Neuropsychologia xxx (2009) xxx–xxx

Contents lists available at ScienceDirect

Neuropsychologia

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

ERP evidence for context congruity effects during simultaneous object–scene processing

Liad Mudrik[a,*], Dominique Lamy[a], Leon Y. Deouell[b]

[a] Department of Psychology, Tel Aviv University, Israel
[b] Department of Psychology, and Interdisciplinary Center for Neural Computation, The Hebrew University of Jerusalem, Israel

A R T I C L E I N F O

Article history:
Received 29 April 2009
Received in revised form 6 October 2009
Accepted 9 October 2009
Available online xxx

Keywords:
Visual scenes
Context effects
N400
Event-related potentials
Incongruency

A B S T R A C T

Contextual regularities help us analyze visual scenes and form judgments on their constituents. The present study investigates the effect of context violation on scene processing using event-related potentials (ERPs). We compared ERPs evoked by congruent vs. incongruent visual scenes (e.g., a man playing a violin vs. a man "playing" a broomstick), when the scene and object are presented simultaneously, so subjects cannot form previous expectations about the object's identity. As expected, an ongoing anterior negativity emerged around 270 ms post-stimulus presentation, lasting for about 330 ms. This negativity, resembling the N300/N400 effect previously associated with semantic integration, was followed by a later and broadly distributed negativity between 650 and 850 ms, possibly related to late processes of semantic evaluation and response preparation. The results confirm that contextual congruity affects scene processing starting from ∼300 ms or earlier, and that this early electrophysiological congruity effect indeed reflects context violation processing, rather than indexing a mismatch between expected vs. actual events, or between prepared vs. correct responses. They also suggest that contextual information may affect object model selection processes, and influence later stages of semantic knowledge activation and/or decision-making.

1. Introduction

Imagine coming to a tea party, and seeing one of the participants dip a hand clock into a tea pot, or watching a croquet game, where the players use flamingoes instead of mallets. Unless you are Alice visiting Wonderland, you would probably be confused. Indeed, in real life we are accustomed to seeing certain objects in specific and recurring contexts: hand watches are normally seen on wrists and flamingoes at the zoo or in their natural habitat. Such contextual regularities help us perform fast judgments about the objects that make up a scene (Antes, Penland, & Metzger, 1981; Biederman, 1972; Biederman, Glass, & Stacy, 1973; Biederman, Rabinowitz, Glass, & Stacy, 1974; Boyce & Pollatsek, 1992; Boyce, Pollatsek, & Rayner, 1989; Friedman, 1979; Kosslyn, 1994; Palmer, 1975; Rayner & Pollatsek, 1992). Behavioral studies have shown that incongruent objects (i.e., objects with low probability of appearing in their current context) are identified slower (Bar & Ullman, 1996; Boyce & Pollatsek, 1992; Chun & Jiang, 1998; Davenport & Potter, 2004; Neider & Zelinsky, 2006; Palmer, 1975; Rieger, Kochy, Schalk, Gruschow, & Heinze, 2008) and less accurately (Antes et al., 1981; Bar & Ullman, 1996; Biederman, 1972; Biederman et al., 1974; Boyce et al., 1989; Davenport & Potter, 2004; Underwood, 2005) than congruent objects (but see Hollingworth & Henderson, 1998). Yet, by defying our expectations, incongruent objects also become salient enough to attract our attention, resulting in earlier, prolonged and more frequent eye fixations (Friedman, 1979; Loftus & Mackworth, 1978; Underwood & Foulsham, 2006; Underwood, Foulsham, van Loon, Humphreys, & Bloyce, 2006; Underwood, Templeman, Lamming, & Foulsham, 2008), and better resistance to change blindness (Hollingworth & Henderson, 2000).

If incongruent objects are identified later than congruent ones, how can they attract attention early enough to induce earlier fixations? Underwood and his colleagues (2008) suggested that Loftus and Mackworth's (1978) model of scene perception may resolve this apparent paradox. According to the model, the gist of the scene can be rapidly extracted, at an early perceptual stage during which the objects that make up the scene are partially recognized. This crude analysis could be based on low spatial frequency ("magnocellular") information, and may involve rapid transmission of information to medial temporal structures and up to the frontal cortex, which subsequently 'guides' analysis in ventral temporal cortex (Bar, 2004; Bar et al., 2006). Partial recognition of a non-fixated object may be sufficient to determine that it violates the gist of the scene and requires further inspection, which triggers an

---

* 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).

0028-3932/$ – see front matter © 2009 Elsevier Ltd. All rights reserved.
doi:10.1016/j.neuropsychologia.2009.10.011

eye movement to its location. During the stage of detailed analysis that follows, the inconsistency between the object and its context delays full identification.

