Neuropsychologia 56 (2014) 447–458

Contents lists available at ScienceDirect

Neuropsychologia

journal homepage: www.elsevier.com/locate/neuropsychologia

Synchronous contextual irregularities affect early scene processing: Replication and extension

Liad Mudrik [a,b,*,1], Shani Shalgi [c], Dominique Lamy [a], Leon Y. Deouell [d]

[a] Department of Psychology, Tel Aviv University, PO Box 39040, Tel Aviv 69978, Israel
[b] Division of Biology, California Institute of Technology, 1200 E California Blvd, Pasadena, CA 91125, USA
[c] Department of Cognitive Science, The Hebrew University of Jerusalem, Jerusalem 91905, Israel
[d] Department of Psychology and the Edmond and Lily Safra Center for brain sciences, The Hebrew University of Jerusalem, Jerusalem 91905, Israel

ARTICLE INFO

Article history:
Received 19 August 2013
Received in revised form
19 February 2014
Accepted 20 February 2014
Available online 2 March 2014

Keywords:
Visual scenes
Context effects
N300/N400
Event-related potentials
Incongruency
Matching models

ABSTRACT

Whether contextual regularities facilitate perceptual stages of scene processing is widely debated, and empirical evidence is still inconclusive. Specifically, it was recently suggested that contextual violations affect early processing of a scene only when the incongruent object and the scene are presented a-synchronously, creating expectations. We compared event-related potentials (ERPs) evoked by scenes that depicted a person performing an action using either a congruent or an incongruent object (e.g., a man shaving with a razor or with a fork) when scene and object were presented simultaneously. We also explored the role of attention in contextual processing by using a pre-cue to direct subjects' attention towards or away from the congruent/incongruent object. Subjects' task was to determine how many hands the person in the picture used in order to perform the action. We replicated our previous findings of frontocentral negativity for incongruent scenes that started ~210 ms post stimulus presentation, even earlier than previously found. Surprisingly, this incongruency ERP effect was negatively correlated with the reaction times cost on incongruent scenes. The results did not allow us to draw conclusions about the role of attention in detecting the regularity, due to a weak attention manipulation. By replicating the 200–300 ms incongruity effect with a new group of subjects at even earlier latencies than previously reported, the results strengthen the evidence for contextual processing during this time window even when simultaneous presentation of the scene and object prevent the formation of prior expectations. We discuss possible methodological limitations that may account for previous failures to find this an effect, and conclude that contextual information affects object model selection processes prior to full object identification, with semantic knowledge activation stages unfolding only later on.

1. Introduction

Biological organisms grasp and interpret visual scenes amazingly fast and effortlessly. How they overcome the formidable challenge of processing the enormous amount of details embedded in natural scenes is one of the greatest puzzles in the study of visual perception. Significant help in achieving this feat may come from the existence of contextual regularities: objects tend to co-appear in particular scenes, allowing for prior knowledge and expectations to narrow the range of probable interpretations, thereby rendering scene analysis easier. Indeed, when such expectations are violated (e.g., a whale showing up in the middle of a football stadium), scene processing is impeded (Biederman, Glass, & Stacy, 1973; Biederman, Rabinowitz, Glass, & Stacy, 1974; Friedman, 1979; Palmer, 1975; Rayner & Pollatsek, 1992), in terms of both speed (Bar & Ullman, 1996; Boyce & Pollatsek, 1992; Chun & Jiang, 1998; Davenport & Potter, 2004) and accuracy (e.g., Antes, Penland, & Metzger, 1981; Bar & Ullman, 1996; Boyce, Pollatsek, & Rayner, 1989).

Evaluation of contextual relations during perceptual stages of scene processing, prior to full identification, would allow maximal benefits and facilitate the ongoing processing of both the scene and its constituents. However, whether contextual evaluation indeed facilitates perception remains controversial. Some theoretical models deny any contextual processing prior to scene and objects identification, and claim that it can occur only at later, post-perceptual stages (i.e., Functional isolation models; De Graef, 1992; Hamm, Johnson, & Kirk, 2002; Hollingworth & Henderson, 1998, 1999), at least 300 ms after the scene has been presented (Ganis & Kutas, 2003). Others posit that contextual processing occurs earlier and influences object identification processes. Such

* Corresponding author at: Department of Psychology, Tel Aviv University, Ramat Aviv, POB 39040, Tel Aviv 69978, ISRAEL.
E-mail address: liadmu@gmail.com (L. Mudrik).
[1] Current address: Division of Biology, California Institute of Technology, 1200 E California Blvd, Pasadena, CA, 91125, USA.

http://dx.doi.org/10.1016/j.neuropsychologia.2014.02.020
0028-3932

influence can take place when object processing commences, during the first 200 ms of scene processing (when initial differences between object catoegories are observed; Thorpe, Fize, & Marlot, 1996; VanRullen & Thorpe, 2001), at the stage of attentional feature selection (i.e., Perceptual schema models; Antes et al., 1981; Biederman, Mezzanotte, & Rabinowitz, 1982; Boyce et al., 1989). Alternatively, contextual processing were suggested to facilitate object identification at somewhat later stages (i.e., Object model selection or Matching models; Bar, 2004; Bar & Aminoff, 2003; Bar & Ullman, 1996; Kosslyn, 1994), between 200 ms and 300 ms post stimulus presentation (Schendan & Kutas, 2002, 2003; Schendan & Maher, 2008), when pre-activated scene-congruent object representations are being matched with upcoming visual information about the scene's constituents.

Relevant empirical evidence has been inconclusive. In particular, a recent series of ERP studies yielded conflicting results. Effects of contextual processing of congruent and incongruent scenes in the 200–300 ms time window, prior to full object identification, were found in three previous studies (Mudrik et al., 2010; Sun, Simon-Dack, Gordon, & Teder, 2011; Võ & Wolfe, 2013). For instance, we (Mudrik et al., 2010) reported an anterior negativity related to incongruent scenes that started around 270 ms post scene presentation, and lasted about 330 ms. This negativity was followed by a later broadly distributed negativity between 650 ms and 850 ms, possibly related to late processes of semantic evaluation and response preparation. The earlier negativity we found was interpreted as a combination of the N300 (McPherson & Holcomb, 1999; Sitnikova, Holcomb, Kiyonaga, & Kuperberg, 2008) and N400 (Kutas & Hillyard, 1980a, 1980b) components. N300, which occurs 200–300 after stimulus onset, was previously suggested to reflect processes that lead to object identification (Ganis & Kutas, 2003). Accordingly, its amplitude is higher for unidentified than for identified objects (Folstein, Van Petten, & Rose, 2008; Holcomb & McPherson, 1994; Schendan & Kutas, 2002), and is modulated by identification difficulty (Doniger et al. 2000; Henson, Rylands, Ross, Vuilleumeir, & Rugg, 2004; Holcomb & McPherson, 1994). Thus, this finding was taken as evidence supporting matching models of contextual processing, which postulate that scenes activate schemas that reduce the amount of perceptual evidence needed to match a particular schema-congruent object with its representation.

However, in a previous study Ganis and Kutas (2003) failed to observe such an N300 effect and reported only a later negativity, namely the "N390 congruency effect", that emerged in the 300–500 ms time window, similarly to the N400 component, albeit with a more frontal distribution. The absence of any earlier differences in either the 200–300 or the 0–200 time-windows was interpreted as ruling out contextual influences on perceptual stages of scene processing, thereby supporting functional isolation models.

The discrepancy between these two findings is especially surprising because it should have been easier to observe earlier differences using Ganis and Kutas' a-synchronous paradigm than using ours. Ganis and Kutas first presented a pre-cue, followed by the scene, and only then added the critical object at the cued location. Thus, subjects had time to form expectations regarding probable objects that matched the scene. By contrast, we presented the scene and object simultaneously in order to prevent subjects from forming prior expectations (see Mudrik et al., 2010 for a detailed argumentation). Nevertheless, we found the early N300 described above.

This discrepancy widens when considering a more recent ERP study (Demiral et al., 2012), which manipulated objects' spatial congruency (e.g., a bus was presented in the sky vs. on the road), rather than their semantic congruency (we use the term "semantic congruency" following Biederman (1981), to denote contextual violations in which the probability of an object to occur in a scene is manipulated. Accordingly, such contextual violations rest on previous knowledge about the co-occurrence of objects and scenes). Demiral et al. conducted two experiments: the first followed Ganis and Kutas' (2003) sequential design, that is, a pre-cue was presented first, followed by the scene, and only then the spatially congruent/incongruent object was presented. Conversely, the second experiment followed our simultaneous design (Mudrik et al., 2010). N300 effects arose in the sequential condition but not in the simultaneous condition, and the N400 component was smaller in the simultaneous than in the sequential condition. The authors concluded that earlier contextual influences are contingent on previously formed expectations about the forthcoming object, in sharp contrast to Mudrik et al.'s (2010) conclusions. Thus, under the premise that direct replications are the best way to establish the reliability of results (Cumming, 2014; Pashler & Harris, 2012), the first aim of our study was to provide a replication of the N300 congruity effects in a new group of subjects, and using more trials to obtain sensitivity to even earlier effects.

The second aim of this study was to examine the role of attention in contextual processing: is focused attention on the critical object necessary for detecting that it is incongruent with its context, or can such detection be performed without focused attention, possibly leading to attention being drawn to the critical object? Loftus and Mackworth's (1978) model of scene perception (see also Underwood, Templeman, Lamming, & Foulsham, 2008) proposed that low-level preattentive extraction of a scene's gist occurs before complete identification of the objects that compose it. Then, partial recognition of an unattended or non-fixated object may be sufficient to determine that it violates the gist of the scene and requires further inspection. Only at that stage does attention come into play, and it triggers an eye movement to the location of the incongruent object. In other words, the incongruent object is labeled as such before it is attended (Underwood et al., 2008). In line with this suggestion, several studies reported object categorization (Evans & Treisman, 2005; Kirchner & Thorpe, 2006; Li, VanRullen, Koch, & Perona, 2002; Potter, Staub, & O'Connor, 2004; Thorpe et al., 1996) as well as contextual processing (Brockmole & Henderson, 2006; Chun & Jiang, 1998, 1999; Hidalgo-Sotelo, Oliva, & Torralba, 2005; Oliva, Wolfe, & Arsenio, 2004), during dual tasks or with very short stimuli exposures, that seem to take place outside the focus of attention, or with very little attentional resources.

