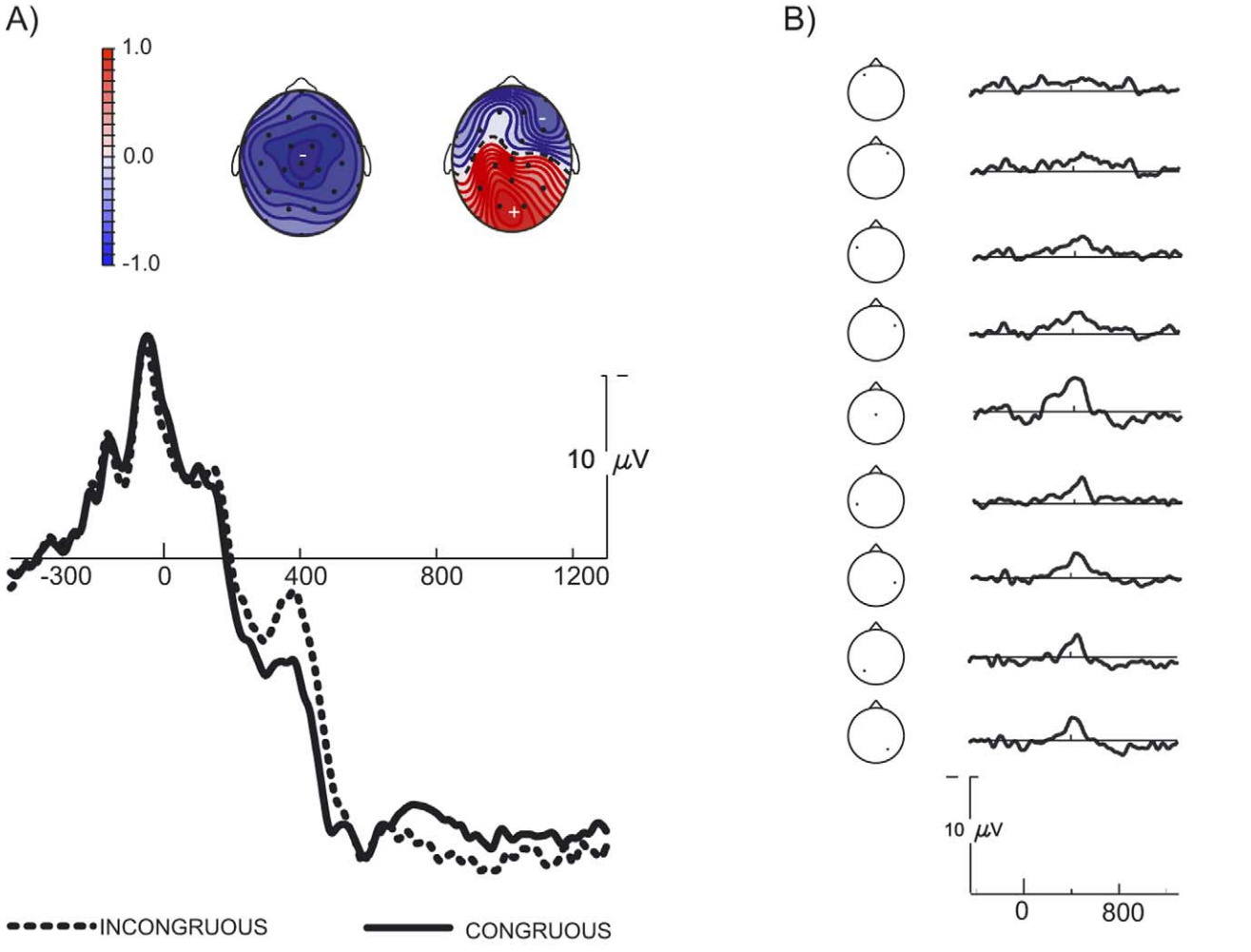

Fig. 5. (A) ERPs elicited by the congruous (solid line) and incongruous (dotted line) stimulus sequences at the midline parietal electrode (23) where both congruity effects are clearly visible (Experiment 2). The normalized scalp distribution of the congruity effects is illustrated by the isopotential maps (see for details Ref. [27]). The dashed line is the zero value isocontour. Positive and negative polarities are indicated with white '+' and '−' signs, respectively. The N390 congruity effect (incongruous minus congruous) is shown on the left (mean between 300 and 500 ms); the congruity effect on the late positivity is shown on the right (mean between 750 and 1250 ms). (B) Difference ERPs (incongruous minus congruous ERPs) at a subset of electrodes.

fixation was on the left, whereas in the other half it was on the right (relative to the center of the image). The average fixation offset was about 2.5° for both left and right trials. When the fixation was on the right (left), most of the scene was flashed onto the left (right) hemifield and was delivered to right (left) striate cortex due to the neuroanatomy of the human visual system [18]. Thus, we expected an effect on early potentials as a function of hemifield. Here, we focus on the occipital P1. The data were collapsed across congruity conditions so that the number of trials in the left and right conditions was comparable to that in the congruous and incongruous conditions in the main analyses (40 trials on average). Consistent with previous studies [18], the occipital P1 was modulated by presentation hemifield (Fig. 7A,B). At the right occipital site, the P1 was larger in the left hemifield condition, while at the left occipital site it was larger in the right hemifield condition (presentation hemifield by site, $F(1,13)=18.14$, $P<0.001$). An ANOVA on the P1 peak revealed no main effect of hemifield, $F(1,13)=0.32$, $P>0.1$.

3.2.2. Experiment 3

Overall, there was no difference in the number of trials rejected due to eye movements across conditions (approximately 12%). Only trials in which correct identification occurred were included in the ERP analyses, an average of 41 trials (out of 48) in each condition per participant.

3.2.2.1. Overview of ERPs. Scenes and target objects elicited stereotypical early visual evoked potentials similar to those observed in Experiment 2. Figs. 8 and 9a show the ERPs elicited by congruous and incongruous stimuli during the entire epoch (−450 to 1250 ms). The sequence of potentials was very similar to that observed in Experi-

Table 2
Summary of the ANOVAs conducted on the mean ERPs in the congruous and incongruous conditions (all 26 sites)

Time window (ms)	Experiment 2				Experiment 3			
	Congr		Congr×Site		Congr		Congr×Site	
	F	*P*	*F*	*P*	*F*	*P*	*F*	*P*
50–75	0.52	0.48	0.46	0.86	0.01	0.90	1.09	0.37
75–100	0.02	0.90	0.30	0.91	0.32	0.58	1.69	0.15
100–125	0.19	0.67	0.33	0.85	0.17	0.68	1.12	0.31
125–150	2.01	0.18	0.69	0.56	0.02	0.90	1.46	0.21
150–175	2.65	0.13	1.47	0.21	0.51	0.49	1.34	0.27
175–200	2.77	0.12	1.09	0.37	1.10	0.31	1.56	0.20
200–225	1.28	0.28	1.38	0.23	0.70	0.42	1.95	0.11
225–250	0.61	0.45	0.99	0.44	2.30	0.15	1.68	0.16
250–275	0.66	0.43	1.04	0.40	3.4	0.10	2.06	0.12
275–300	3.87	0.07	2.14	0.09	1.83	0.20	0.57	0.57
300–325	7.17	0.019*	3.07	0.01*	2.64	0.13	0.81	0.49
325–350	8.71	0.013*	3.56	0.007*	5.35	0.03*	2.66	0.02*
350–375	8.31	0.013*	4.03	0.003*	9.64	0.01*	2.71	0.02*
375–400	8.79	0.011*	3.49	0.01*	10.12	0.01*	1.70	0.14
400–425	9.17	0.009*	2.72	0.038*	8.45	0.01*	2.90	0.01*
425–450	10.30	0.007*	2.34	0.05*	6.52	0.02*	1.87	0.10
450–475	9.99	0.007*	1.54	0.19	3.06	0.10	1.90	0.10
475–500	8.14	0.013*	0.62	0.65	2.90	0.11	1.77	0.11
500–750	0.57	0.46	1.74	0.18	0.02	0.89	3.63	0.02*
750–1000	0.16	0.69	3.52	0.004*	5.97	0.03*	4.61	0.003*
1000–1250	0.01	0.97	2.29	0.04*	3.08	0.10	4.96	0.002*

The congruity factor is indicated as 'Congr'. Degrees of freedom are 1 and 13 for the main effects and 25 and 325 for the interactions.

ment 2. In particular, there was an N390, larger for incongruous than congruous stimuli, and a later slow positive potential, also modulated by congruity.