In line with this model, extensive empirical work has shown that the gist of scenes and objects can be rapidly identified. Human observers are able to recognize scenes with exposure durations as short as 26 ms (Rieger, Braun, Bulthoff, & Gegenfurtner, 2005; Rousselet, Joubert, & Fabre-Thorpe, 2005), possibly before perceptual processing is complete (Biederman et al., 1974; Carr, Mccauley, Sperber, & Parmelee, 1982; Hochstein & Ahissar, 2002; Intraub, 1981; Mccauley, Parmelee, Sperber, & Carr, 1980; Oliva & Schyns, 1997, 2000; Oliva & Torralba, 2006). Similarly, images of isolated objects presented for no more than 20 ms can typically be categorized with high accuracy (94% correct or more), median reaction times neighboring 400 ms and shortest response latencies around 250 ms (e.g., Delorme, Richard, & Fabre-Thorpe, 2000; Fabre-Thorpe, Delorme, Marlot, & Thorpe, 2001; Fabre-Thorpe, Richard, & Thorpe, 1998; Rousselet, Mace, & Fabre-Thorpe, 2003).

However, detection of context violation requires not only scene and object identification, but also semantic computation of the object's probability to appear in the scene (Loftus & Mackworth, 1978). Can such computation occur so early during perceptual processing, as suggested by Underwood et al. (2008)? The current literature is divided on this issue.

According to the *perceptual schema model* (Antes et al., 1981; Biederman, 1981; Biederman, Mezzanotte, & Rabinowitz, 1982; Biederman, Teitelbaum, & Mezzanotte, 1983; Boyce et al., 1989; Palmer, 1975), global, low-resolution contextual information leads to rapid activation of a scene schema (Loftus, Nelson, & Kallman, 1983; Metzger & Antes, 1983; Schyns & Oliva, 1994). This, in turn, prompts feature-selective attention that facilitates the subsequent detection of perceptual features associated with objects specified within the schema itself (Antes et al., 1981; Friedman, 1979). So, at very early perceptual stages, context violations impede the analysis of schema-incongruent objects, relative to schema-congruent ones.

Object model selection or *matching models* (Bar, 2004; Bar & Aminoff, 2003; Bar & Ullman, 1996; Friedman, 1979; Kosslyn, 1994; Ullman, 1996) similarly suggest that contextual information activates a scene schema, which primes identification of schema-congruent object types by modulating the observer's criterion regarding the amount of perceptual information needed to match a particular object representation. When the input image has been sufficiently analyzed to be compared to stored representations, contextual activation facilitates convergence into the most probable interpretation (Bar, 2004). Thus, as opposed to the perceptual schema model, the matching models propose that processing of context information influences mainly the decision about object identity by producing top-down 'informed guesses', but does not affect the actual perceptual analysis of the object. However, both models predict that context violations have their effect *prior* to objects' complete identification.

On the other hand, the *functional isolation model* (De Graef, 1992; De Graef, Christiaens, & Dydewalle, 1990; Hamm, Johnson, & Kirk, 2002; Henderson, Weeks, & Hollingworth, 1999; Hollingworth & Henderson, 1998, 1999) denies any early contextual influences on object recognition. According to this model, contextual information comes into effect only after object identification, and influences semantic knowledge activation and/or decision-making (Henderson & Hollingworth, 1999). In addition, the model suggests that attention is not initially drawn to an incongruent object, and that once such an object has been fixated, the eyes tend to linger on it (De Graef et al., 1990; Henderson et al., 1999).

Findings from several behavioral studies have been put forward as evidence against the functional isolation model (e.g., Auckland, Cave, & Donnelly, 2007; Davenport & Potter, 2004). These showed that context violations impaired subjects' performance with exposure durations shorter than 100 ms (and as short as 26 ms, Joubert, Rousselet, Fize, & Fabre-Thorpe, 2007).

However, such behavioral measures may not be appropriate to uncover the time course of context violation processing. Exposure durations cannot signify when context violation processing actually occurs, because processing is likely to continue after the stimulus is no longer in view. Reaction times also do not indicate when the violation was processed, as they merely represent an upper estimate of processing time that also includes the time needed for decisional and motor mechanisms to unfold (Ganis & Kutas, 2003; Luce, 1986; VanRullen & Thorpe, 2001). By contrast, recordings of event-related potentials (ERPs) provide a suitable tool for contrasting early vs. late contextual violation processing accounts.

Ganis and Kutas (2003) used this approach to investigate processing of context violations within a given visual scene. They presented subjects with a pre-cue, followed by a real-life visual scene (e.g., players in a football field). After 300 ms, an object that was either congruent (a football) or incongruent (a toilet paper roll) with the surrounding scene, appeared at the cued location. Subjects were asked to identify the object. The authors reported an "N390" congruency effect, beginning at ∼300 ms and lasting until ∼500 ms with larger amplitude for incongruent than for congruent objects. This component resembled the N400 effect typically found for a verbal (e.g., Holcomb, 1993; Kutas & Hillyard, 1980a, 1980b, 1980c) or a visual stimulus (Barrett & Rugg, 1990; Federmeier & Kutas, 2002; Ganis, Kutas, & Sereno, 1996; McPherson & Holcomb, 1999; Nigam, Hoffman, & Simons, 1992; Sitnikova, Kuperberg, & Holcomb, 2003; West & Holcomb, 2002) that violates the semantic context set by preceding stimuli, although it was not as frontal as in previous studies using pictorial stimuli (e.g., Federmeier & Kutas, 2001; Ganis et al., 1996). The similarity to the semantic N400 and the absence of any earlier differences in the brain activity associated with the congruent vs. incongruent conditions led the authors to suggest that contextual effects take place only at the stage of semantic analysis.