However, whether the *semantic relationship* that links an object to its context can also be processed in the absence of attention remains under debate. While several eye fixation studies reported earlier fixations on incongruent than on congruent objects (Friedman, 1979; Loftus & Mackworth, 1978; Underwood & Foulsham, 2006; Underwood, Foulsham, van Loon, Humphreys, & Bloyce, 2006), others observed only *prolonged* but not *earlier* fixations on incongruent objects (De Graef, Christiaens, & Dydewalle, 1990; Henderson, Pollatsek, & Rayner, 1989; Henderson, Weeks, & Hollingworth, 1999; Võ & Henderson, 2009, 2011), suggesting that attention is engaged by incongruent objects, but not drawn to them. Using binocular rivalry (for review, see Logothetis, Leopold, & Sheinberg, 1996), we found support for the latter view (Mudrik, Deouell, & Lamy, 2011).

To examine the role of spatial attention, in the current study we used exogenous cues (Posner, 1980) to direct subjects' attention towards or away from the location of a critical congruent/incongruent object and measured the effects of this manipulation on the electrophysiological markers of congruency processing (i.e., the N300/N400 component). We reasoned that if attention is needed for congruency processing, N300/N400 should be found for attended but not for unattended objects, and larger behavioral incongruency effects should be observed with attended than with unattended objects.

In summary, the aim of the current study was twofold: (a) to replicate the N300 effects found in our previous ERP study using a simultaneous object-scene presentation (Mudrik et al., 2010) in a new group of subjects, and thereby to provide critical support for contextual effects prior to full object identification in the face of conflicting data (Demiral et al., 2012) and (b) to directly manipulate attention in order to examine its influence on the

electrophysiological markers of congruency processing (i.e., the N300/N400 component).

2. Methods and materials

2.1. Participants

Twenty-three healthy students of the Hebrew University of Jerusalem, with reportedly normal or corrected-to-normal sight and no psychiatric or neurological history, volunteered to participate in the study for payment (~5$ per hour). Seven subjects were excluded from the analysis due to excessive eye movements or muscular artifacts, resulting in too few trials in each condition (fewer than 20). The remaining 16 subjects (nine males) were 20–30 years old (mean=23.8), 15 of them right handed. The experiment was approved by the ethics committee of the faculty of Social Sciences at the Hebrew University of Jerusalem, and informed consent was obtained after the experimental procedures were explained to the subjects.

2.2. Stimuli and apparatus

Subjects sat in a dimly lit room. The stimuli were presented on a 17″ CRT monitor with a 100-Hz refresh rate, using E-prime (version 1.2) software. They appeared on a black background at the center of the computer screen and subtended 6.96° (width) × 9.65° (height) of visual angle. The screen was located 100 cm away from subjects' eyes.

One hundred and sixty one pairs of colored pictures were used in the experiment (following Mudrik et al. (2010)). Each pair was based on a real-life scene taken from Internet sources and depicted a human action involving an object. To create the incongruent scenes, the original object was replaced with a different, unrelated object that was also taken from Internet sources. Images' congruency was rated in a pretest experiment on a scale of 0 (not unusual at all) to 4 (very unusual) (congruent images: $M=0.22$, SD=0.39, incongruent images: $M=3.22$, SD=0.74). The pictures' luminance and contrast levels were digitally equated using Adobe Photoshop software. Low-level differences in saliency, chromaticity and spatial frequency were tested using two different perceptual models (Itti & Koch, 2000; Neumann & Gegenfurtner, 2006). No such differences were found (for details on both low-level feature analysis and congruency rating, as well as on a followup experiment conducted to control for the fact that only incongruent objects were pasted on the scenes, see Mudrik et al., 2010).

In addition, valid and invalid cues were tailored to each stimulus of the original stimuli database using Adobe Photoshop software. Cues were yellow rectangles (RGB values: 255, 255, 0) that were rotated at different angles. Valid cues were drawn so that the rectangle surrounded the largest object in the pair (congruent/incongruent). Invalid cues were created by flipping the valid cues horizontally, so that the valid and invalid cues appeared at the same eccentricity (i.e., equally distant from image center), and in opposite orientations (Fig. 1). Eighty-one cues were tailored to incongruent objects, and 80 to congruent objects.

2.3. Procedure

The experiment included 322 trials, half of which were valid-cue (attended) trials, and half invalid-cue (unattended) trials. In both the attended and unattended conditions, half of the trials were congruent and the other half incongruent. Conditions of attention and congruency were randomly intermixed, with the constraint that the same trial type (attended/unattended, congruent/incongruent) was never presented in four consecutive trials. The session began with five practice trials, where subjects performed the task in the experimenter's presence, to ensure that they followed the instructions correctly.

At the beginning of each trial, the valid/invalid pre-cue was presented for 100 ms and was immediately followed by the critical stimulus (congruent/incongruent scene) that appeared for 200 ms (Fig. 1). After stimulus presentation, a question immediately appeared: "How many hands were used by the person in the picture to perform the action?" The question was aimed at having subjects focus on the action performed by the person in the image (that involved the congruent/incongruent object), without explicitly asking about the congruency of the objects. Accordingly, congruency-related ERPs were held to index spontaneous rather than task-induced congruency processing. Subjects were asked to type their responses as quickly as possible, using the keys 0, 1 and 2. If they did not respond after 5 s, the question disappeared from the screen. Trial presentation was self-paced.

2.4. ERP methods

2.4.1. ERP recording

The EEG was recorded using an Active 2 system (BioSemi, the Netherlands) from 64 electrodes distributed based on the extended 10–20 system (Fig. 2) connected to a cap, and 7 external electrodes. Four of the external electrodes recorded the EOG: two located at the outer canthi of the right and left eyes and two above and below the center of the right eye. Two external electrodes were located on the mastoids, and one electrode was placed on the tip of the nose. All electrodes were referenced during recording to a common-mode signal (CMS) electrode between POz and PO3. The EEG was continuously sampled at 1024 Hz and stored for offline analysis.

2.4.2. ERP analysis

ERP analysis was conducted using the "Brain Vision Analyzer" software (Brain Products, Germany). For consistency with several studies addressing the N400 effects, data from all channels were referenced offline to the average of the mastoid channels. The data were digitally high-pass filtered at 0.1 Hz (24 dB/octave) to remove slow drifts, using a Butterworth zero-shift filter. Bipolar EOG channels were calculated by subtracting the left from the right horizontal EOG channel, and the inferior from the superior vertical EOG channels. This bipolar derivation accentuates horizontal and vertical eye movement artifacts, respectively, which serves the artifact detection procedure described below. The signal was cleaned of blink artifacts using Independent Component Analysis (ICA) (Jung et al. 2000). Segments contaminated by other artifacts were detected as amplitudes exceeding ± 100 μV, differences beyond 100 μV within a 200 ms interval, or activity below 0.5 μV for over 100 ms (the latter was never found), in any channel, including the bipolar EOG channels. Segments including such artifacts were discarded from further analysis (leaving an average number of 69 with a range of 45–80 trials in each condition).

The EEG was segmented into 1000-ms long epochs starting 100 ms prior to the scene onset, and the segments were averaged separately for each condition (congruent / incongruent). The average waveforms were low-pass filtered using a Butterworth zero-shift filter with a cutoff of 30 Hz, and the baseline was adjusted by subtracting the mean amplitude of the pre-stimulus period of each ERP from all the data points in the segment. Difference waves were computed by subtracting the response to the congruent trials from the response to the incongruent trials.

The effect of context violations was assessed in two ways. First, we used a 'time of interest' approach (following the findings of Ganis and Kutas (2003), Mudrik et al. (2010)), and analyzed the average amplitude within two time windows: 200–300 ms and 300–500 ms. Differences were assessed for each time window using 4-way ANOVAs with Attention (Attended, Unattended), Congruity (Congruent, Incongruent), Region (Frontal, Central, Occipitoparietal; see Fig. 2) and Laterality (Left, Midline, Right) as factors. Second, we investigated the onset latency of the early incongruency effect by conducting point-by-point t-tests on the difference between context-congruent and context-incongruent conditions within a 100–400 ms time window, irrespective of attention, after down-sampling the data to 256 Hz. To determine epochs of significant difference without inflating the probability of error due to multiple comparisons, we used the cluster-based non-parametric permutation statistical test described in Maris and Oostenveld (2007), including all 9 regions. Under the null hypothesis of no difference between the conditions, in each iteration of the procedure, the labels of the conditions were switched in some of the subjects. This was repeated for all possible permutations ($N=2\backslash$widehat15, considering a 2-tailed test). In each permutation, sequential time points (clusters) exceeding the critical t value corresponding to $p<0.05$ (uncorrected) were identified in all 9 regions. The t values in each cluster were summed, and the largest of these cluster sums across all regions was noted. Finally, the sum of t values in the clusters identified in the original data were considered significant if they were larger than 95% of the maximal sums obtained under the null hypothesis. For analyses designed to examine spatial distribution differences, we normalized the amplitudes using the vector scaling method as described in McCarthy and Woods (1985). As recommended by Picton et al. (2000) (p. 147), results from both normalized and non-normalized analyses are presented, as scaling is necessary to evaluate possible differences in spatial distributions, but obscures the effects of experimental manipulations. Greenhouse-Geisser correction was used where appropriate. The uncorrected degrees of freedom are reported along with the Greenhouse-Geisser epsilon values (Picton et al., 2000).

2.5. Follow-up behavioral experiment

In the main experiment, subjects' task was to enumerate the number of hands used by the person in the image, while their attention was directed towards or away from the congruent/incongruent object. Since in ~20% of the trials (33 pairs out of 161) the object and the hands appeared at different locations, this behavioral measure turned out not to be ideal for assessing the effectiveness of our attentional manipulation. To directly test the manipulation we conducted a follow-up experiment with a simple perceptual task, namely discriminating between the letters T and L (T/L discrimination task; see Braun & Julesz, 1998; Lee, Koch, & Braun, 1999). We presented each of 12 naïve subjects with the same procedure as in the main experiment, but in one third of the trials (henceforth, T/L trials) a white letter – either T or L (1.15 × 0.92°) – was presented instead of the scene for 200 ms, at the place where the critical object was supposed to appear. Then, a question was presented, referring either to the letter's identity (in T/L trials) or to the number of hands (in the rest of the trials). To keep all other conditions equal to the original experiment, in 20% of the valid trials the cue did not refer to the hands' location. We hypothesized that if the attentional manipulation was indeed effective, T/L discrimination should be better when the pre-cue was valid than when it was invalid.