3.2.2.2. ERP congruity effects. ANOVAs on the lateral sites in the N400 time window revealed that the ERPs were more positive for congruous than incongruous stimuli (5.16 vs. 3.87 μV; main effect of congruity $F(1,13)=6.89$, $P<0.05$), though the congruity effect was smaller that in the previous experiment (1.29 vs. 2.5 μV). The size of the congruity effect varied with site pair ($F(1,13)=6.87$, $P<0.01$). ANOVAs on the midline sites revealed only a trend for the ERPs to be more positive for congruous than incongruous stimuli (6.1 vs. 5.1 μV; main effect of congruity: $F(1,13)=3.70$, $P=0.07$; interaction congruity by site: $F(3,39)=2.88$, $P=0.07$). Additional ANOVAs performed on each pair of laterally symmetric sites showed that the congruity effect was not reliable at frontal pair 2–10 ($F(1,13)=0.76$, $P>0.1$; $F(1,13)=0.38$, $P>0.1$) and at occipital pair 5–7 ($F(1,13)=3.93$, $P>0.05$; $F(1,13)=0.62$, $P>0.1$). There were no hemispheric asymmetries (see Fig. 9b for difference waves).

ANOVAs on the mean ERPs in 25 ms windows between 300 and 500 ms showed that there was a main effect of congruity between 325 and 450 ms (Table 2) and no hemispheric asymmetries in any time window (all *F* values <0.31; all *P* values >0.5). There were neither main effects of congruity nor reliable interactions of congruity with site prior to 325 ms (Table 2).

As in Experiment 2, early congruity effects first were assessed with ANOVAs covering adjacent 25-ms time segments, beginning at 50 ms and ending at 300 ms. Neither main effects of congruity nor interactions of congruity with site were found before 300 ms (Table 2). Early congruity effects were also assessed with ANOVAs performed on the amplitude of the same ERP components analyzed in Experiment 2. No reliable congruity effects were found in any of these comparisons (all *P* values >0.2).

Congruity effects after the N400 window were assessed within three adjacent 250-ms time segments, between 500 and 1250 ms. There was a main effect of congruity between 750 and 1000 ms and significant interactions with site throughout (Table 2). Follow-up ANOVAs showed that this late congruity effect tended to be larger over the left hemisphere sites between 1000 and 1250 ms (congruity by hemisphere interaction: $F(1,13)=5.61$, $P=0.05$).

3.2.2.3. Early ERPs to intact and scrambled scenes. The first reliable difference between intact and scrambled scenes occurred on the occipital P1, between 75 and 125 ms after scene onset (Fig. 10 and Table 3). An ANOVA conducted on the P1 (mean amplitude between 100 and 125 ms) at occipital site pair 5–7 revealed that intact scenes elicited a larger P1 than scrambled ones ($F(1,13)=21.35$, $P<0.001$); the P1 was larger over the right than left hemisphere (main effect of hemisphere: $F(1,13)=8.82$,

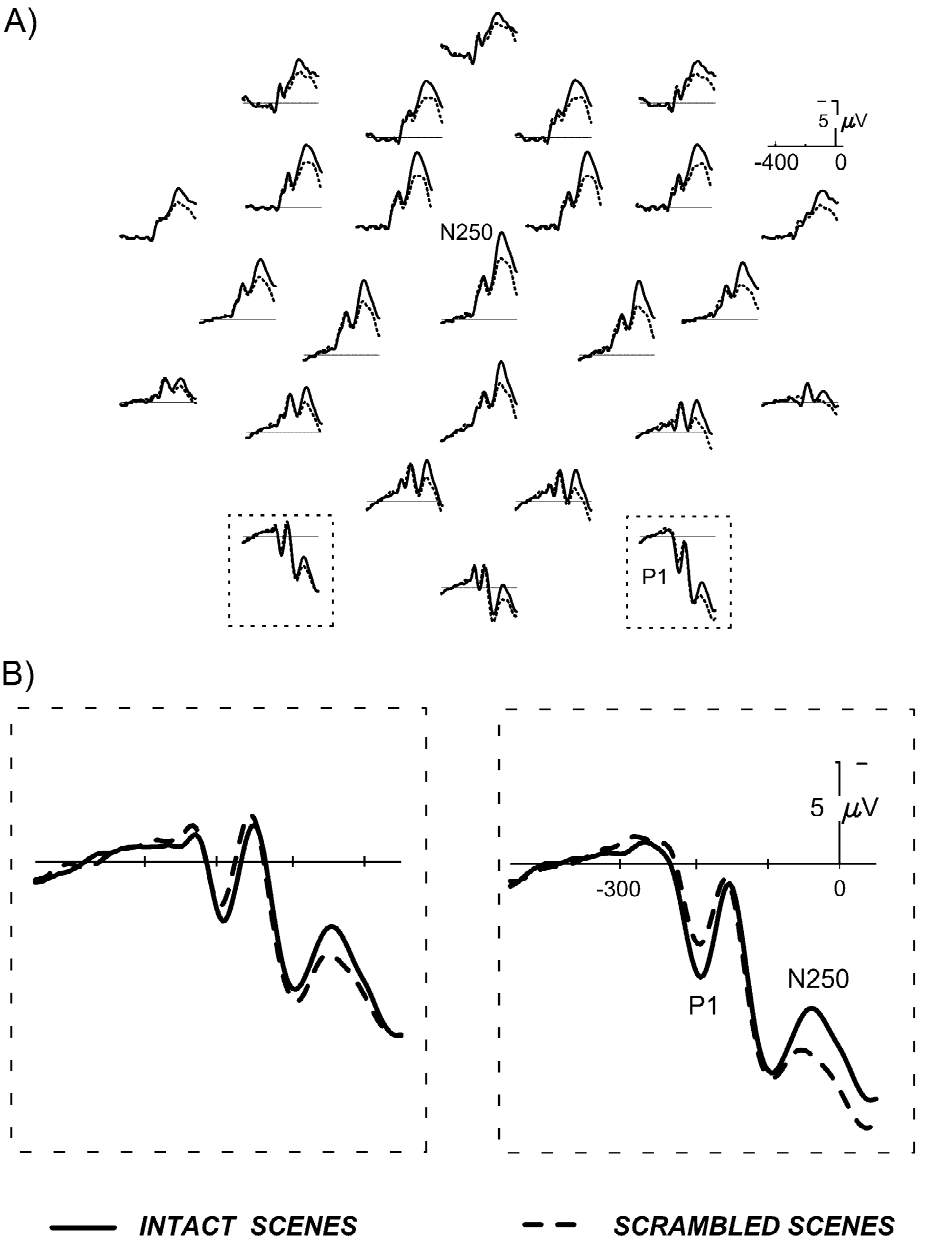

Fig. 6. (A) Overview of the ERPs elicited by the intact (solid line) and scrambled (dotted line) scenes before the onset of the target object. Intact scenes elicited a larger occipital P1 and a larger fronto-central N250 than scrambled ones. (B) Close-up of the effects at occipital sites 5 and 7.

$P<0.05$). The N250 (175–300 ms) was larger for intact than scrambled scenes, especially over fronto-central sites.

3.2.2.4. Effects of hemifield of presentation on early visual evoked potentials. As in Experiment 2, the occipital P1 was modulated by hemifield of presentation, as shown in Fig. 11 for occipital sites 5 and 7. At the right occipital site the P1 was larger in the left hemifield condition, while at the left occipital site it was larger in the right hemifield condition. Analyses as in Experiment 2 showed no main effect of hemifield ($F(1,13)=0.17$, $P>0.1$). The hemispheric asymmetry as a function of fixation was indicated by an interaction between hemifield of presentation and site ($F(1,13)=12.98$, $P<0.01$).

3.2.2.5. Comparison with sentence contexts. This experiment was similar to many previous studies with sentences in that no explicit congruity decision was required, allowing a direct comparison of the scalp distribution of the N390 scene congruity effect with that of the N400 sentence congruity effect [27]. The best comparison was with the blocked condition from [27], where pictures or words in independent blocks completed sentences presented one word at a time. Fig. 12 shows that, overall, the

Table 3
Summary of the ANOVAs on the mean ERPs as a function of scrambling (all 26 sites)

Time window (ms)	Experiment 2				Experiment 3			
	Scramb		Scramb×Site		Scramb		Scramb×Site	
	F	*P*	*F*	*P*	*F*	*P*	*F*	*P*
50–75	0.21	0.65	1.65	0.41	0.45	0.51	0.54	0.66
75–100	1.90	0.19	3.49	0.14	3.63	0.08	4.76	0.006*
100–125	5.23	0.04*	1.13	0.35	11.25	0.005*	2.35	0.04*
125–150	0.08	0.78	2.38	0.08	3.81	0.07	0.54	0.71
150–175	1.73	0.21	0.70	0.55	0.61	0.44	0.87	0.43
175–200	1.10	0.31	0.64	0.55	9.12	0.01*	1.34	0.28
200–225	8.72	0.01*	1.76	0.17	53.17	0.001*	2.18	0.13
225–250	23.09	0.001*	2.68	0.06	41.61	0.001*	3.59	0.02*
250–275	26.59	0.001*	2.85	0.05	25.74	0.001*	3.45	0.03*
275–300	10.03	0.007*	1.71	0.18	17.52	0.001*	4.35	0.01*

The scrambling factor is indicated as ‘Scramb’. Degrees of freedom are 1 and 13 for the main effects and 25 and 325 for the interactions.