However, two important characteristics of Ganis and Kutas' (2003) procedure may complicate the interpretation of the observed N390 component. First, the context and location of the critical object were known before this object appeared. Thus, subjects probably formed perceptual expectations with regard to the critical object. For example, it seems reasonable that after seeing a football field and fixating their gaze on the players' legs following the pre-cue, subjects immediately expect a football to appear next. When they are presented with a toilet paper roll instead, subjects do not encounter only a violation of the context itself but also a violation of their previously formed expectations. Second, the subjects' task was to identify the critical object. Having formed an expectation as to its identity, subjects are likely to prepare the associated response. Then, when a different object appears, they need to change their response accordingly. Thus, rather than indexing processing of context violation per se, the N390 in Ganis and Kutas' study might reflect either a mismatch between a previously formed expectation and the object that actually appears, or a mismatch between the prepared response and the response that turns out to be correct. In that sense, the N390 is similar to the traditional N400 effect which is typically triggered by a stimulus that is incongruent with the expectation and associated response created by a sentence or visual sequence (Ganis et al., 1996; e.g., Holcomb, 1993; Kutas & Hillyard, 1980a, 1980b, 1980c).

The objective of the present study was to track the time course of visual context violation processing, while avoiding the effects of prior expectation. Accordingly, in the present study the scene was presented simultaneously with the object and the subjects' task did

Please cite this article in press as: Mudrik, L., et al. ERP evidence for context congruity effects during simultaneous object–scene processing. *Neuropsychologia* (2009), doi:10.1016/j.neuropsychologia.2009.10.011

not involve identification of the congruent/incongruent object, or object naming.

Subjects saw images depicting a person performing an action that involved a congruent or an incongruent object (e.g., a man drinking from a can or from a potato, respectively), and reported how many hands the person used to perform the action (one hand, two or none). In order to direct the subjects' gaze to the subsequent location of the critical object, a pre-cue was presented at the beginning of each trial. Note that aside from avoiding the described mismatch-related confounds, simultaneous presentation of the critical object and its context may be more akin to real-life situations in which objects are frequently perceived within a given context rather than suddenly popping out into a scene, although such situations also exist.

Behaviorally, we expected to replicate the well-documented performance advantage for congruent relative to incongruent objects. On the ERP measure, if the N390 effect indeed reflects context violation processing rather than a response/expectation mismatch, we expected to find an enhanced negativity for scenes containing an incongruent object in the 300–500 ms time window, as reported by Ganis and Kutas (2003). In addition, any earlier difference in the ERP waveform between scenes that contain congruent vs. incongruent objects would support the notion that context processing interacts with perceptual stages of object processing (e.g., Bar, 2004; Biederman, 1981).

2. Methods and materials

2.1. Participants

Sixteen healthy volunteers, students of the Hebrew University of Jerusalem, with reportedly normal or corrected-to-normal sight and no psychiatric or neurological history, participated in the study for payment (~5$ per hour). Three subjects were excluded from analysis due to noisy recordings. The remaining 13 subjects (6 males, 7 females) were 19–29 years old (mean = 23.8), 11 of them right handed.

The experiment was approved by the ethics committee of the department of psychology at the Hebrew University, 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 in. CRT monitor with a 100-Hz refresh rate, using E-prime 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 eighty-eight pairs of colored pictures were designed for the experiment, using Adobe Photoshop software. Each pair included a congruent scene and an incongruent scene (Fig. 1). The congruent scene was a real-life picture taken from Internet sources and depicted a human action involving an object (e.g., a man playing a violin, a woman using a microscope and so on). The incongruent scene was created by replacing the original object of the action with a different, unrelated object (e.g., instead of a violin, the man was "playing" a broomstick). The objects were also chosen from real-life pictures taken from Internet sources. Special care was given to the assimilation of the out-of-context object into the new scene, avoiding unnatural edges, adding or removing shadows as necessary, etc. For example, if a tennis racquet was pasted to a new scene, the background seen through the strings of the racquet was painstakingly replaced by that of the new scene (Fig. 1, Image 2). The pictures' luminance and contrast levels were digitally equated using the Adobe Photoshop software.

Prior to the main experiment, all pictures were rated by 24 subjects as either highly incongruent or congruent. Twenty-seven pairs among the pre-tested pairs were excluded because they elicited inconsistent ratings across subjects. Thus, the final set included 161 pairs.