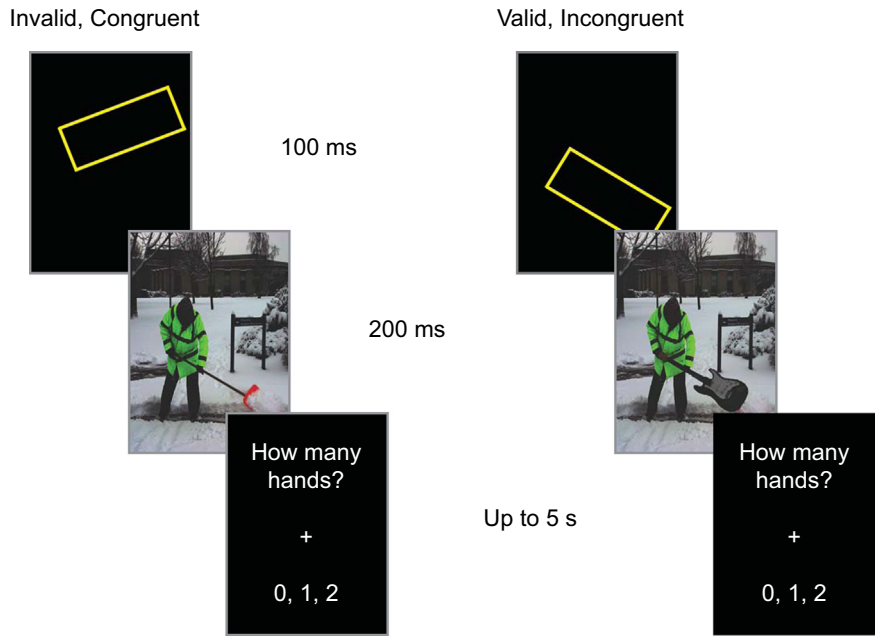

Fig. 1. Experimental procedure; a valid or an invalid pre-cue was presented for 100 ms, followed by a congruent or an incongruent scene for 200 ms. After stimulus presentation, subjects were asked how many hands were used by the person in the picture to perform the action. On the left, an invalid congruent trial (a man shoveling snow with a shovel, but the pre-cue is not around the future location of the shovel). On the right, a valid incongruent trial (a man shoveling snow with an electric guitar, and the pre-cue location is consistent with the location of the guitar).

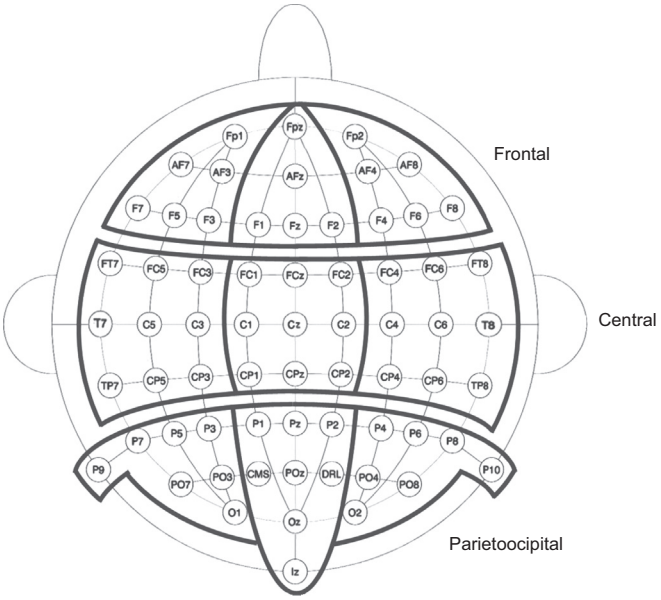

Fig. 2. Electrode array and division of the electrodes into regions.

3. Results

3.1. Behavioral results

Subjects reported whether the object in the image was manipulated by zero, one, or two hands. Their mean RTs (on correct trials) and accuracy scores were analyzed in 2-way, within-subject, Attention × Congruity ANOVAs. The main effect of congruity was significant on both the RTs and the accuracy measures. Responses were faster for congruent scenes ($M=859$ ms, SD=165 ms) than for incongruent scenes ($M=903$ ms, SD=181 ms; $F(1,15)=18.04$, $p=0.001$) and accuracy was higher for congruent scenes ($M=0.86$, SD=0.08) than for incongruent ones ($M=0.82$, SD=0.09; $F(1,15)=8.68$, $p=0.01$). The main effect of attention on accuracy was significant, but in the unexpected direction: performance in the unattended condition ($M=0.85$, SD=0.08) was slightly better than in the attended one ($M=0.83$, SD=0.09; $F(1,15)=5.8$, $p=0.029$). No difference was found in reaction times between attended ($M=888$ ms, SD=175 ms) and unattended trials ($M=873$ ms, SD=175 ms; $F(1,15)=1.84$, $p=0.195$). No interaction was found between attention and congruity for either RTs or accuracy.

The unexpected main effect of attention may have resulted from the fact that while the valid and invalid cues were created according to the location of the critical object, the task itself pertained to the number of hands used by the person in the image. Since in about 20% of the trials object and hands appeared at different locations (see Section 2.5), the valid cues on those trials were in fact invalid with regard to the task. However, the main findings were replicated even when such trials were excluded, providing no support for this post-hoc interpretation: responses were again faster for congruent scenes ($M=843$ ms, SD=164 ms) than for incongruent scenes ($M=888$ ms, SD=186 ms; $F(1,15)=8.20$, $p=0.012$) and in unattended trials ($M=852$ ms, SD=179 ms) than in attended trials ($M=879$ ms, SD=174 ms; $F(1,15)=12.12$, $p=0.003$). There were no significant effects of attention ($F(1,15)<1$, $p=0.73$) or congruity ($F(1,15)=2.03$, $p=0.17$) on accuracy. This analysis, however, does not exclude the possibility that subjects interpreted the cues as being invalid because of their lower predictive value with regard to the hands.

3.2. ERP results

Overall, incongruent images elicited prolonged frontocentral negativity both for the attended and the unattended conditions (Figs. 3 and 4). The subsequent analyses were done to (a) compare the incongruency effect in each attentional condition and (b) determine the latency of this congruency-related negativity difference irrespective of attention (i.e., whether it emerges only in the N400 time window, or earlier, in the N300 window).

3.2.1. "Spatio-temporal region of interest" analyses

3.2.1.1. N300 time window. The amplitudes of the ERPs were averaged within the 200–300 ms time window separately for the three areas and three lateralities (see ERP methods, Section 2.4.2). A 4-way Attention × Congruity × Region × Laterality ANOVA revealed a main effect of Congruity ($F(1,15)=8.2$, $p=0.012$), of Region ($F(2,30)=52.05$, $p<0.0001$, $\varepsilon=0.528$) and of Laterality ($F(2,30)=13.87$, $p<0.0001$, $\varepsilon=0.96$). No interaction involving Attention and Congruity was found. A three-way interaction of Attention × Region × Laterality was observed ($F(4,60)=3.56$, $p=0.032$, $\varepsilon=0.59$), yet when inspecting the source of this interaction, none of the two-way ANOVAs conducted either between Attention and Laterality at each region or between Attention and Region at each laterality level yielded any significant interaction (all $p>0.2$). Notably, the three-way interaction did not hold after vector-scaling normalization (McCarthy & Wood, 1985), suggesting that attention did not alter the distribution across the scalp.

3.2.1.2. N400 time window. The amplitudes of the ERPs were averaged within the 300–500 ms time window. A 4-way Attention x Congruity × Region × Laterality ANOVA revealed a main effect of Congruity ($F(1,15)=14.03$, $p=0.002$) and of Region ($F(2,30)=20.51$, $p<0.0001$, $\varepsilon=0.532$). The two-way interactions between Congruity and Region and Congruity and Laterality were significant ($F(2,30)=10.41$, $p=0.003$, $\varepsilon=0.611$, and $F(2,30)=4.44$, $p=0.025$, $\varepsilon=0.894$, respectively), but not after vector-scaling normalization, thus precluding conclusions about differences in distribution. Post-hoc contrasts indicated that the congruity effect (Incongruent–Congruent) was smaller in the parieto-occipital regions (−0.38 μV) compared with the frontal (−1.17 μV, $t(15)=3.21$, $p=0.006$) and the central (−1.01 μV, $t(15)=4.41$, $p=0.001$) regions. In line with these findings, one sample *t*-tests confirmed that the congruity effect was different from zero in the frontal ($t(15)=4.14$, $p=0.001$) and central regions ($t(15)=4.18$, $p=0.001$), but not in parieto-occipital regions ($t(15)=1.67$, $p=0.115$).