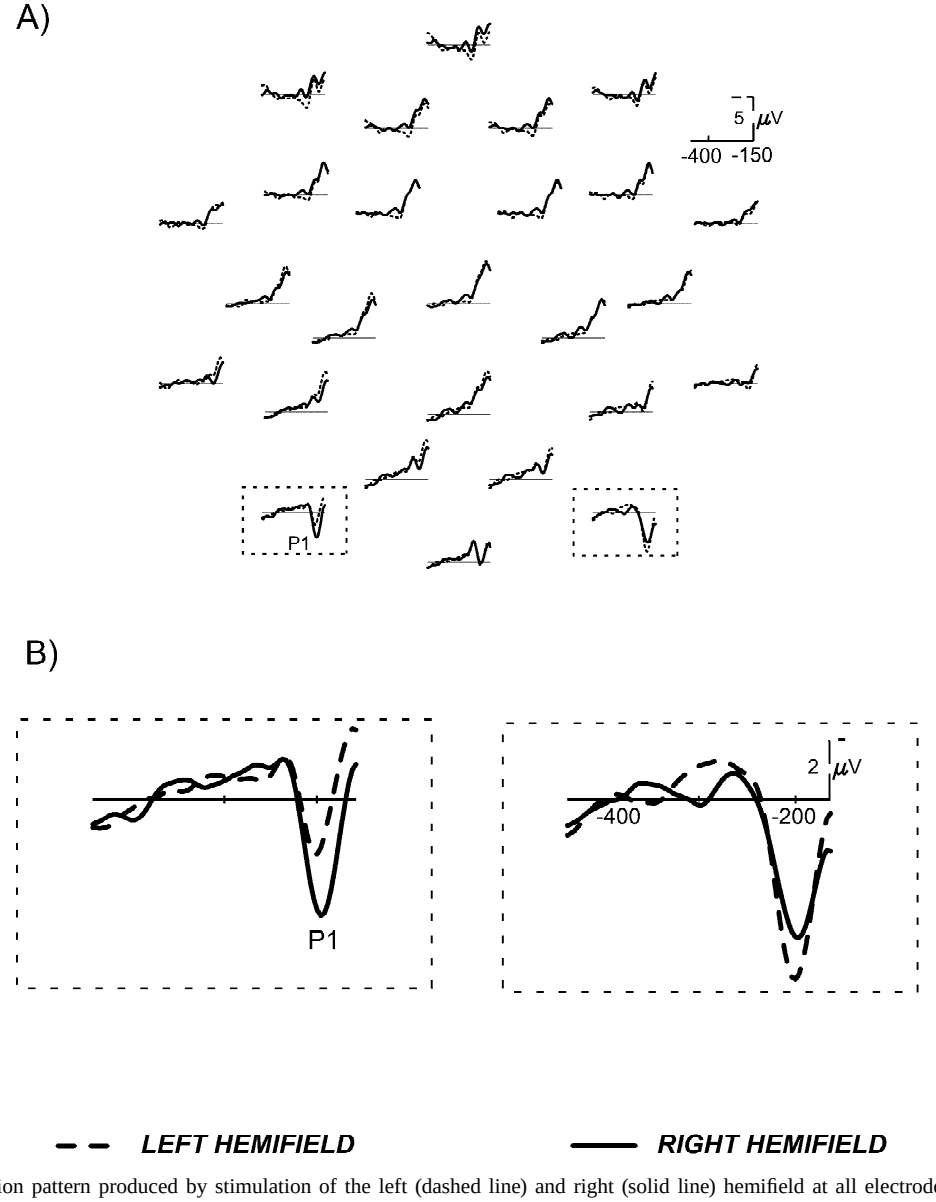

Fig. 7. (A) P1 modulation pattern produced by stimulation of the left (dashed line) and right (solid line) hemifield at all electrodes (Experiment 2). (B) Close-up of the effect at occipital sites 5 and 7.

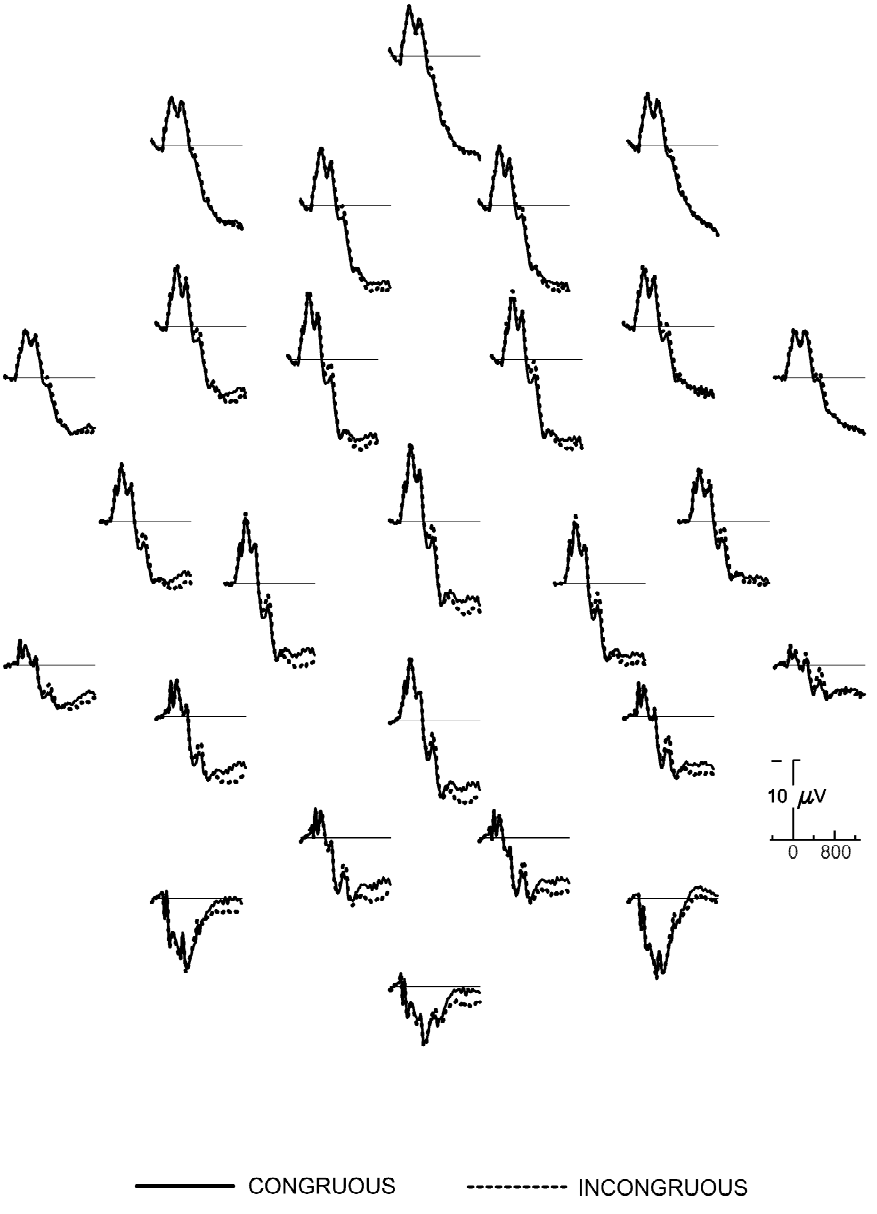

Fig. 8. Overview of ERPs elicited by the congruous (solid line) and incongruous (dotted line) sequences (Experiment 3).

scalp distributions of the scene and sentence congruity effects are quite similar. In fact, an analysis of the normalized scalp distributions of objects in scenes (for 12 of the participants, to equate group size) did not reliably differ from that of objects in sentences (experiment by site interaction: $F(25,550)=1.26$, $P>0.1$) but was different from that of words in sentences (experiment by site interaction: $F(25,550)=2.72$, $P<0.05$).