Post-hoc, several analyses were conducted in order to ensure that digitally pasting the critical object onto the image for the incongruent but not for the congruent scenes did not create systematic differences in early stages of processing between the two types of scenes. First, all pairs were examined using the visual saliency model developed by Itti and Koch (2000). The model combines information across dimensions, namely orientation, intensity and color information, in a purely stimulus-driven manner, to create an explicit two-dimensional map that encodes the visual saliency or conspicuity of objects. The saliency values of all objects were determined, and objects were divided accordingly (low, medium or high saliency). No differences were found between the saliency levels of congruent vs. incongruent objects. In addition, no saliency difference was found between the non-natural edges of the objects in the digitally processed incongruent scenes and the natural edges of the objects in the congruent scenes.

Fig. 1. Examples of congruent (right) and incongruent (left) scenes. (From top to bottom) A woman putting either food or chess board in the oven. A boy holding a bow and either an arrow or a tennis racket. Two players playing basketball with either a ball or a watermelon.

Second, we evaluated the images' perceived similarity in chromaticity and spatial frequency according to Neumann and Gegenfurtner's (2006) model that relies on psychophysically based indexes. The indexes are constructed in accordance with some of the known properties of the early stages of human vision. The color codes in the "red-green" and "blue-yellow" channels are modeled using the color-opponent axes of a DKL color space and a logarithmic-radial scaling for the histogram bins. 2D discrete Fourier transform is used to create an orientation and spatial frequency histogram analogous to similar representations in the visual cortex. We applied the model to calculate similarity scores between every two images used in the experiment, and then compared the average similarity measures of all within-group pairs (congruent–congruent and incongruent–incongruent), to the similarity calculated

Please cite this article in press as: Mudrik, L., et al. ERP evidence for context congruity effects during simultaneous object–scene processing. *Neuropsychologia* (2009), doi:10.1016/j.neuropsychologia.2009.10.011

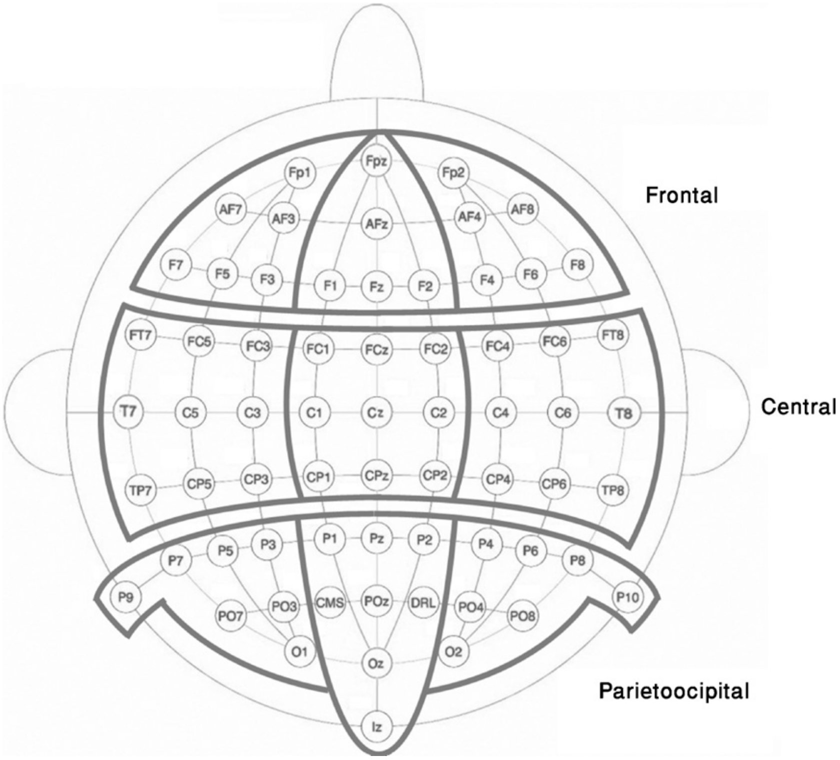

Fig. 2. Electrode array and division of the electrodes into regions.

between pairs of congruent and incongruent images, using Student's t-test. If congruent images systematically differ from incongruent images along these measures of chromaticity and spatial frequency, higher similarity measures should be found for within-group pairs, than across groups. Under this assumption, we conducted a non-parametric bootstrap analysis, to assess whether the difference between within and between-group similarities was significant, without any assumption on the normality of the distributions. The rational of this analysis was that under the null hypothesis of no difference between groups, the designation of an image as congruent or incongruent is immaterial. Thus, for 2500 iterations, we randomly permuted the similarity matrix so that a random number of similarity measures were designated 'between-' instead of 'within-' and vice versa. Thus, on each iteration we produced surrogate 'within-group' and 'between-group' similarity ratings and we compared their mean ratings using the Student's t-test. The frequency distribution of the t values thus obtained represents the distribution of t values under the null hypothesis of no difference. A t value in the true comparison (i.e., without mixing the similarity matrix) exceeding 95% of these values would be violating the assumption of no difference. In the spatial frequency domain, the comparison between the similarity scores within and between groups ($t(51518) = 0.16$) was not significantly different from the null distribution. In the chromaticity domain the images were in fact significantly more similar *between* groups than within ($t(51518) = 2.19$, $p < 0.03$). This is likely to reflect the fact that each image had a comparable image in the other group but not within group. Note that similar qualitative results were obtained using parametric statistics (i.e., comparing the t values to the textbook t distribution assuming normality). Thus, congruent and incongruent images could not be divided by luminance, contrast, chromaticity, spatial frequency, or object saliency.