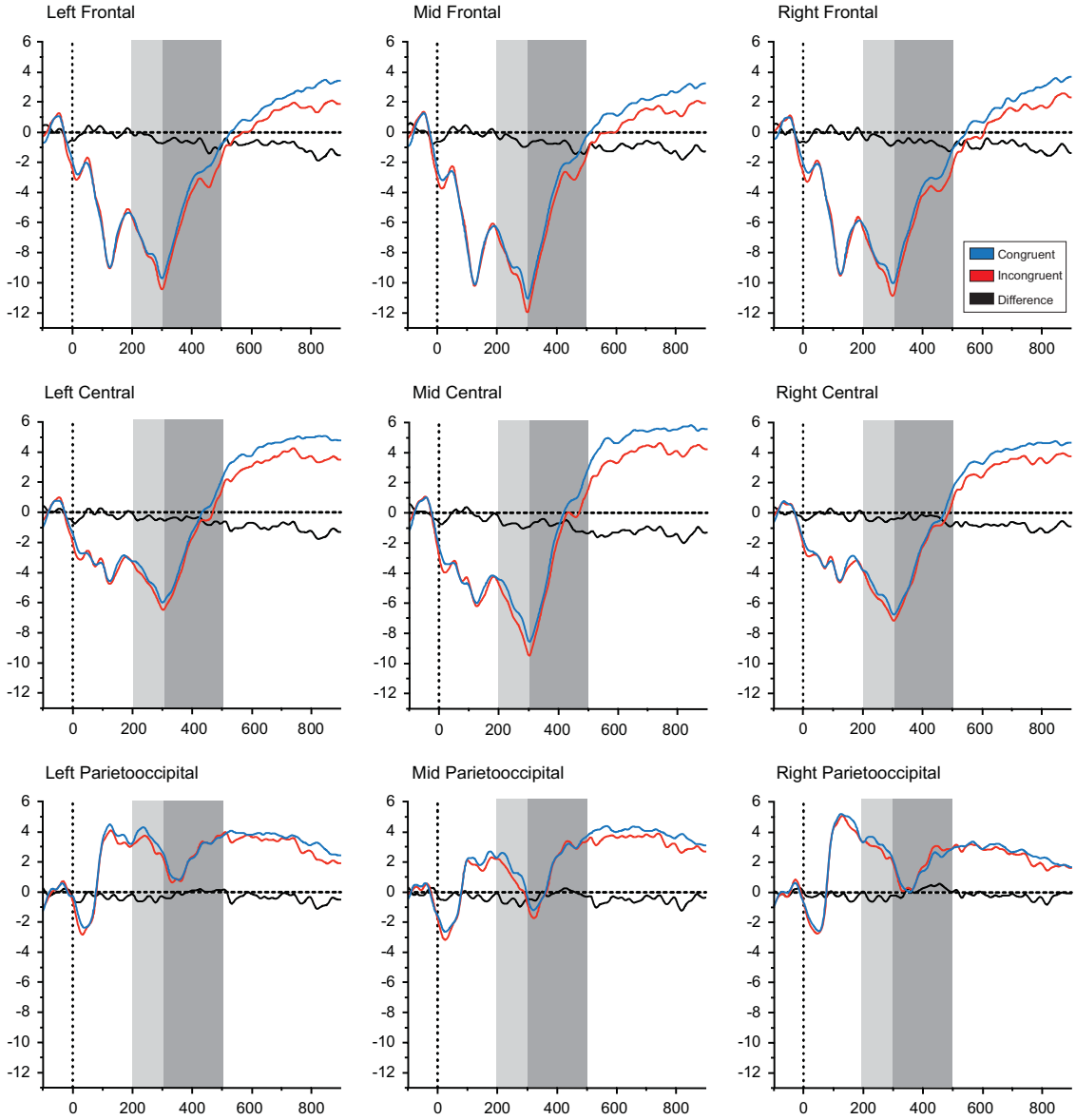

Fig. 3. Response to congruent (blue) and incongruent (red) images in *attended* trials, and difference waves (black waveforms below). Light gray rectangles denote the 200–300 ms time-window, dark gray denotes the 300–500 ms time-window. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

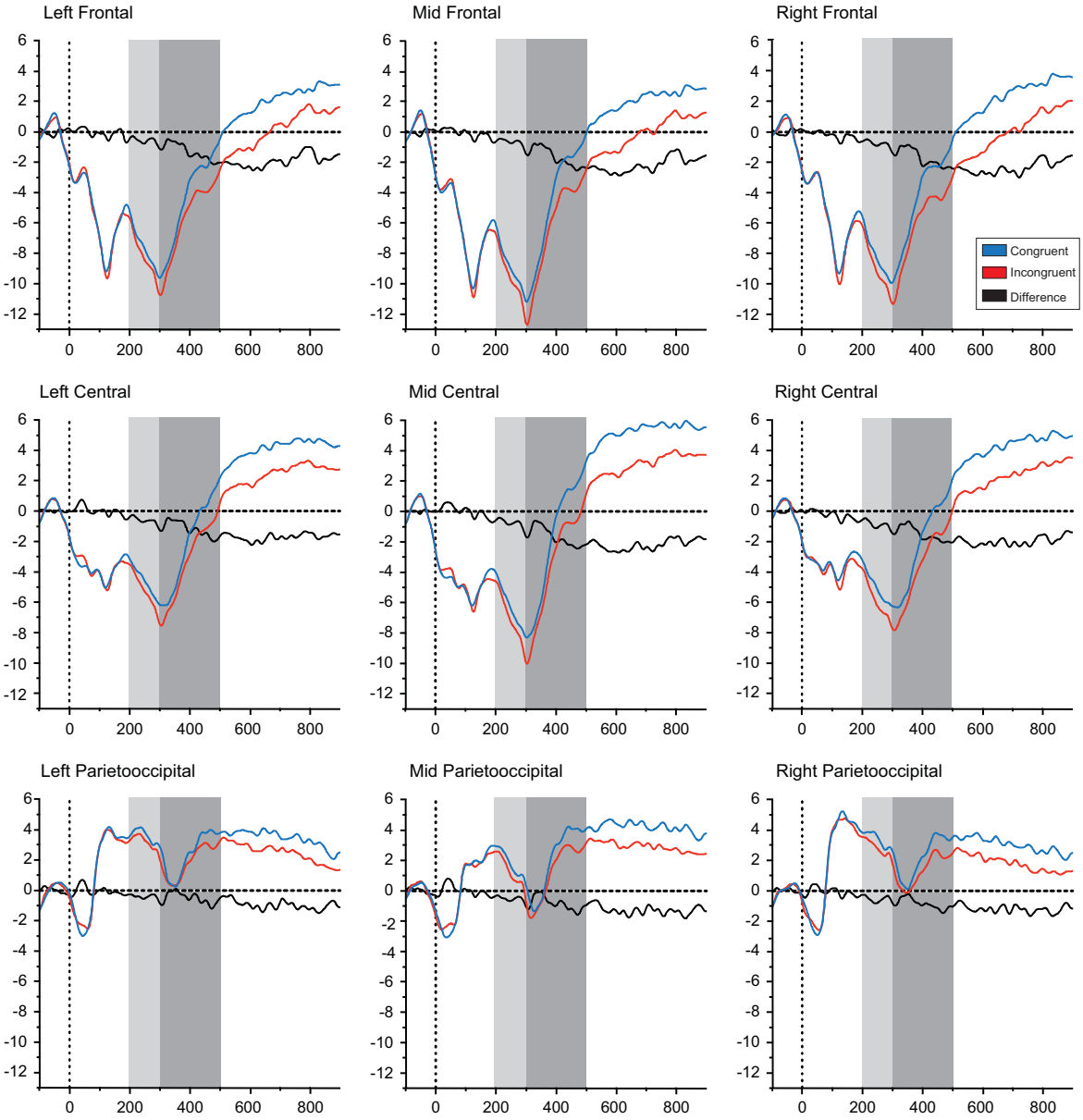

Fig. 4. Response to congruent (blue) and incongruent (red) images in *unattended* trials, and difference waves (black waveforms below). Light gray rectangles denote the 200–300 ms time-window, dark gray denotes the 300–500 ms time-window. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

3.2.2. *Latency analysis for congruent vs. incongruent trials*

To assess the point in time where the congruity effect started to emerge (see Fig. 5 for scalp distribution over time), a cluster-based non-parametric permutation statistical test (Maris & Oostenveld, 2007) was conducted (see Section 2). The results show that in the middle central region, the congruent and incongruent waveforms started to differ significantly as early as 209 ms post stimulus presentation, and continued to show a difference for ~110 ms (Fig. 6). In other regions, although earlier effects were found using uncorrected point-by-point *t*-tests, the early effects did not survive the cluster-based permutation analysis.

3.2.3. *ERPs and behavior analysis*

After identifying the N300 and N400 components evoked by incongruent scenes, we conducted a post-hoc analysis to investigate the relations between the amplitudes of these components and subjects reaction times in reporting the number of hands used by the person in the scene. For each subject, we computed (a) the mean voltage of the difference waves between waveforms elicited by congruent and incongruent scenes, for each region, during the 200–300 ms and 300–500 ms time windows and (b) the difference between reaction times following congruent and incongruent scenes. Then, we computed Pearson's *R* correlations between ERPs amplitudes and RT differences for each region in each time window. To determine significance, we conducted a permutation analysis in which the assignment of the behavioral data to subjects was randomly permuted 10,000 times while the assignment of the EEG data was unaltered. In all frontal regions and the left central one, positive correlations were found ($r=0.599$, $p=0.014$, $r=0.677$, $p=0.004$, $r=0.554$, $p=0.026$, $r=0.567$, $p=0.022$ for left, middle and right frontal regions and the left central one, respectively), such

that the larger the incongruency negativity, the smaller the difference in reaction times following congruent and incongruent scenes. A similar trend was found for the middle central region ($r=0.511$, $p=0.043$), though it was only at the 93 percentile of the bootstrapping analysis. None of the correlations in the 300–500 ms time window were significant.

3.3. Follow-up behavioral experiment results

The follow-up behavioral experiment was designed to test the effectiveness of the attentional manipulation, using a simpler, and more direct, perceptual task (T/L discrimination), known to be sensitive to attention. Since in valid trials the T/L stimulus always appeared within the rectangular cue (as opposed to the original hands task, where the hands were on the same side as the cue but not always within the rectangle), we expected this task to be more sensitive to the attentional manipulation. Thus, if the cue was effective in drawing attention, we expected faster and more accurate responses in valid than invalid trials. For comparison, we also tested the original task in which subjects had to decide how many hands (0–2) were used. Participants were required to perform the T/L discrimination task on one third of the trials, and to perform the original task from the main experiment on the remaining trials.

3.3.1. T/L discrimination task

There was no attention effect on either accuracy ($M=0.67$, SD=0.09 vs. $M=0.65$, SD=0.10 for attended vs. unattended trials, respectively; $t(11)=1.09$, $p=0.30$) or RTs (M =721 ms, SD= 167 ms; vs. $M=718$ ms, SD=168 ms, respectively; $t(11)=0.20$, $p=0.85$), suggesting that the attentional manipulation was ineffective in diverting subjects' attention towards or away from the critical object, in an observable way.

3.3.2. Original one- vs. two-hands task

Subjects' mean accuracy scores and RTs (on correct trials) were analyzed in 2-way Attention × Congruity ANOVAs. Subjects were significantly more accurate for congruent than for incongruent scenes ($M=0.85$, SD=0.07 vs. $M=0.80$, SD=0.05, respectively, $F(1,11)=5.37$, $p=0.04$), and marginally faster on congruent than on incongruent trials ($M=1060$ ms, SD=296 ms vs. $M=1098$ ms, SD=327 ms, respectively, $F(1,11)=3.04$, $p=0.11$). No other effect was significant. Thus, in the follow-up experiment, we did not replicate the unexpected finding of better accuracy on unattended than on attended trials.

4. Discussion

Contextual regularities play a key role in scene perception and interpretation (Bar, 2004; Biederman et al., 1982), but the stage at which their influence takes effect and the role of attention in these processes, are still under debate. The findings of this study provide a clear replication of our previous results (Mudrik et al., 2010; see also Võ & Wolfe, 2013) showing even earlier contextual processing, probably because using twice as many trials considerably improved signal-to-noise ratio. Incongruent scenes evoked ongoing frontocentral negativity at both the N300 (as early as ~210 ms in the present study) and the N400 time windows. Importantly, these differences emerged during simultaneous processing of scene and object. Thus, our results reflect real-time processing of the semantic relationship between a scene and its constituents, rather than the effect of confirming or violating expectations set prior to object presentation. We found no effect of attention. However, as discussed below, it seems that our spatial manipulation of attention was ineffective.