4. Discussion

4.1. N390 congruity effect

The first reliable effect of congruity, modulation of an N400-like component, begins ~300 ms, peaks~ 390 ms, and lasts until ~500 ms. Throughout this interval, congruous targets show more positivity than incongruous ones. The time course of this congruity effect is roughly similar to that reported for written words in sentences, suggesting that the neural processes underlying the interaction between a visual stimulus and the context in which it appears may operate under similar time constraints during reading and scene processing: both domains require the quick integration of highly processed visual information. The scalp distribution of the scene congruity effect is also similar to that observed with sentential contexts [27], more so for pictures than for words, suggesting proximity of the underlying neural generators (i.e., brain regions that generate the effects). We suggest that the N390 congruity effect

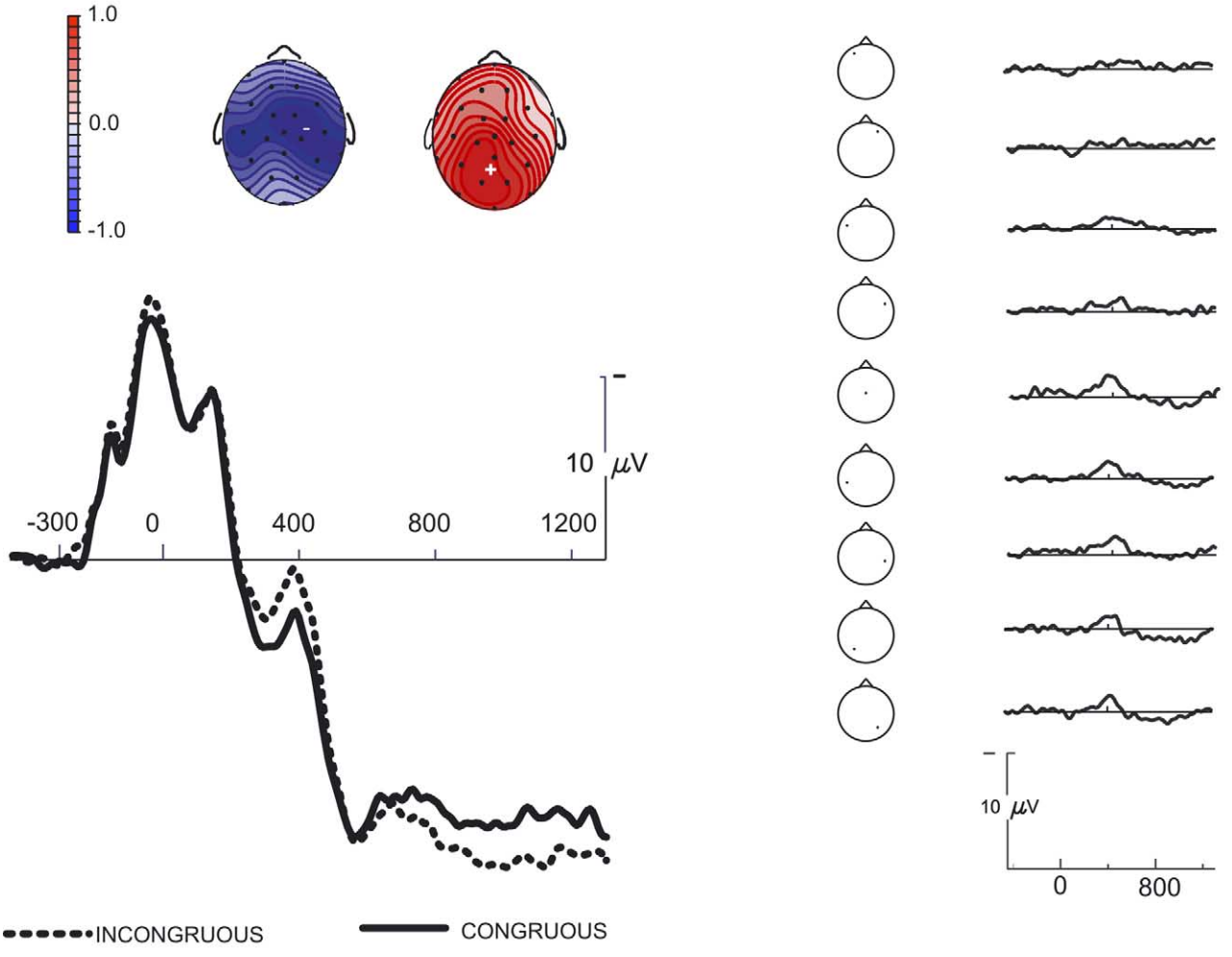

Fig. 9. (A) ERPs elicited by the congruous (solid line) and incongruous (dotted line) stimulus sequences at the midline parietal electrode (23) where both congruity effects are clearly visible (Experiment 3). The normalized scalp distribution of the congruity effects is illustrated by the isopotential maps (see for details Ref. [27]). The dashed line is the zero value isocontour. Positive and negative polarities are indicated with white '+' and '−' signs, respectively. The N390 congruity effect (incongruous minus congruous) is shown on the left (mean between 300 and 500 ms); the congruity effect on the late positivity is shown on the right (mean between 750 and 1250 ms). (B) Difference ERPs (incongruous minus congruous ERPs) at a subset of electrodes.

reflects scene influences on object processing at the level of semantic analysis.

While the spatial distribution of the N390 effect across the scalp cannot be used to determine its brain origin unequivocally, there is evidence to suggest that it is supported largely by structures in the anterior temporal lobes. A similar proposal has been made for the N400 sentence congruity effect based primarily on recordings of intracranial electrophysiological potentials from epilepsy patients. Like the scalp N400, intracranial ERPs recorded from regions of the anterior medial temporal lobe (aMTL potentials) are modulated by the semantic content and semantic context conveyed by written words and by factors influencing memory (e.g., Ref. [60]). AMTL potentials are typically recorded bilaterally near the anterior parts of the collateral sulcus, including perirhinal and entorhinal cortices [60,79]. The time course and pattern of modulation of aMTL potentials are similar to those of the N400s to written words recorded at the scalp [42]. Furthermore, the broad, bilateral distribution of the scalp N400 is consistent with the possibility that a large part of it may originate from within deep bilateral brain structures, such as the anterior temporal lobe. These properties of the N400 sentence congruity effect, together with the proximity of their proposed generators to medial temporal lobe structures for memory, have implicated them in the representation and/or activation of associations among written words [61].

Thus, by analogy with the sentence N400, we propose that anterior temporal regions may also be involved in generating the N390 scene congruity effect and thereby in the representation and/or activation of associations between non-linguistic visual stimuli. Several lines of evidence support this proposal. First, the time course and scalp distribution of the scene context effects are roughly similar to those obtained with sentences, suggesting a functional similarity and proximity of the underlying neural generators. Second, research with monkeys has