2.3. Procedure

The experiment included 161 trials. For half the subjects, 80 trials were congruent and 81 incongruent and vice versa for the other half. Congruent and incongruent trials were intermixed, with the constraint that the same condition (congruent/incongruent scene) was never presented in four consecutive trials, and that only one picture from each pair was presented during the experiment. In other words, each subject saw either the congruent or the incongruent version from each pair, with counterbalancing between subjects.

At the beginning of each trial, a pre-cue (white cross) was presented for 200 ms and served as a fixation point. That is, subjects were instructed to direct their gaze to the cued location. One hundred ms after the pre-cue offset, a scene was presented, with the object occupying the pre-cued location (i.e., the center of gaze) being either congruent or incongruent with its context (henceforth, congruent vs. incongruent scene, respectively). The scene, including the object, remained on the screen for 1000 ms.

After stimulus presentation, a question appeared: how many hands were used by the person in the picture to perform the action? 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.

Following the ERP session, the pictures were viewed again, and the subjects were required to rate how unusual they thought each picture was, on a scale ranging from 1 (completely usual) to 5 (completely unusual). This procedure was used in order to validate the congruency level assigned to the stimuli during the pretest that was conducted on different subjects.

2.4. ERP methods

2.4.1. ERP recording

The EEG was recorded using the Active 2 system (BioSemi) from 64 electrodes distributed based on the extended 10–20 system (see 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 and were subsequently re-referenced digitally. The EEG was continuously sampled at 1024 Hz with 24 bit resolution, low-pass filtered online with a cutoff of 256 Hz and stored for off-line analysis.

2.4.2. ERP analysis

ERP analysis was conducted using the "Brain Vision Analyzer" software (version 1.05; Brain Products, Germany). For consistency with several studies addressing the N400 effects, data from all channels were referenced off-line 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).[1] 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

---

[1] Subjects blinked at the beginning of only 1.6% and 1.4% of congruent and incongruent trials, respectively (difference not significant).

Please cite this article in press as: Mudrik, L., et al. ERP evidence for context congruity effects during simultaneous object–scene processing. *Neuropsychologia* (2009), doi:10.1016/j.neuropsychologia.2009.10.011

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. Ten percent of segments were rejected by this procedure on the average.

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 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 using the average amplitude within a pre-defined window of 300–500 ms, based on the results reported by Ganis and Kutas (2003). Differences were assessed using a three-way ANOVA with Region (Frontal, Central, Parieto-occipital; Fig. 2), Laterality (Left, Midline, Right) and Congruity (Congruent, Incongruent) as factors. Second, because a crucial issue was how early context violations come into effect, onset latency was investigated. We ran a point-by-point t-test on the difference wave between context-congruent and context-incongruent conditions (Picton et al., 2000). To reduce the number of t-tests the waveforms were down-sampled to 256 Hz. To control for multiple comparisons in the temporal domain, we required that at least 20 consecutive t-tests achieve $p < 0.05$ (two-tailed) (Guthrie & Buchwald, 1991). Note that this correction method renders the analysis less sensitive to transient, short-lived effects. For analyses designed to examine distribution differences, we normalized the amplitudes using the vector-scaling method as described in McCarthy and Wood (1985).[2] 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

Incongruent images in the ERP experiment were created by pasting a new object into a scene, whereas congruent images were not digitally manipulated. Although we were very careful in eliminating evidence of the photomontage process (see Fig. 1 for examples) we conducted a behavioral follow-up experiment to investigate the possibility that subjects could nonetheless perceive the digital processing (pasting) of objects based on some visual cue. To dissociate the 'pasted' vs. 'non-pasted' from the 'congruent' vs. 'incongruent' dichotomy, we created a new set of 40 pairs of congruent images (taken from the main experiment); for each congruent image, another image was prepared in which the critical object had been replaced by another congruent object copied from a different picture (see Fig. 3 below). This cut-and-paste procedure was undertaken using the exact same software and techniques, and by the same author (LM) who performed the manipulation of the incongruent images in the main experiment. We presented each of eight naïve subjects with two experimental blocks (the order was counterbalanced). In the critical block, 40 congruent images, half of them with a foreign (but congruent) object pasted and half natural, were presented in the same way they were presented in the main experiment (200 ms cue–100 ms blank screen–1000 ms of image), and subjects were explicitly asked to determine if the object whose location was cued was pasted or not. Each subject saw either the pasted version or the natural version of each scene, in a counterbalanced fashion. In another block, 20 congruent and 20 incongruent images from the main experiment were presented following a cue, with the same temporal sequence as in the main experiment, and subjects were instructed to determine if the image was congruent or not.