4.1. Stage of congruency processing

Our findings validate our previous results by showing them to be replicable even with shorter stimulus presentations, and therefore lend strong support for matching models of contextual influences on scene perception (e.g., Bar, 2004; Bar & Aminoff, 2003; Bar & Ullman, 1996; Kosslyn, 1994).

Matching models hold that a coarse, low spatial-frequency representation of a visual scene suffices to activate an experience-based prediction about the scene's context in the parahippocampal cortex (PHC) (Bar, 2003, 2004). This coarse representation is projected early and rapidly from the visual cortex to the PHC and orbitofrontal cortex (OFC), presumably through the magnocellular pathway (Graboi & Lisman, 2003; Merigan & Maunsell, 1993), triggering a set of relevant pre-existing schemata. This results in top-down activation of a set of schema-congruent object representations – arguably in the Inferior Temporal Cortex (ITC) (see again Bar, 2003, 2004) – so that the level of sensitization of each representation depends on the strength of its association with the specific context. Then, the upcoming visual information about the scene's constituents is matched with the already activated schema-congruent representations in the ITC, until a single identity for each object is reliably selected. When the upcoming visual information does not match the pre-activated schemata, object identification is impeded. Thus, object incongruency involves lack of correspondence between the object's perceptual features, such as its spatial frequency (Bar et al., 2006) or color (Goffaux et al., 2005), and semantic knowledge about the expected scene constituents.

The N300 observed in our study is thought to stem from a large collection of areas around the Middle Occipitotemporal Gyrus (MOG), including the ITC (Schendan & Maher, 2008) and to index the difficulty of object-selection matching processes, with greater matching difficulty being associated with higher amplitudes (Ganis & Kutas, 2003; Schendan & Maher, 2008). Accordingly, scrambled and pseudo-objects with unknown perceptual structures induce a greater N300 (Folstein et al., 2008; Holcomb & McPherson, 1994; Schendan & Kutas, 2002), much like related vs. unrelated objects (Barrett & Rugg, 1990; McPherson & Holcomb, 1999) and new vs. repeated objects in memory tasks (Henson et al., 2004; Schendan & Kutas, 2003, 2007). Thus, the N300 component we found is in line with the claim that contextual relations are processed during perceptual stages of scene processing, probably prior to objects' full identification.

The argument that full object identification is achieved in the 200–300 ms time window may seem at odds with the fact that faces elicit the N170 effect (Bentin, Allison, Puce, Perez, & McCarthy, 1996). However, N170 does not signal complete face recognition but probably the detection of physiognomic features, whereas ERP indices of face identification occur later than 200 ms (Bentin, Deouell, & Soroker, 1999; Gosling & Eimer, 2011; Zheng, Mondloch, & Segalowitz, 2012). The argument may also seem incompatible with the fact that target vs. non-target scenes show differential effects as early as 150 ms post scene identification (Codispoti, Ferrari, Junghöfer, & Schupp, 2006; Goffaux et al., 2005; Thorpe et al., 1996), and that information on object categories may be detected in EEG or MEG as early as 80 ms post stimulus presentation (e.g., Carlson, Tovar, Alink, & Kriegeskorte, 2013; Hung, Kreiman, Poggio, & DiCarlo, 2005; VanRullen & Thorpe, 2001). However, early category-specific activations by scenes and objects may reflect mostly low-level differences between these categories: extracting a scene's gist is held to rely on low-level global features, like statistical properties of object sets (Ariely, 2001; Chong & Treisman, 2003; Fiser & Aslin, 2001), spatial distribution of colored regions (Goffaux et al., 2005; Oliva & Schyns, 2000), or the mean of global image features at a coarse spatial resolution (Oliva & Torralba, 2001, 2006; Torralba, Oliva, Castelhano, & Henderson, 2006). Object

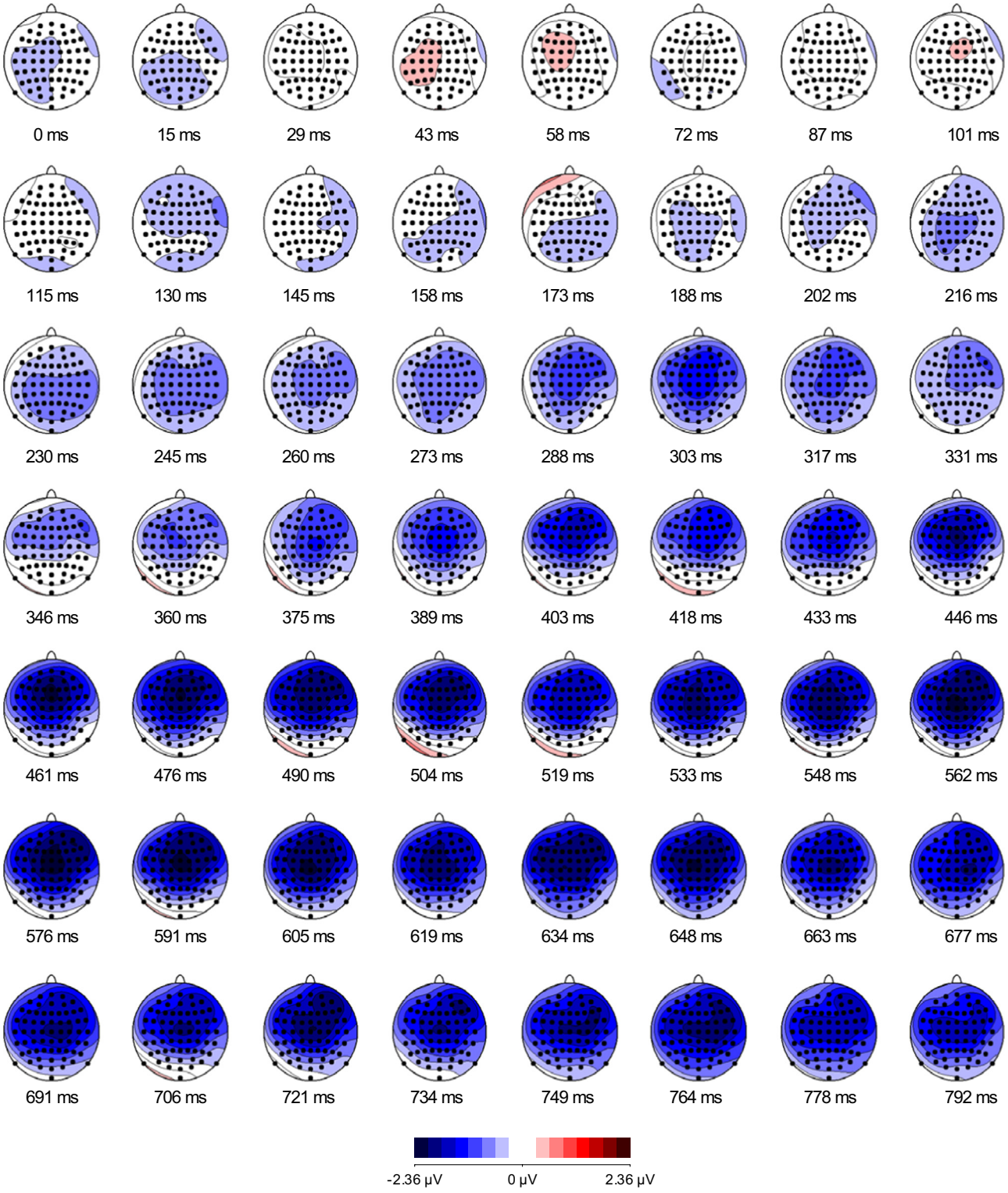

Fig. 5. Scalp distribution maps of the incongruency effect (incongruent–congruent) over time.

categorization probably also starts with crude classification that rests on low-level features, which is followed by processing of more complex features, aided by contextual and semantic expectations (Bar, 2004). Indeed, different object categories – like animals and vehicles (VanRullen & Thorpe, 2001) – have very different low-level features. Thus, while differences between scenes and between objects may be evident in neural activity prior to 200 ms, they do not contradict the conjecture that object categorization (partly based on these early differences) continues within the 200–300 ms window, where we show that it is affected by context.

Further support for this conjecture comes from a recent study that compared ERPs for objects embedded in natural scenes vs. phase-randomized backgrounds (Sun et al., 2011). Subjects performed an animal/non-animal go/no-go categorization task. Behavioral performance was faster and more accurate for objects appearing in their natural scenes. Crucially, onset latency of the differential activity