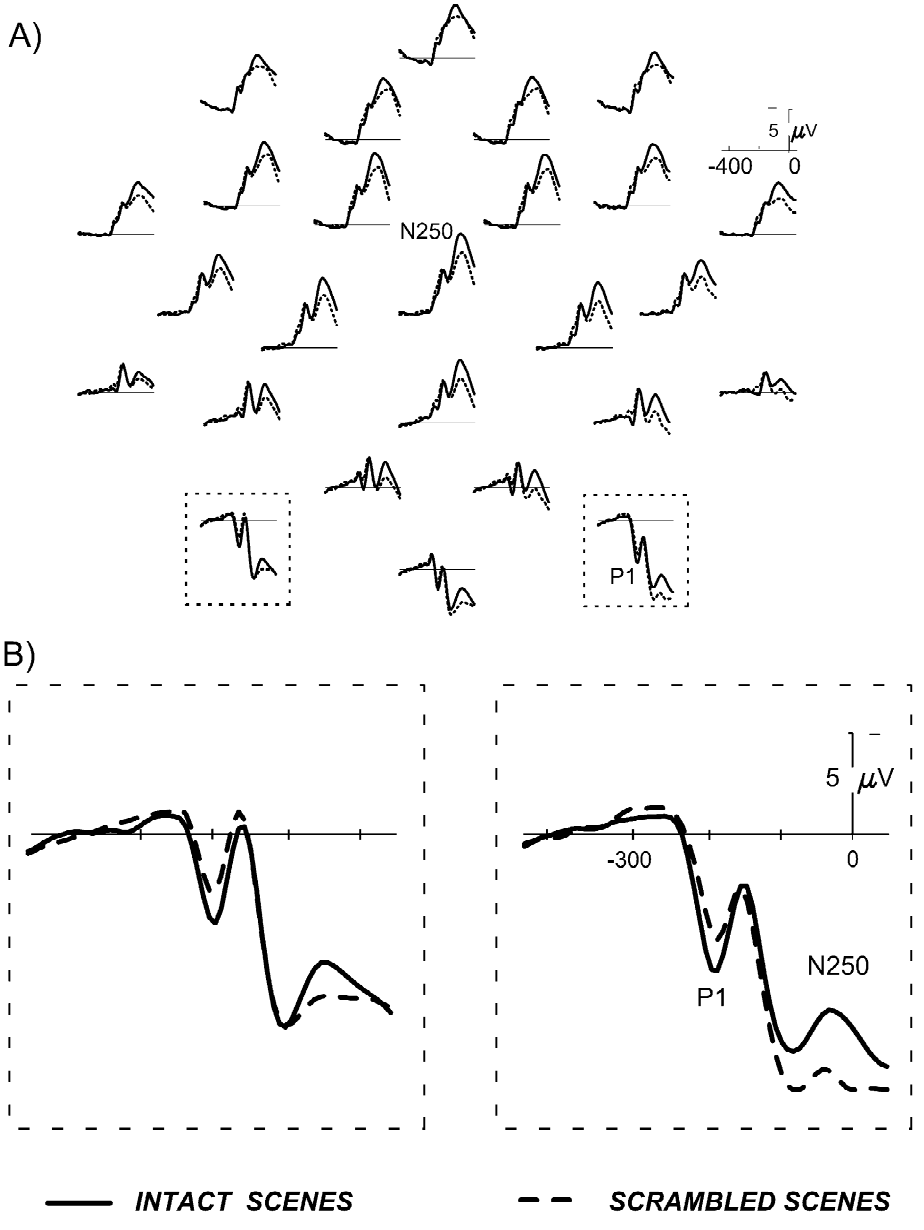

Fig. 10. (A) Overview of the ERPs elicited by the intact (solid line) and scrambled (dotted line) scenes before the onset of the target object. Intact scenes elicited a larger occipital P1 and a larger fronto-central N250 than scrambled ones. (B) Close-up of the effects at occipital sites 5 and 7.

implicated the perirhinal cortex in the maintenance of associations between representations of visual objects stored in high-level visual cortices [34,56,65,75]. Third, in humans, loss or degeneration of aMTL regions, as in 'Semantic Dementia', is associated with dramatic loss of associative visual and verbal knowledge but generally not of structural knowledge about objects [11,35,57,74].

In contrast, evidence from brain imaging studies suggests that structural object knowledge is represented in posterior temporal and occipital regions. Studies using stimuli with novel visual structures and an object decision task find activation in the posterior inferior temporal cortex [28,66]. Several studies [29,30,44] have identified a set of object-related areas in occipito-temporal regions whose response profile suggests a role for encoding structural information in terms of spatially localized structural descriptions of objects, but not necessarily for the full objects.

Overall, these data suggest a segregation of structural and semantic knowledge between the posterior and anterior regions of the temporal lobe, respectively. Together with the finding that patients with anterior temporal damage have difficulty in interpreting picture anomalies [53,54], these findings all support the idea that anterior temporal

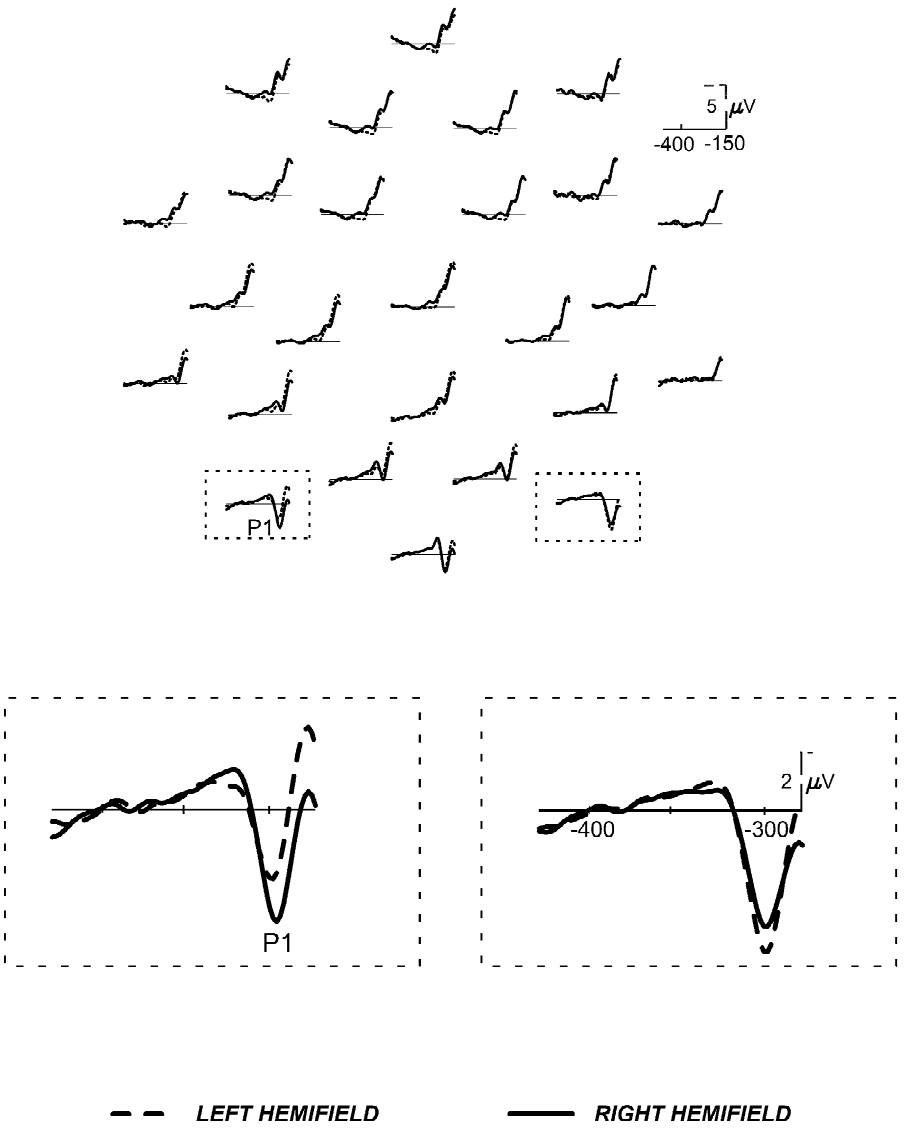

Fig. 11. (A) P1 modulation pattern produced by stimulation of the left (dashed line) and right (solid line) hemifield at all electrodes (Experiment 3). (B) Close-up of the effect at occipital sites 5 and 7.

areas may be involved in the ''quick access to semantic information derived from the arrangement of multiple objects into scenes'' (Ref. [11], p. 285).

Recent studies suggest that object identification relies upon many areas in a distributed associative network [20,48]. According to Damasio [19], aMTL regions are involved in reactivating associative information stored elsewhere. In addition, Martin et al. [48] showed that semantic information about categories of objects (e.g., animals and manipulable objects) may be stored in different sets of brain areas that are engaged during the interactions with such objects. Such proposals notwithstanding, there seems to be little doubt that anterior temporal lobe regions contain neural machinery used during the comprehension of written language and visual scenes. We suggest that the engagement of these regions is reflected in N400 sentence congruity effect and the N390 scene congruity effect.

4.2. Late positivities

The N390 effect was followed by a congruity effect on a late posterior positivity that was larger for incongruous than congruous objects. This congruity effect outlasted the overt response. The positivity may be a member of the P3 family [23], probably a P3b. Incongruous object-scene combinations are much more infrequent than congruous ones in people's visual experience. By some views of the P3b, then, these infrequent stimuli would produce a larger P3b than more frequent ones. A larger P3b could reflect