3. Results

3.1. Behavioral results

3.1.1. Task performance

Subjects answered whether the object at fixation was manipulated by zero, one, or two hands. Their mean accuracy scores and RTs (on correct trials) were computed and compared for congruent and incongruent scenes. Although the question did not address the congruity of the scene, the subjects answered it more accurately for congruent scenes ($M = 0.92$, $SD = 0.04$) than for incongruent scenes ($M = 0.90$, $SD = 0.04$) ($t(12) = 3.53$, $p < 0.005$) and RTs were shorter for congruent scenes ($M = 783.51$ ms, $SD = 320.74$ ms) than for incongruent scenes ($M = 855.48$ ms, $SD = 340.25$ ms) ($t(12) = 3.73$, $p < 0.005$).

Fig. 3. Examples of pasted congruent (left) and natural congruent (right) images used in the follow-up behavioral experiment.

3.1.2. Post-test congruency ratings

On a scale ranging from 1 (very usual) to 5 (very unusual), congruent scenes were rated as usual ($M = 1.15$, $SD = 0.12$), while incongruent scenes were rated as unusual ($M = 4.55$, $SD = 0.27$). For all subjects, the number of congruent scenes that were rated as unusual ($M = 2.23$, $SD = 1.69$), or of incongruent scenes that were rated as usual ($M = 7.08$, $SD = 2.81$) was low (less than 4% of all scenes on average). Notably, which individual scenes were "mis-rated" differed from one subject to another.

3.2. ERP results

3.2.1. "Time of interest" analysis

The amplitudes of the ERPs were averaged within the pre-determined time window of 300–500 ms. Responses to scenes containing incongruent objects were more negative than responses to scenes containing congruent objects in frontal and central electrodes within this time window (Fig. 4). This observation was confirmed by a three-way Region × Laterality × Congruity ANOVA, which revealed an interaction between Region and Congruity ($F(2,24) = 6.48$, $p < 0.05$, $\varepsilon = 0.621$), and between Laterality and Congruity ($F(2,24) = 4.27$, $p < 0.05$, $\varepsilon = 0.951$). Post-hoc contrasts indicated that the congruity effect (Incongruent–Congruent) was different in the parieto-occipital regions ($-0.16\ \mu$V) compared with the frontal regions ($-1.36\ \mu$V, $t(12) = 2.72$, $p < 0.05$) and the central

---

[2] Since ANOVAs diagnose additive differences between groups whereas differences in ERP data tend to be multiplicative, a statistical effect might indicate either a difference in amplitude or a difference in distribution. To remove this ambiguity the cells that yield a significant effect on any factor other than the electrodes are separately scaled by dividing the scores on the effect by the square root of their summed squares. A topographical difference is held to exist between the conditions if the ANOVA still yields a significant effect after normalizations.

Please cite this article in press as: Mudrik, L., et al. ERP evidence for context congruity effects during simultaneous object–scene processing. *Neuropsychologia* (2009), doi:10.1016/j.neuropsychologia.2009.10.011