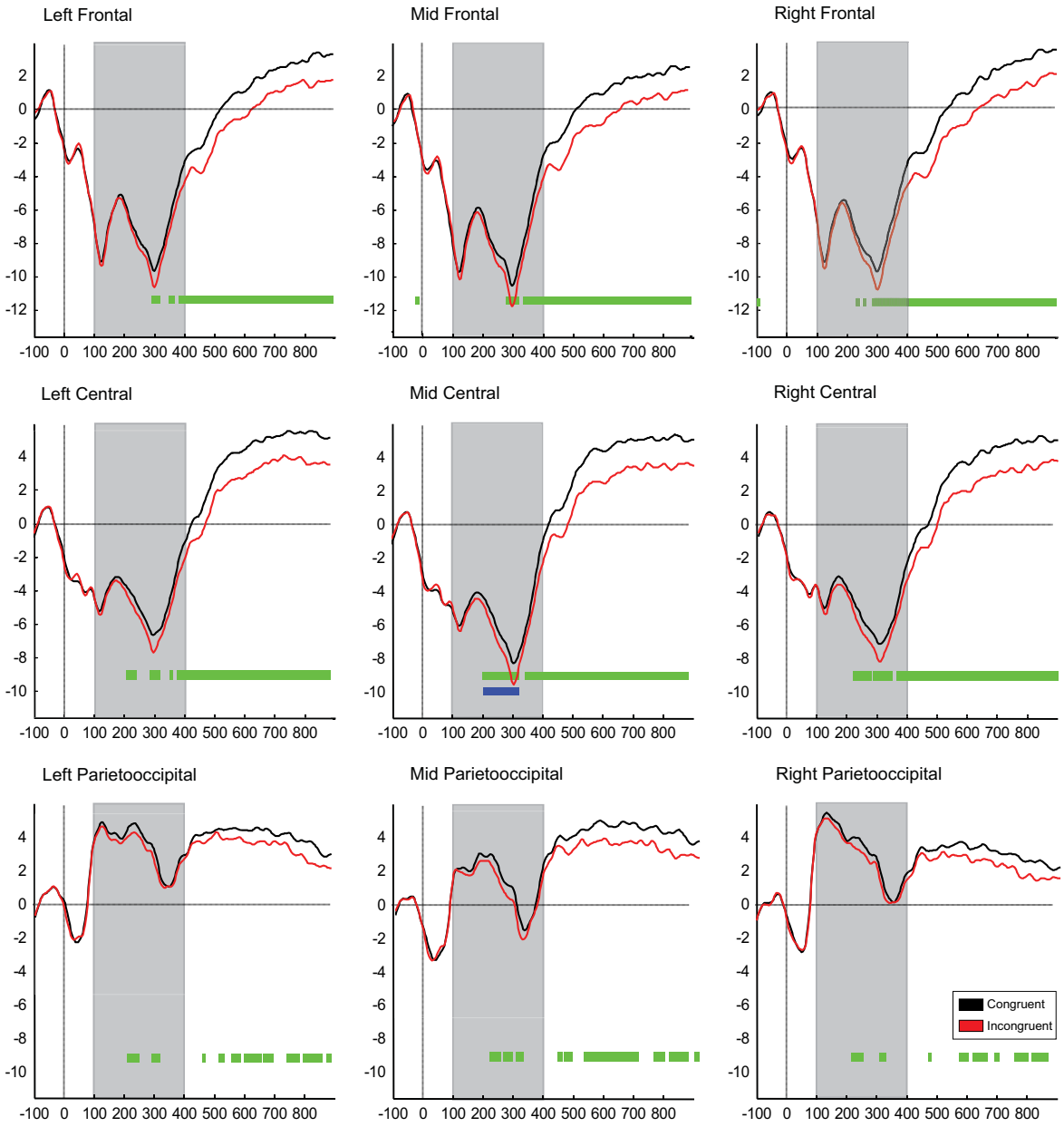

Fig. 6. Congruent and incongruent waveforms (irrespective of attention) for each region. Green lines on the bottom represent time points in which the waveforms differ, according to an uncorrected *t*-test ($p < 0.05$) performed point-by-point *t*-test on all data. Blue lines represent time points during the 100–400 ms time-window (shaded) which were significant after correction according to the cluster-based non-parametric permutation statistical procedure. In midline central electrodes, the earliest significant (corrected) differences are found at ~210 ms. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

between animal and vehicle items found in frontal electrodes was delayed by about 20 ms in the phase-randomized condition. These findings suggest that scene context indeed facilitates object processing prior to its identification.

It is therefore somewhat surprising that the post-hoc correlation analysis revealed that the reaction time cost of incongruent scenes was in fact smaller in subjects for whom the earlier ERP incongruency effect was larger, suggesting that having a stronger N300 incongruency effect may reduce the deleterious effects of incongruency on performance in an irrelevant task. This might be conjecturally explained by noticing that the earlier incongruency effect occurred between 200 ms and 300 ms, some 600 ms before the response. Possibly, subjects who fail to process the incongruency at earlier stages, face the conceptual difficulty of deciphering the incongruent scene at later stages, when they also need to choose the appropriate response regarding how many hands are used. It would have been informative to know whether slow responses are associated also with longer onset latencies of the incongruency effect. Unfortunately determining onsets on a single subject basis as required for such correlation was hindered due to signal-to-noise issues and the fact that the congruency effect is a slow wave rather than a component with a clear morphology (which would allow for example to determine the latency at some percent from the peak). As a post-hoc explanation, this hypothesis is merely suggestive at this point, and calls for more research.

4.2. Simultaneous vs. sequential presentation of the scene and the object

As noted in the Introduction, the possibility that early contextual effects occur even when the context does not temporally precede the critical object was recently contested by Demiral et al. (2012).

These authors reported earlier congruity effects only when the scene was presented prior to the critical object (sequential condition), but not when the two were presented simultaneously, while later effects were found for both simultaneous and sequential presentations. Considering that we now replicated our results with simultaneous presentations and found even earlier effects, how can the failure to find such effects in the simultaneous condition of Demiral et al. be explained?

A potentially important difference between their study and ours is that we examined semantic violations (i.e., manipulating the probability of an object to appear in a given scene), whereas they investigated "syntactic" violations (i.e., manipulating the probability of an object to appear at a specific location within a semantically congruent scene). Semantic violations may evoke greater identification difficulties than syntactic violations, since according to matching models, a scene's context pre-activates all schema-congruent representations. Thus, the search for the identity of a *syntactically* incongruent object (e.g., a bus in the sky) would still be facilitated to some extent by the scene's congruent context, even if the object is not in its typical position within the scene, and the mismatch between incoming visual information and preactivated representations should be weaker (and possibly be detected later) than for semantic violations. Corroboration for this argument comes from a recent study that directly compared semantic and syntactic violations (with sequential scene-object presentation); while the former elicited an N300 effect, the latter did not (Võ & Wolfe, 2013).

Demiral et al.'s failure to observe early congruency effects in the simultaneous condition may have also stemmed from the relatively long cue-target onset asynchrony they used. The pre-cue was 725–1075 ms long whereas it was 100-ms long in the current study and 200-ms long in Mudrik et al.'s (2010) study. The long cue-target onset asynchrony in Demiral el al.'s study may have placed the subject in an Inhibition Of Return (IOR; Posner & Cohen, 1984) situation: after attention is exogenously directed to a peripheral location, observers tend to shift their attention *away from that location*, arguably to allow for further exploration of the visual field (Klein, 2000; Klein & MacInnes, 1999). As a result, processing of a target at the cued vs. uncued location is impeded, as reflected by slower RTs (Bichot & Schall, 2002; Klein, 1988; Posner & Cohen, 1984) and ocular saccades away from the cued location (e.g., Hooge & Frens, 2000; Hooge, Over, van Wezel, & Frens, 2005; Posner, Rafal, Choate, & Vaughan, 1985). While IOR was originally found in visual search tasks, it was also reported during free viewing (Bays & Husain, 2012; Hooge et al., 2005). Importantly, intracranial recordings in primates showed that IOR attenuates neural responses to cued targets in visual areas (Dorris, Klein, Everling, & Munoz, 2002; Fecteau & Munoz, 2005; Mirpour, Arcizet, Ong, & Bisley, 2009). Therefore, the N400 attenuation and lack of early effects in the simultaneous condition of Demiral et al.'s (2012) study may reflect degraded processing of the critical stimuli due to IOR. Their simultaneous condition was more likely to suffer from IOR than was their sequential condition, because in the latter, cue-target asynchrony was shorter (600 ms), and more importantly, the onset of the object served as a second exogenous cue, which could reorient subjects' attention to the object and mitigate IOR effects. By contrast, in the simultaneous condition, no second cue could draw subjects' attention back to the object's location.

4.3. The role of attention in congruency processing

We found no difference in congruency effects for validly or invalidly cued objects. Because no attentional effect on performance was found in either the main task or in the simpler T/L perceptual task known to be sensitive to attention (Lee et al., 1999; Li et al., 2002), it is likely that our attentional manipulation was ineffective. Thus, we cannot draw any clear conclusions about the involvement of attention in congruency processing.

Three factors might have rendered our attention manipulation ineffective. First, the predictive value of the cues was low because the task pertained to the hands rather than to the critical objects. Accordingly, 60% of the cues were in fact invalid with regard to the task (see Section 2.5). This might have led subjects to either ignore the cues or direct attention away from them. Indeed, attentional attraction by salient stimuli is often contingent on task demands, so that involuntary shifts of attention towards a stimulus occur only when that stimulus is relevant to subjects' explicit or implicit set of perceptual goals (Corbetta & Shulman, 2002; Folk, Leber, & Egeth, 2002; Folk, Remington, & Johnston, 1992; Serences et al., 2005). Second, the sizes and locations of our cues may have hindered their effectiveness. Typically, exogenous cues are small and located at the periphery of the visual field (Eriksen & Hoffman, 1973; Muller & Rabbitt, 1989; Nakayama & Mackeben, 1989; Posner, 1980). In our study, since the attended region had to encompass the space occupied by the object, cues sometimes occupied almost half of the visual field (21% of the cues), or were centrally located (29% of the cues). Note however that in fact, a recent study that used much smaller cues also did not find any effect of focused attention on congruency processing (Munneke, Brentari, & Peelen, 2013). Third, subjects' reaction times in the task were relatively long (860 ms and 900 ms for congruent and incongruent scenes, respectively), presumably because the short presentation of the scenes rendered the relatively high-level task difficult to perform. Subjects apparently took their time in answering the question, possibly because our emphasis on speed was not strong enough, which may have diluted the effect of attention.

4.4. Conclusions

In this study we obtained the earliest reported electrophysiological markers for contextual processing during simultaneous scene and object processing. This finding puts an upper limit to the earliest effects of incongruency (which might start even earlier) and substantially strengthens matching models that argue for early contextual influences on object identification (e.g., Bar, 2004; Bar & Ullman, 1996; Kosslyn, 1994). Contrary to recent claims, it shows that such effects in the 200–300 ms time window can be found when scene and object are presented simultaneously rather than sequentially (Demiral et al., 2012), hereby validating and extending previous results from our lab (Mudrik et al., 2010). Our results thus present a challenge to functional isolation models, which allow for contextual influences only during later, semantic stages of scene processing. Together with mounting evidence for the crucial role of top-down mechanisms during early perceptual stages (e.g., Barcelo, Suwazono, & Knight, 2000; Hopfinger, Buonocore, & Mangun, 2000; Humphreys, Riddoch, & Price, 1997; Mechelli, Price, Friston, & Ishai, 2004; Miyashita & Hayashi, 2000; Pascual-Leone & Walsh, 2001; Ranganath, DeGutis, & D'Esposito, 2004), our findings contest traditional dichotomies between perception and cognition and call for a more integrative approach that allows for ongoing interactions between perceptual "low-level" and cognitive, knowledge-based processing.

Acknowledgments

The study was supported by a grant from the National Institute of Psychobiology in Israel, founded by the Charles E. Smith foundation. L.M. was supported by the Human Frontiers Science Program, and the Weizmann Institute of Science – National Postdoctoral Award Program for Advancing Women in Science.

References

Antes, J. R., Penland, J. G., & Metzger, R. L. (1981). Processing global information in briefly presented pictures. *Psychological Research—Psychologische Forschung*, *43*(3), 277–292.

Ariely, D. (2001). Seeing sets: Representation by statistical properties. *Psychological Science*, *12*(2), 157–162.

Bar, M. (2003). A cortical mechanism for triggering top-down facilitation in visual object recognition. *Journal of Cognitive Neuroscience*, *15*(4), 600–609.

Bar, M. (2004). Visual objects in context. *Nature Reviews Neuroscience*, *5*(8), 617–629.

Bar, M., & Aminoff, E. (2003). Cortical analysis of visual context. *Neuron*, *38*(2), 347–358.