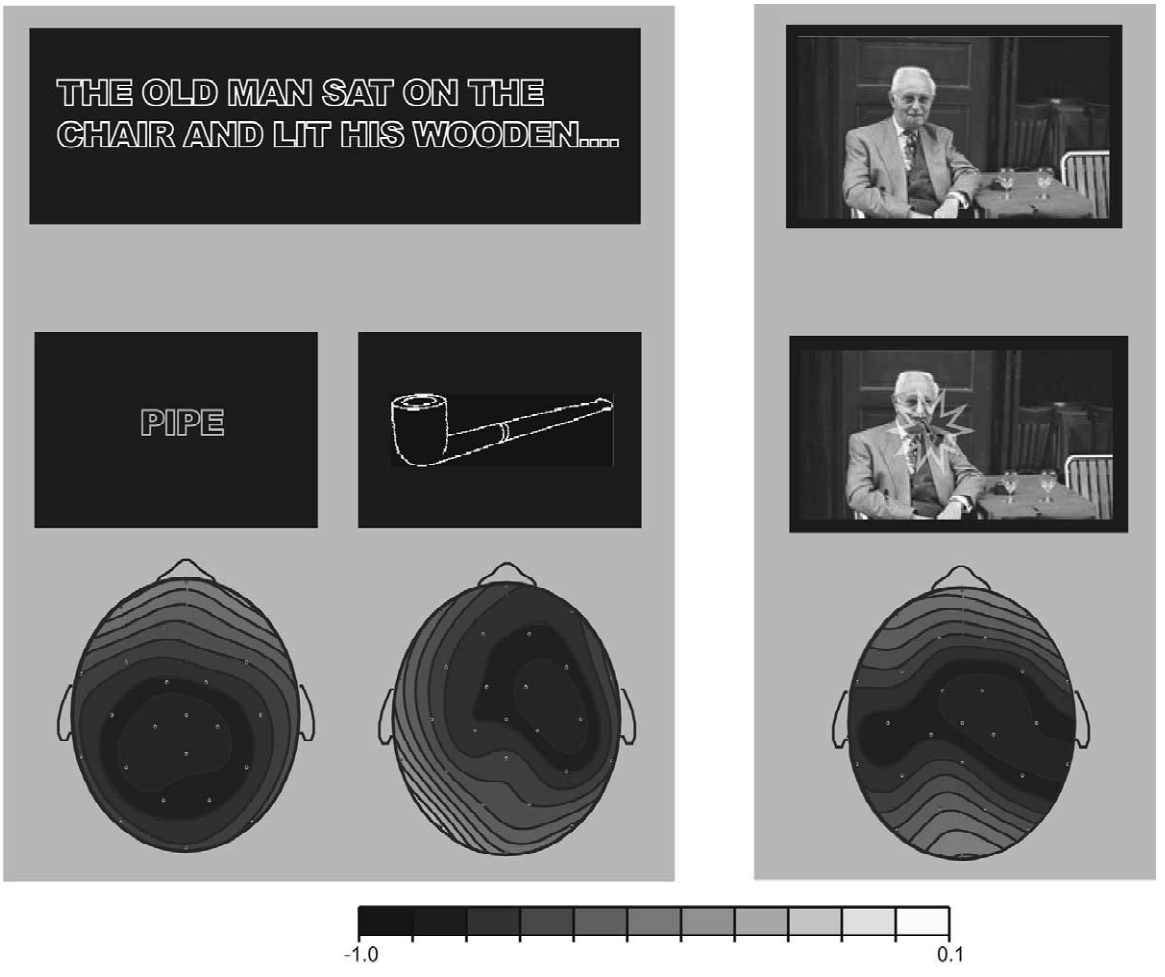

Fig. 12. Isopotential maps of the normalized distribution of the congruity effect between 300 and 500 ms for word in sentences (left) objects in sentences (middle) and objects in scenes (right). The specific stimuli shown (congruous versions) are used only for illustrative purposes and were not employed in the actual experiments. The original scattered data were interpolated with a spherical spline algorithm.

the engagement of processes required to integrate the incongruous object and scene into a mental model [23], perhaps to enable efficient episodic encoding. This interpretation is consistent with a study by Neville et al. [58] where scenes that were rated 'high' in surprise elicited more positive-going late ERPs than stimuli that were rated 'low' in surprise. In Experiment 2, the N390 was followed also by a frontal positivity, larger for congruous than incongruous objects. This frontal positivity is probably related to the (delayed) ratings required by the task. Indeed, such frontal positivity is absent in Experiment 3, where confidence and congruity ratings were not required. We suggest that the higher confidence people have in the identification of congruous than incongruous scenes (Experiment 2) may be responsible for the pattern of modulation of this positivity [38].

4.3. Early ERPs

We did not find any congruity effects in the ERPs to target objects before 300 ms poststimulus. Before addressing the implications of this result further, we note that this conclusion is based on a null finding. In general, the absence of an effect should be interpreted with caution for several reasons. For one, processing differences at a columnar or hyper-columnar level, may produce identical average electric fields at the scalp, or there may be 'silent' sources, i.e., neural generators whose activity cannot be detected at the scalp [1]. For another, the signal-to-noise ratio of early congruity effects may be lower than that of effects caused by other manipulations (e.g., hemifield of stimulus presentation). Further experiments with technologies able to zoom onto weaker neural signals will likely be needed to provide stronger evidence on this issue. A third, related, possibility is that the spatial sampling may be inadequate. Early components, such as the P1, are believed to be generated in posterior extrastriate areas, the same areas one would expect to be engaged by early scene effects on object identification. The finding that variables known to influence early perceptual processes modulate early ERPs in our study suggests that the spatial sampling per se is adequate in the present experiments. Specifically,

the amplitudes of the occipital P1 and the parietal N250 were both larger for intact than for scrambled scenes. Moreover, early ERPs, including the P1, were modulated by stimulus position in the visual field, as expected.

The P1 modulation for scrambled versus intact scenes may reflect spatial attention processes (e.g., Ref. [45]). Real scenes have regions of low and high informativeness (e.g., Ref. [51]) which their scrambled versions lack, and so real scenes might elicit differential patterns of spatial attention allocation. P1 modulation also may reflect differences in local contrast or high-level statistical regularities leading to different patterns of perceptual grouping. Consistent with this view, during the processing of intact and scrambled images, such as the ones herein, fMRI has shown differential activity in occipito-temporal extrastriate areas [44]. The N250 modulation may be related to the selection negativity (SN) observed in selective attention studies (see for review Ref. [2]). The visual system is continually exposed to scenes, thereby acquiring vast amounts of information about the statistical regularities of the visual world. These regularities may capture selective attention mechanisms or serve to guide the allocation of attention during visual scene processing [16]. These regularities (and their neural sequelae) distinguish the intact scenes from their scrambled versions, which do not contain them, as reflected in the N250 differences.

Regardless of their interpretation, these findings provide evidence that our experiments have sufficient power to detect small modulations of early visual evoked potentials (~1 μV), that is, early processing differences between conditions, when present. Based on this evidence, and bearing the aforementioned caveats in mind, we consider it reasonable to accept (a) the absence of early ERP differences between the congruous and incongruous conditions as a reliable non-effect, and (b) that 300 ms is a reasonable estimate of the actual onset of scene effects (at least at the neocortex). Such a finding can be explained more easily by Post-Identification accounts of object identification in scenes (e.g., Refs. [33,37]) than by accounts that stipulate an early locus of scene effects. As articulated in the introduction, these early locus accounts postulate that, once a schema is activated, it can affect the analysis of the perceptual features of target objects (Early Perception Accounts) or the matching of structural descriptions of these objects with internal models (Matching Process Accounts). We found no evidence for either process: scene congruity did not modulate either ERP components implicated in the encoding and analysis of perceptual features (P1, N1, P2, SN) or ERPs that may index the matching process (Nsd).

Our finding that congruity does not modulate early ERPs suggests that the type of information that can be extracted during the brief presentation of a scene is not precise enough to support congruity effects in early visual areas. A glance at a scene may provide enough information to predict classes of likely objects but not specific object exemplars. For example, a briefly-presented office scene will activate associated knowledge about desks. It is unlikely, however, that without prior exposure to this particular office ensemble the activated structural representations would be identical to the specific desk flashed in the scene. In other words, even if top-down mechanisms activated by the semantic content of a briefly presented scene were capable of priming representations of a generic desk in early visual areas, any such representation is likely to differ considerably from that activated by the actual instance of the target desk. This is because the actual desk can appear in a non-canonical orientation, with an unusual shape, and so on. After all, unlike higher-level visual areas, such as TE, early visual areas are very sensitive to these kinds of stimulus variations [76]. Accordingly, the high-level constraints provided by semantically related but unfamiliar contexts are not exact enough to prime the neural representation in early visual areas which the actual target object activates. In our example, then, priming by an office scene can occur reliably only at a level where different desks would activate sufficiently similar neural representations; that is, at a fairly high, semantic-level.

This interpretation is consistent with prior studies on semantic priming between isolated objects where congruity effects have been observed before 300 ms [36,50]. In these studies, a frontal N300 is often seen that is modulated by semantic relatedness. Such earlier effects may not have been present in our study for several reasons. First, it takes longer to identify a small object embedded in a scene than a large object in isolation, or against a uniform background [13]; thus when using scenes, as herein, early ERP congruity effects implicated in object-object priming may just be delayed. Second, the delay between prime and target onset in the object priming studies is typically longer than that used in the current study; during this delay, additional information elicited by the prime may become available to early visual areas, resulting in earlier congruity effects. Third, the surface features of our objects were quite varied, being constrained by the scenes which were shot from different angles, in different lighting, etc. The target object sets used in the object priming studies tend to have more uniform surface features, perhaps enabling earlier visual areas to predict some of these attributes. A strong prediction of this explanation is that early congruity effects should occur if people see the target object set in advance.