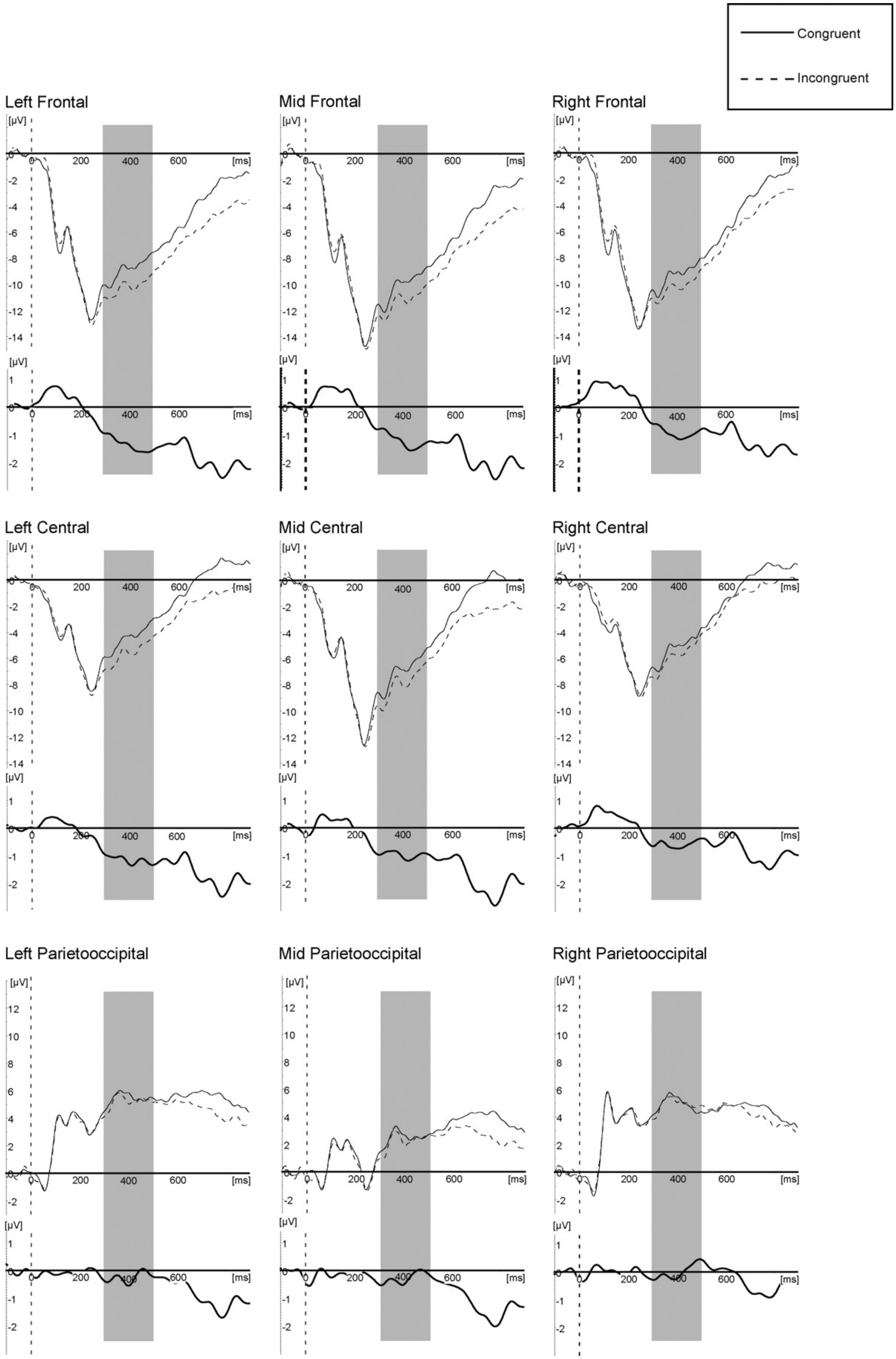

Fig. 4. Response to congruent (solid) and incongruent (dashed) images, for each region, and difference waves (thick waveforms below). Gray rectangles denote the 'time of interest' window of 300–500 ms.

regions (−0.96 μV, $t(12) = 3.11$, $p < 0.01$). In line with these findings, one sample *t*-tests confirmed that the congruity effect was significantly different from zero in the frontal ($t(12) = 2.29$, $p = 0.04$) and central regions ($t(12) = 2.40$, $p = 0.03$), but not in parieto-occipital regions.

Analysis of the Congruity × Laterality interaction also revealed that the congruity effects in both the left hemisphere (−1.01 μV) and the midline electrodes (−0.99 μV) were different than in the right hemisphere (−0.49 μV) ($t(12) = 2.57$, $p < 0.05$, $t(12) = 2.28$, $p < 0.05$, respectively). One sample *t*-tests confirmed that while the congruity effect was significantly different from zero in the left hemisphere ($t(12) = 2.58$, $p < 0.05$) and approached significance in midline electrodes ($t(12) = 2.03$, $p = 0.065$) it was not significant in the right hemisphere ($t(12) = 1.13$, $p = 0.46$). However, the observed interactions between Congruity and either Region or Laterality did not hold after vector-scaling normalization following McCarthy and Wood (1985) ($F(2,24) = 0.377$, $p = 0.61$, $\varepsilon = 0.673$, and $F(2,24) = 1.50$, $p = 0.25$, $\varepsilon = 0.844$, respectively), which suggests that the neural responses associated with the congruent vs. incongruent conditions may differ in their intensity rather than in their distribution.

3.2.2. *Exploratory analysis*

To assess the possibility of a congruity effect outside the predefined time window, a point-by-point *t*-test was conducted on the down-sampled (256 Hz) data. Fig. 5 depicts the distribution of the effect over time. Electrodes in which a significant effect was found are highlighted (two-tailed $p < 0.05$, corrected; see Section 2). The results confirm the significant incongruity effect between 300 and 500 ms described in the previous analysis: a significant negativity was observed at left frontal and central electrodes.

However, this analysis shows that the congruity effect in the left fronto-central region started at ∼270 ms (on the C5 electrode), and lasted until about 600 ms post-stimulus (Fig. 6). A later congruity effect was also seen from about 650 to about 840 ms post-stimulus, although the transition point between the two effects cannot be clearly determined. We tentatively suggest that this later effect was indeed a separate response, based on the observation that its spatial distribution was different—a much broader fronto-central-parietal negativity extending more also to right hemisphere electrodes (Fig. 5).