Bar, M., Kassam, K. S., Ghuman, A. S., Boshyan, J., Schmidt, A. M., Dale, A. M., et al. (2006). Top-down facilitation of visual recognition. *Proceedings of the National Academy of Sciences of the United States of America*, *103*(2), 449–454.

Bar, M., & Ullman, S. (1996). Spatial context in recognition. *Perception*, *25*(3), 343–352.

Barcelo, F., Suwazono, S., & Knight, R. T. (2000). Prefrontal modulation of visual processing in humans. *Nature Neuroscience*, *3*(4), 399–403.

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

Bays, P. M., & Husain, M. (2012). Active inhibition and memory promote exploration and search of natural scenes. *Journal of Vision*, *12*, 8.

Bentin, S., Allison, T., Puce, A., Perez, E., & McCarthy, G. (1996). Electrophysiological studies of face perception in humans. *Journal of Cognitive Neuroscience*, *8*(6), 551–565.

Bentin, S., Deouell, L. Y., & Soroker, N. (1999). Selective visual streaming in face recognition: Evidence from developmental prosopagnosia. *Neuroreport*, *10*(4), 823–827.

Bichot, N. P., & Schall, J. D. (2002). Priming in macaque frontal cortex during popout visual search: Feature-based facilitation and location-based inhibition of return. *The Journal of Neuroscience*, *22*(11), 4675–4685.

Biederman, I. (1981). On the semantics of a glance at a scene. In: M. Kubovy, & J. R. Pomerantz (Eds.), *Perceptual organization* (pp. 213–253). Hillsdale, New Jersey: Lawrence Erlbaum Associates Inc.

Biederman, I., Glass, A. L., & Stacy, E. W. (1973). Searching for objects in real-world scences. *Journal of Experimental Psychology*, *97*(1), 22–27.

Biederman, I., Mezzanotte, R. J., & Rabinowitz, J. C. (1982). Scene perception: Detecting and judging objects undergoing relational violations. *Cognitive Psychology*, *14*(2), 143–177.

Biederman, I., Rabinowitz, J. C., Glass, A. L., & Stacy, E. W. (1974). Information extracted from a glance at a scene. *Journal of Experimental Psychology*, *103*(3), 597–600.

Boyce, S. J., & Pollatsek, A. (1992). Identification of objects in scenes—the role of scene background in object naming. *Journal of Experimental Psychology—Learning Memory and Cognition*, *18*(3), 531–543.

Boyce, S. J., Pollatsek, A., & Rayner, K. (1989). Effect of background information on object identification. *Journal of Experimental Psychology—Human Perception and Performance*, *15*(3), 556–566.

Braun, J., & Julesz, B. (1998). Withdrawing attention at little or no cost: Detection and discrimination tasks. *Perception and Psychophysics*, *60*(1), 1–23.

Brockmole, J. R., & Henderson, J. M. (2006). Recognition and attention guidance during contextual cueing in real-world scenes: Evidence from eye. *Quarterly Journal of Experimental Psychology*, *59*(7), 1177–1187.

Carlson, T., Tovar, D. A., Alink, A., & Kriegeskorte, N. (2013). Representational dynamics of object vision: The first 1000 ms. *Journal of Vision*, *13*(10).

Chong, S. C., & Treisman, A. (2003). Representation of statistical properties. *Vision Research*, *43*(4), 393–404 (doi: Pii S0042-6989(02)00596-5).

Chun, M. M., & Jiang, Y. H. (1998). Contextual cueing: Implicit learning and memory of visual context guides spatial attention. *Cognitive Psychology*, *36*(1), 28–71.

Chun, M. M., & Jiang, Y. H. (1999). Top-down attentional guidance based on implicit learning of visual covariation. *Psychological Science*, *10*, 360–365.

Codispoti, M., Ferrari, V., Junghöfer, M., & Schupp, H. T. (2006). The categorization of natural scenes: Brain attention networks revealed by dense sensor ERPs. *Neuroimage*, *32*(2), 583–591.

Corbetta, M., & Shulman, G. L. (2002). Control of goal-directed and stimulus-driven attention in the brain. *Nature Reviews Neuroscience*, *3*(3), 201–215.

Cumming, G. (2014). The new statistics: Why and how. *Psychological Science*, *25*(1), 7–29.

Davenport, J. L., & Potter, M. C. (2004). Scene consistency in object and background perception. *Psychological Science*, *15*(8), 559–564.

De Graef, P. (1992). Scene-context effects and models of real-world perception. In: K. Rayner (Ed.), *Eye movements and visual cognition: Scene perception and reading* (pp. 243–259). New York: Springer.

De Graef, P., Christiaens, D., & Dydewalle, G. (1990). Perceptual effects of scene context on object identification. *Psychological Research—Psychologische Forschung*, *52*(4), 317–329.

Demiral, Ş. B., Malcolm, G. L., & Henderson, J. M. (2012). ERP correlates of spatially incongruent object identification during scene viewing: Contextual expectancy versus simultaneous processing. *Neuropsychologia*, *50*, 1271–1285.

Doniger, G. M., Foxe, J. J., Murray, M. M., Higgins, B. A., Snodgrass, J. G., Schroeder, C. E., et al. (2000). Activation timecourse of ventral visual stream object-recognition areas: High density electrical mapping of perceptual closure processes. *Journal of Cognitive Neuroscience*, *12*(4), 615–621.

Dorris, M. C., Klein, R. M., Everling, S., & Munoz, D. P. (2002). Contribution of the primate superior colliculus to inhibition of return. *Journal of Cognitive Neuroscience*, *14*(8), 1256–1263.

Eriksen, C. W., & Hoffman, J. E. (1973). Extent of processing of noise elements during selective encoding from visual-displays. *Perception and Psychophysics*, *14*(1), 155–160.

Evans, E., & Treisman, A. (2005). Perception of objects in natural scenes: Is it really attention free? *Journal of Experimental Psychology: Human Perception and Performance*, *31*, 1476–1492.

Fecteau, J. H., & Munoz, D. P. (2005). Correlates of capture of attention and inhibition of return across stages of visual processing. *Journal of Cognitive Neuroscience*, *17*(11), 1714–1727.

Fiser, J., & Aslin, R. (2001). Unsupervised statistical learning of higher-order spatial structures from visual scenes. *Psychological Science*, *12*(6), 499.

Folk, C. L., Leber, A. B., & Egeth, H. E. (2002). Made you blink! Contingent attentional capture produces a spatial blink. *Perception & Psychophysics*, *64*(5), 741–753.

Folk, C. L., Remington, R. W., & Johnston, J. C. (1992). Involuntary covert orienting is contingent on attentional control settings. *Journal of Experimental Psychology—Human Perception and Performance*, *18*(4), 1030–1044.

Folstein, J. R., Van Petten, C., & Rose, S. A. (2008). Novelty and conflict in the categorization of complex stimuli. *Psychophysiology*, *45*(3), 467–479.

Friedman, A. (1979). Framing pictures—Role of knowledge in automatized encoding and memory for gist. *Journal of Experimental Psychology—General*, *108*(3), 316–355.

Ganis, G., & Kutas, M. (2003). An electrophysiological study of scene effects on object identification. *Cognitive Brain Research*, *16*(2), 123–144.

Goffaux, V., Jacques, C., Mouraux, A., Oliva, A., Schyns, P., & Rossion, B. (2005). Diagnostic colours contribute to the early stages of scene categorization: Behavioural and neurophysiological evidence. *Visual Cognition*, *12*(6), 878–892.

Gosling, A., & Eimer, M. (2011). An event-related brain potential study of explicit face recognition. *Neuropsychologia*, *49*(9), 2736–2745.

Graboi, D., & Lisman, J. (2003). Recognition by top-down and bottom-up processing in cortex: The control of selective attention. *Journal of Neurophysiology*, *90*(2), 798–810, http://dx.doi.org/10.1152/jn.00777.2002.

Hamm, J. P., Johnson, B. W., & Kirk, I. J. (2002). Comparison of the N300 and N400 ERPs to picture stimuli in congruent and incongruent contexts. *Clinical Neurophysiology*, *113*(8), 1339–1350.

Henderson, J. M., Pollatsek, A., & Rayner, K. (1989). Covert visual-attention and extrafoveal information use during object identification. *Perception and Psychophysics*, *45*(3), 196–208.

Henderson, J. M., Weeks, P. A., & Hollingworth, A. (1999). The effects of semantic consistency on eye movements during complex scene viewing. *Journal of Experimental Psychology: Human Perception and Performance*, *25*(1), 210–228.

Henson, R. N., Rylands, A., Ross, E., Vuilleumeir, P., & Rugg, M. D. (2004). The effect of repetition lag on electrophysiological and haemodynamic correlates of visual object priming. *Neuroimage*, *21*, 1674–1689.

Hidalgo-Sotelo, B., Oliva, A., & Torralba, A. (2005). Human learning of contextual priors for object search: Where does the time go? *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, *3*, 86–93.

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

Hollingworth, A., & Henderson, J. M. (1998). Does consistent scene context facilitate object perception? *Journal of Experimental Psychology—General*, *127*(4), 398–415.

Hollingworth, A., & Henderson, J. M. (1999). Object identification is isolated from scene semantic constraint: Evidence from object type and token discrimination. *Acta Psychologica*, *102*(2-3), 319–343.

Hooge, I. T. C., & Frens, M. A. (2000). Inhibition of saccade return (ISR): Spatio-temporal properties of saccade programming. *Vision Research*, *40*(24), 3415–3426.

Hooge, I. T. C., Over, E. A., van Wezel, R. J., & Frens, M. A. (2005). Inhibition of return is not a foraging facilitator in saccadic search and free viewing. *Vision Research*, *45*(14), 1901–1908.

Hopfinger, J. B., Buonocore, M. H., & Mangun, G. R. (2000). The neural mechanisms of top-down attentional control. *Nature Neuroscience*, *3*(3), 284–291.

Humphreys, G., Riddoch, M., & Price, C. (1997). Top-down processes in object identification: Evidence from experimental psychology, neuropsychology and functional anatomy. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *352*(1358), 1275.

Hung, C. P., Kreiman, G., Poggio, T., & DiCarlo, J. J. (2005). Fast readout of object identity from macaque inferior temporal cortex. *Science*, *310*(5749), 863–866.

Itti, L., & Koch, C. (2000). A saliency-based search mechanism for overt and covert shifts of visual attention. *Vision Research*, *40*(10–12), 1489–1506.

Jung, T. P., Makeig, S., Humphries, C., Lee, T. W., McKeown, M. J., Iragui, V., et al. (2000). Removing electroencephalographic artifacts by blind source separation. *Psychophysiology*, *37*(2), 163–178.