4.4. Behavioral findings

Our behavioral results replicate the known advantage for the identification of congruous relative to incongruous stimuli (e.g., Ref. [13]). Following the logic used in prior behavioral studies [14], the RTs in the ‘scrambled condition’ are more similar to those in the congruous than incongruous condition (Table 1), and so it is tempting to conclude that the behavioral congruity effect is due to

inhibition in the incongruous condition instead of facilitation in the congruous one. To the contrary, both Experiments 2 and 3 demonstrate ERP differences between the scrambled and the intact scenes occurring well before the onset of the target object (Table 3). Thus, RT differences between the scrambled and intact conditions can be due to their differential processing before the onset of the target objects and so may have little to do with congruity per se. Further work will be required to address this challenging issue.

5. Conclusions

The current experiments, together with prior findings, suggest the following picture of object identification in briefly flashed scenes. Within the first 250–300 ms, visual inputs are processed without any notable impact of top-down influences from the semantic content of a scene. The effects of a scene or a sentence context on object identification processes occur somewhat later, ~300 ms, as reflected in the N390 scene congruity effect; here, information in memory accessed by the scene context is related to a high level (semantic) representation of the object. At this point, an appropriate context may prime relevant semantic representations via activity in the anterior temporal lobes.

Acknowledgements

Work reported herein was supported by grants MH52893, HD22614, and AG08313 to M. Kutas. During the revision process of this article the first author was supported by a McDonnell-Pew Program in Cognitive Neuroscience Award and by grants NMA 201-01-C-0032 and 5R01 MH 60734-03 to Stephen M. Kosslyn. We would like to thank Haline E. Schendan and Stephen M. Kosslyn for helpful discussion.

References

[1] A. Amir, Uniqueness of the generators of brain evoked potential maps, IEEE Trans. Biomed. Eng. 41 (1994) 1–11.
[2] L. Anllo-Vento, S.A. Hillyard, Selective attention to the color and direction of moving stimuli: electrophysiological correlates of hierarchical feature selection, Percept. Psychophys. 58 (1996) 191–206.
[3] L. Anllo-Vento, S.J. Luck, S.A. Hillyard, Spatio-temporal dynamics of attention to color: evidence from human electrophysiology, Hum. Brain Mapp. 6 (1998) 216–238.
[4] J.R. Antes, J.G. Penland, R.L. Metzger, Processing global information in briefly presented pictures, Psychol. Res. 43 (1981) 277–292.
[5] S.E. Barrett, M.D. Rugg, Asymmetries in event-related potentials during rhyme-matching: confirmation of the null effects of handedness, Neuropsychologia 27 (1989) 539–548.
[6] S.E. Barrett, M.D. Rugg, Event-related potentials and the phonological matching of picture names, Brain Lang. 38 (1990) 424–437.
[7] S.E. Barrett, M.D. Rugg, Event-related potentials and the semantic matching of pictures, Brain Cogn. 14 (1990) 201–212.
[8] S.E. Barrett, M.D. Rugg, D.I. Perrett, Event-related potentials and the matching of familiar and unfamiliar faces, Neuropsychologia 26 (1988) 105–117.
[9] I. Biederman, Aspects and extensions of a theory of human image understanding, in: Z.W. Pylyshin (Ed.), Computational Processes In Human Vision: An Interdisciplinary Approach, Ablex, Norwood, NJ, 1988.
[10] I. Biederman, R.J. Mezzanotte, J.C. Rabinowitz, Scene perception: detecting and judging objects undergoing relational violations, Cognit. Psychol. 14 (1982) 143–177.
[11] I. Biederman, P.C. Gerhardstein, E.E. Cooper, C.A. Nelson, High level object recognition without an anterior inferior temporal lobe, Neuropsychologia 35 (1997) 271–287.
[12] M.A. Bobes, M. Valdes-Sosa, E. Olivares, An ERP study of expectancy violation in face perception, Brain Cogn. 26 (1994) 1–22.
[13] S.J. Boyce, A. Pollatsek, Identification of objects in scenes: the role of scene background in object naming, J. Exp. Psychol. Learn. Mem. Cogn. 18 (1992) 531–543.
[14] S.J. Boyce, A. Pollatsek, K. Rayner, Effect of background information on object identification, J. Exp. Psychol. Hum. Percept. Perform. 15 (1989) 556–566.
[15] H. Buchner, U. Weyen, R.S. Frackowiak, J. Romaya, S. Zeki, The timing of visual evoked potential activity in human area V4, Proc. R. Soc. Lond. B Biol. Sci. 257 (1994) 99–104.
[16] M.M. Chun, Contextual cueing of visual attention, Trends Cogn. Sci. 4 (2000) 170–178.
[17] V.P. Clark, S. Fan, S.A. Hillyard, Identification of early visual evoked potential generators by retinotopic analyses, Hum. Brain Mapp. 2 (1995) 170–187.
[18] V.P. Clark, S. Fan, S.A. Hillyard, Identification of early visual evoked potential generators by retinotopic and topographic analyses, Hum. Brain Mapp. 2 (1995) 170–187.
[19] A.R. Damasio, The brain binds entities and events by multiregional activation from convergence zones, Neural Comput. 1 (1989) 123–132.
[20] A.R. Damasio, Time-locked multiregional retroactivation: a systems-level proposal for the neural substrates of recall and recognition, Cognition 33 (1989) 25–62.
[21] P. De Graef, Scene-context effects and models of real-world perception, in: K. Rayner (Ed.), Eye Movements and Visual Cognition: Scene Perception and Reading, Springer, New York, 1992, pp. 243–259.
[22] P. De Graef, A. De Troy, G. D'Ydewalle, Local and global contextual constraints on the identification of objects in scenes. Special Issue: Object perception and scene analysis, Can. J. Psychol. 46 (1992) 489–508.
[23] E. Donchin, M.G.H. Coles, Is the P300 component a manifestation of context updating?, Behav. Brain Sci. 11 (1988) 357–374.
[24] G.M. Doninger, J.J. Foxe, M.M. Murray, B.A. Higgins, J.G. Snodgrass, C.E. Schroeder, D.C. Javitt, Activation timecourse of ventral visual stream object-recognition areas: high density electrical mapping of perceptual closure processes, J. Cogn. Neurosci. 12 (2000) 615–621.
[25] A. Friedman, Framing pictures: the role of knowledge in automatized encoding and memory for gist, J. Exp. Psychol. Gen. 108 (1979) 316–355.
[26] G. Ganis, H.E. Schendan, M. Kutas, Reading pictures, viewing text: an event-related potential study, Abstr. Soc. Neurosci. 21 (1995) 279.1.
[27] G. Ganis, M. Kutas, M.I. Sereno, The search for common sense—an electrophysiological study of the comprehension of words and pictures in reading, J. Cogn. Neurosci. 8 (1996) 89–106.

[28] C. Gerlach, I. Law, A. Gade, O.B. Paulson, Perceptual differentiation and category effects in normal object recognition: a PET study, Brain 122 (1999) 2159–2170.
[29] K. Grill-Spector, T. Kushnir, S. Edelman, Y. Itzchak, R. Malach, Cue-invariant activation in object-related areas of the human occipital lobe, Neuron 21 (1998) 191–202.
[30] K. Grill-Spector, T. Kushnir, T. Hendler, R. Malach, The dynamics of object-selective activation correlate with recognition performance in humans, Nat. Neurosci. 3 (2000) 837–843.
[31] M.R. Harter, F.H. Previc, Size-specific information channels and selective attention: visual evoked potential and behavioral measures, Electroencephalogr. Clin. Neurophysiol. 45 (1978) 628–640.
[32] H.J. Heinze, G.R. Mangun, W. Burchert, H. Hinrichs, M. Scholz, T.F. Munte, A. Gos, M. Scherg, S. Johannes, H. Hundeshagen et al., Combined spatial and temporal imaging of brain activity during visual selective attention in humans, Nature 372 (1994) 543–546.
[33] J.M. Henderson, A. Hollingworth, High-level scene perception, Annu. Rev. Psychol. 50 (1999) 243–271.
[34] S. Higuchi, Y. Miyashita, Formation of mnemonic neuronal responses to visual paired associates in inferotemporal cortex is impaired by perirhinal and entorhinal lesions, Proc. Natl. Acad. Sci. USA 93 (1996) 739–743.
[35] J.R. Hodges, N. Graham, K. Patterson, Charting the progression in semantic dementia: implications for the organisation of semantic memory, Memory 3 (1995) 463–495.
[36] P.J. Holcomb, W.B. McPherson, Event-related brain potentials reflect semantic priming in an object decision task, Brain Cogn. 24 (1994) 259–276.
[37] A. Hollingworth, J.M. Henderson, Object identification is isolated from scene semantic constraint: evidence from object type and token discrimination, Acta Psychol. 102 (1999) 319–343.
[38] D. Karis, M. Fabiani, E. Donchin, 'P300' and memory: Individual differences in the von Restorff effect, Cogn. Psychol. 16 (1984) 177–216.
[39] J.L. Kenemans, A. Kok, F.T. Smulders, Event-related potentials to conjunctions of spatial frequency and orientation as a function of stimulus parameters and response requirements, Electroencephalogr. Clin. Neurophysiol. 88 (1993) 51–63.
[40] J.L. Kenemans, J.M. Baas, G.R. Mangun, M. Lijffijt, M.N. Verbaten, On the processing of spatial frequencies as revealed by evoked-potential source modeling, Clin. Neurophysiol. 111 (2000) 1113–1123.
[41] S.M. Kosslyn, in: Image and Brain, MIT Press, Cambridge, MA, 1994.
[42] M. Kutas, C. Van Petten, Psycholinguistics electrified: event-related potential investigations, in: M. Gernsbacher (Ed.), Handbook of Psycholinguistics, Academic Press, New York, 1994, pp. 83–143.
[43] G.R. Loftus, W.W. Nelson, H.J. Kallman, Differential acquisition rates for different types of information from pictures, Q. J. Exp. Psychol. 35A (1983) 187–198.
[44] R. Malach, J.B. Reppas, R.R. Benson, K.K. Kwong, H. Jiang, W.A. Kennedy, P.J. Ledden, T.J. Brady, B.R. Rosen, R.B. Tootell, Object-related activity revealed by functional magnetic resonance imaging in human occipital cortex, Proc. Natl. Acad. Sci. USA 92 (1996) 8135–8139.
[45] G.R. Mangun, S.A. Hillyard, Allocation of visual attention to spatial locations: Tradeoff functions for event-related brain potentials and detection performance, Percept. Psychophys. 47 (1990) 532–550.
[46] M.G. Manolas, T.D. Stamoulos, P.A. Anninos, Differences in human visual evoked potentials during the perception of colour as revealed by a bootstrap method to compare cortical activity. A prospective study, Neurosci. Lett. 270 (1999) 21–24.
[47] D. Marr, in: Vision: A Computational Investigation Into the Human Representation and Processing of Visual Information, W.H. Freeman, San Francisco, CA, 1982.
[48] A. Martin, C.L. Wiggs, L.G. Ungerleider, J.V. Haxby, Neural correlates of category-specific knowledge, Nature 379 (1996) 649–652.
[49] G. McCarthy, C.C. Wood, Scalp distributions of event-related potentials: An ambiguity associated with analysis of variance models, Electroencephalogr. Clin. Neurophysiol. 62 (1985) 203–208.
[50] W. McPherson, P. Holcomb, An electrophysiological investigation of semantic priming with pictures of real objects, Psychophysiology 36 (1999) 53–65.
[51] R.L. Metzger, J.R. Antes, The nature of processing early in picture perception, Psychol. Res. 45 (1983) 267–274.
[52] J. Miller, Reaction time analysis with outlier exclusion: Bias varies with sample size, Q. J. Exp. Psychol. 43 (1991) 907–912.
[53] B. Milner, Psychological deficits produced by temporal-lobe excision, Proc. Assoc. Res. Nerv. Ment. Dis. 36 (1958) 244–257.
[54] B. Milner, Visual recognition and recall after right temporal-lobe excision in man, Neuropsychologia 6 (1968) 191–209.
[55] Y. Miyashita, Inferior temporal cortex: where visual perception meets memory, Annu. Rev. Neurosci. 16 (1993) 245–263.
[56] Y. Miyashita, H. Okuno, W. Tokuyama, T. Ihara, K.-i. Nakajima, Feedback signal from medial temporal lobe mediates visual associative mnemonic codes of inferotemporal neurons, Brain Res. Cogn. Brain Res. 5 (1996) 81–86.
[57] E.A. Murray, T.J. Bussey, Perceptual-mnemonic functions of the perirhinal cortex, Trends Cogn. Sci. 3 (1999) 142–151.
[58] H. Neville, E. Snyder, D. Woods, R. Galambos, Recognition and surprise alter the human visual evoked response, Proc. Natl. Acad. Sci. USA 79 (1982) 2121–2123.
[59] M. Niedeggen, E.R. Wist, Characteristics of visual evoked potentials generated by motion coherence onset, Brain Res. Cogn. Brain Res. 8 (1999) 95–105.
[60] A.C. Nobre, G. McCarthy, Language-related ERPs: scalp distributions and modulation by word type and semantic priming, J. Cogn. Neurosci. 6 (1994) 233–255.
[61] A.C. Nobre, G. McCarthy, Language-related field potentials in the anterior-medial temporal lobe: II. Effects of word type and semantic priming, J. Neurosci. 15 (1995) 1090–1099.
[62] F.H. Previc, M.R. Harter, Electrophysiological and behavioral indicants of selective attention to multifeature gratings, Percept. Psychophys. 32 (1982) 465–472.
[63] M.J. Riddoch, G.W. Humphreys, Visual object processing in optic aphasia: a case of semantic access agnosia, Cogn. Neuropsychol. 4 (1987) 131–185.
[64] D.E. Rumelhart, P. Smolensky, J.L. McClelland, G.E. Hinton, Schemata and sequential thought processes in PDP models, in: E.E. Smith (Ed.), Readings in Cognitive Science: A Perspective From Psychology and Artificial Intelligence, Kaufmann, San Mateo, CA, 1988, pp. 224–249.
[65] K.S. Saleem, K. Tanaka, Divergent projections from anterior inferotemporal area TE to the perirhinal and entorhinal cortices in the macaque monkey, J. Neurosci. 16 (1996) 4757–4775.
[66] D.L. Schacter, E. Reiman, A. Uecker, M.R. Polster, L.S. Yun, L.A. Cooper, Brain regions associated with retrieval of structurally coherent visual information, Nature 376 (1995) 587–590.
[67] H.E. Schendan, M. Kutas, Neurophysiological evidence for two processing times for visual object identification, Neuropsychologia 40 (2002).
[68] H.E. Schendan, G. Ganis, M. Kutas, Neurophysiological evidence for visual perceptual categorization of words and faces within 150 ms, Psychophysiology 35 (1998) 240–251.
[69] S.R. Schweinberger, How Gorbachev primed Yeltsin: analyses of associative priming in person recognition by means of reaction times and event-related brain potentials, J. Exp. Psychol. Learn. Mem. Cogn. 22 (1996) 1383–1407.
[70] S.R. Schweinberger, E.-M. Pfutze, W. Sommer, Repetition priming and associative priming of face recognition: evidence from event-related potentials, J. Exp. Psychol. Learn. Mem. Cogn. 21 (1995) 722–736.
[71] P.G. Schyns, From blobs to boundary edges: evidence for time- and

spatial-scale-dependent scene recognition, Psychol. Sci. 5 (1994) 195–200.
[72] H.G. Smid, G. Mulder, L.J. Mulder, G.J. Brands, A psychophysiological study of the use of partial information in stimulus-response translation, J. Exp. Psychol. Hum. Percept. Perform. 18 (1992) 1101–1119.
[73] R. Srebro, A bootstrap method to compare the shapes of two scalp fields, Electroencephalogr. Clin. Neurophysiol. 100 (1996) 25–32.
[74] K. Srinivas, S.D. Breedin, H.B. Coslett, E.M. Saffran, Intact perceptual priming in a patient with damage to the anterior inferior temporal lobes, J. Cogn. Neurosci. 9 (1997) 490–511.
[75] W.A. Suzuki, D.G. Amaral, Perirhinal and parahippocampal cortices of the macaque monkey: cortical afferents, J. Comp. Neurol. 350 (1994) 497–533.
[76] K. Tanaka, Inferotemporal cortex and object vision, Annu. Rev. Neurosci. 19 (1996) 109–139.
[77] S. Thorpe, D. Fize, C. Marlot, Speed of processing in the human visual system, Nature 381 (1996) 520–522.
[78] S. Ullman, in: High-level Vision: Object Recognition and Visual Cognition, MIT Press, Cambridge, MA, 1996, p. 412.
[79] G.W. Van Hoesen, Anatomy of the medial temporal lobe, Magn. Reson. Imaging 13 (1995) 1047–1055.
[80] C. Van Petten, Words and sentences: event-related brain potential measures, Psychophysiology 32 (1995) 511–525.