Kirchner, H., & Thorpe, S. J. (2006). Ultra-rapid object detection with saccadic eye movements: Visual processing speed revisited. *Vision Research*, *46*(11), 1762–-1776, http://dx.doi.org/10.1016/j.visres.2005.10.002.

Klein, R. M. (1988). Inhibitory tagging system facilitates visual search. *Nature*, *334*, 430–431.

Klein, R. M. (2000). Inhibition of return. *Trends in Cognitive Sciences*, *4*(4), 138–147.

Klein, R. M., & MacInnes, W. J. (1999). Inhibition of return is a foraging facilitator in visual search. *Psychological Science*, *10*(4), 346–352.

Kosslyn, S. M. (1994). *Image and brain*. Cambridge, MA: MIT Press.

Kutas, M., & Hillyard, S. A. (1980a). Event-related brain potentials to semantically inappropriate and surprisingly large words. *Biological Psychology*, *11*(2), 99–116.

Kutas, M., & Hillyard, S. A. (1980b). Reading Senseless Sentences—Brain potentials reflect semantic incongruity. *Science*, *207*(4427), 203–205.
Lee, D. K., Koch, C., & Braun, J. (1999). Attentional capacity is undifferentiated: Concurrent discrimination of form, color, and motion. *Perception and Psychophysics*, *61*(7), 1241–1255.
Li, F. F., VanRullen, R., Koch, C., & Perona, P. (2002). Rapid natural scene categorization in the near absence of attention. *Proceedings of the National Academy of Sciences of the United States of America*, *99*(14), 9596–9601.
Loftus, G. R., & Mackworth, N. H. (1978). Cognitive determinants of fixation location during picture viewing. *Journal of Experimental Psychology—Human Perception and Performance*, *4*(4), 565–572.
Logothetis, N. K., Leopold, D. A., & Sheinberg, D. L. (1996). What is rivalling during binocular rivalry? *Nature*, *380*(6575), 621–624.
Maris, E., & Oostenveld, R. (2007). Nonparametric statistical testing of EEG-and MEG-data. *Journal of Neuroscience Methods*, *164*(1), 177–190.
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.
McPherson, W. B., & Holcomb, P. J. (1999). An electrophysiological investigation of semantic priming with pictures of real objects. *Psychophysiology*, *36*(1), 53–65.
Mechelli, A., Price, C. J., Friston, K. J., & Ishai, A. (2004). Where bottom-up meets top-down: Neuronal interactions during perception and imagery. *Cerebral Cortex*, *14*(11), 1256–1265, http://dx.doi.org/10.1093/cercor/bhh087.
Merigan, W. H., & Maunsell, J. H. R. (1993). How parallel are the primate visual pathways. *Annual Review of Neuroscience*, *16*, 369–402.
Mirpour, K., Arcizet, F., Ong, W. S., & Bisley, J. W. (2009). Been there, seen that: A neural mechanism for performing efficient visual search. *Journal of Neurophysiology*, *102*(6), 3481–3491.
Miyashita, Y., & Hayashi, T. (2000). Neural representation of visual objects: Encoding and top-down activation. *Current Opinion in Neurobiology*, *10*(2), 187–194.
Mudrik, L., Deouell, L. Y., & Lamy, D. (2011). Scene congruency biases binocular rivalry. *Consciousness and Cognition*, *20*, 756–767.
Mudrik, L., Lamy, D., & Deouell, L. Y. (2010). ERP evidence for context congruity effects during simultaneous object-scene processing. *Neuropsychologia*, *48*(2), 507–517, http://dx.doi.org/10.1016/j.neuropsychologia.2009.10.011.
Muller, H. J., & Rabbitt, P. M. A. (1989). Reflexive and voluntary orienting of visual-attention—Time course of activation and resistance to interruption. *Journal of Experimental Psychology—Human Perception and Performance*, *15*(2), 315–330.
Munneke, J., Brentari, V., & Peelen, M. V. (2013). The influence of scene context on object recognition is independent of attentional focus. *Frontiers in Psychology*, *4*, 552.
Nakayama, K., & Mackeben, M. (1989). Sustained and transient components of focal visual-attention. *Vision Research*, *29*(11), 1631–1647.
Neumann, D., & Gegenfurtner, K. (2006). Image retrieval and perceptual similarity. *ACM Transactions on Applied Perception (TAP)*, *3*(1), 31–47.
Oliva, A., & Schyns, P. G. (2000). Diagnostic colors mediate scene recognition. *Cognitive Psychology*, *41*(2), 176–210.
Oliva, A., & Torralba, A. (2001). Modeling the shape of the scene: A holistic representation of the spatial envelope. *International Journal of Computer Vision*, *42*(3), 145–175.
Oliva, A., & Torralba, A. (2006). Building the gist of a scene: The role of global image features in recognition. *Progress in Brain Research*, *155*, 23–36.
Oliva, A., Wolfe, J. M., & Arsenio, H. C. (2004). Panoramic search: The interaction of memory and vision in search through a familiar scene. *Journal of Experimental Psychology—Human Perception and Performance*, *30*(6), 1132–1146, http://dx.doi.org/10.1037/0096-1523.30.6.1132.
Palmer, S. E. (1975). Effects of contextual scenes on identification of objects. *Memory and Cognition*, *3*(5), 519–526.
Pascual-Leone, A., & Walsh, V. (2001). Fast back projections from the motion to the primary visual area necessary for visual awareness. *Science*, *292*(5516), 510–512.
Pashler, H., & Harris, C. R. (2012). Is the replicability crisis overblown? Three arguments examined. *Perspectives on Psychological Science*, *7*(6), 531–536.
Picton, T. W., Bentin, S., Berg, P., Donchin, E., Hillyard, S. A., Johnson, R., et al. (2000). Guidelines for using human event-related potentials to study cognition: Recording standards and publication criteria. *Psychophysiology*, *37*(2), 127–152.
Posner, M. I. (1980). Orienting of attention. *Quarterly Journal of Experimental Psychology*, *32*(Feb), 3–25.
Posner, M. I., & Cohen, Y. (1984). Components of visual orienting. In: H. Bouma, & D. Bowhuis (Eds.), *Attention and performance*, Vol. X (pp. 531–556). Hillsdale, NJ: Erlbaum.
Posner, M. I., Rafal, R. D., Choate, L. S., & Vaughan, J. (1985). Inhibition of return: Neural basis and function. *Cognitive Neuropsychology*, *2*(3), 211–228.
Potter, M. C., Staub, A., & O'Connor, D. H. (2004). Pictorial and conceptual representation of glimpsed pictures. *Journal of Experimental Psychology-Human Perception and Performance*, *30*(3), 478–489.
Ranganath, C., DeGutis, J., & D'Esposito, M. (2004). Category-specific modulation of inferior temporal activity during working memory encoding and maintenance. *Cognitive Brain Research*, *20*(1), 37–45, http://dx.doi.org/10.1016/j.cogbrainres.2003.11.017.
Rayner, K., & Pollatsek, A. (1992). Eye movements and scene perception. *Canadian Journal of Psychology*, *46*(3), 342–376.
Schendan, H. E., & Kutas, M. (2002). Neurophysiological evidence for two processing times for visual object identification. *Neuropsychologia*, *40*(7), 931–945.
Schendan, H. E., & Kutas, M. (2003). Time course of processes and representations supporting visual object identification and memory. *Journal of Cognitive Neuroscience*, *15*(1), 111–135.
Schendan, H. E., & Kutas, M. (2007). Neurophysiological evidence for the time course of activation of global shape, part, and local contour representations during visual object categorization and memory. *Journal of Cognitive Neuroscience*, *19*(5), 734–749.
Schendan, H. E., & Maher, S. M. (2008). Object knowledge during entry-level categorization is activated and modified by implicit memory after 200 ms. *Neuroimage*, *44*(4), 1423–1438.
Serences, J. T., Shomstein, S., Leber, A. B., Golay, X., Egeth, H. E., & Yantis, S. (2005). Coordination of voluntary and stimulus-driven attentional control in human cortex. *Psychological Science*, *16*(2), 114–122.
Sitnikova, T., Holcomb, P. J., Kiyonaga, K. A., & Kuperberg, G. R. (2008). Two neurocognitive mechanisms of semantic integration during the comprehension of visual real-world events. *Journal of Cognitive Neuroscience*, *20*(11), 2037–2057.
Sun, H.-M., Simon-Dack, S. L., Gordon, R. D., & Teder, W. A. (2011). Contextual influences on rapid object categorization in natural scenes. *Brain Research*, *1398*, 40–54.
Thorpe, C., Fize, D., & Marlot, C. (1996). Speed of processing in the human visual system. *Nature*, *381*, 520–522.
Torralba, A., Oliva, A., Castelhano, M. S., & Henderson, J. M. (2006). Contextual guidance of eye movements and attention in real-world scenes: The role of global features in object search. *Psychological Review*, *113*(4), 766–786.
Underwood, G., & Foulsham, T. (2006). Visual saliency and semantic incongruency influence eye movements when inspecting pictures. *Quarterly Journal of Experimental Psychology*, *59*(11), 1931–1949.
Underwood, G., Foulsham, T., van Loon, E., Humphreys, L., & Bloyce, J. (2006). Eye movements during scene inspection: A test of the saliency map hypothesis. *European Journal of Cognitive Psychology*, *18*(3), 321–342.
Underwood, G., Templeman, E., Lamming, L., & Foulsham, T. (2008). Is attention necessary for object identification? Evidence from eye movements during the inspection of real-world scenes. *Consciousness and cognition*, *17*(1), 159–170.
VanRullen, R., & Thorpe, S. J. (2001). The time course of visual processing: From early perception to decision-making. *Journal of Cognitive Neuroscience*, *13*(4), 454–461.
Võ, M. L.-H., & Henderson, J. M. (2009). Does gravity matter? Effects of semantic and syntactic inconsistencies on the allocation of attention during scene perception. *Journal of Vision*, *9*(3).
Võ, M. L.-H., & Henderson, J. M. (2011). Object–scene inconsistencies do not capture gaze: Evidence from the flash-preview moving-window paradigm. *Attention, Perception, & Psychophysics*, *73*(6), 1742–1753.
Võ, M. L.-H., & Wolfe, J. M. (2013). Differential electrophysiological signatures of semantic and syntactic scene processing. *Psychological science*, *24*, 1816–1823.
Zheng, X., Mondloch, C. J., & Segalowitz, S. J. (2012). The timing of individual face recognition in the brain. *Neuropsychologia*, *50*(7), 1451–1461.