To test the possibility that the later response indeed involves a different set of sources, we defined two time windows: 270–600 and 650–850 ms, based on the results of the exploratory *t*-tests. Then we conducted a four-way ANOVA on the normalized data (McCarthy & Wood, 1985), with Time (early, late), Region (Frontal, Central, Parieto-occipital; Fig. 2), Laterality (Left, Midline, Right) and Congruity (Congruent, Incongruent). Significant interaction involving Time, Region and/or Laterality, and Congruity, would indicate an effect of time on the distribution of the congruity effect.

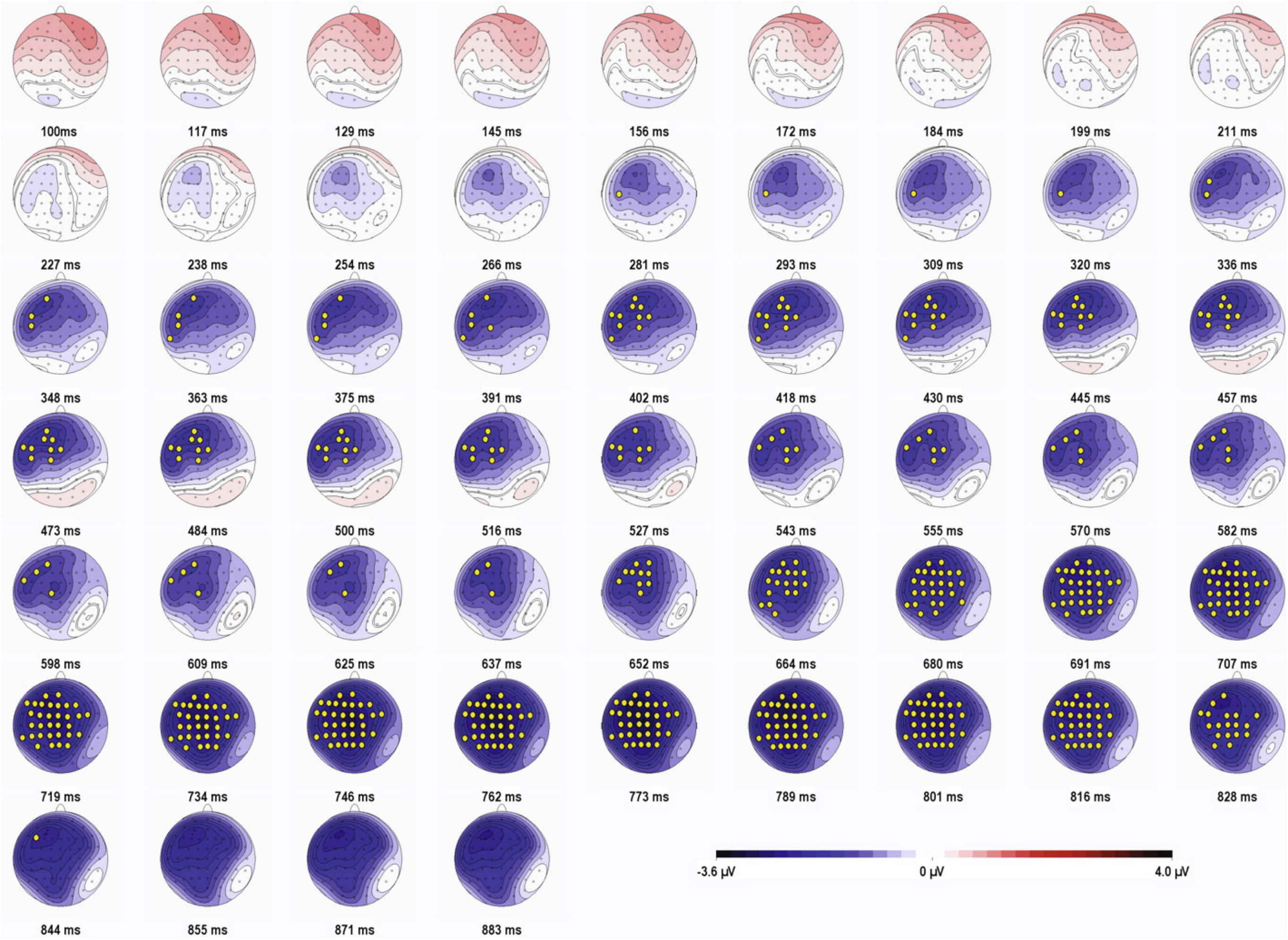

Fig. 5. Distribution maps of incongruency effect (incongruent–congruent) over time. Electrodes in which a significant effect ($p < 0.05$) was found at each time point are highlighted in yellow, with the constraint that the difference was significant for at least 20 consecutive data points. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of the article.)

Please cite this article in press as: Mudrik, L., et al. ERP evidence for context congruity effects during simultaneous object–scene processing. *Neuropsychologia* (2009), doi:10.1016/j.neuropsychologia.2009.10.011

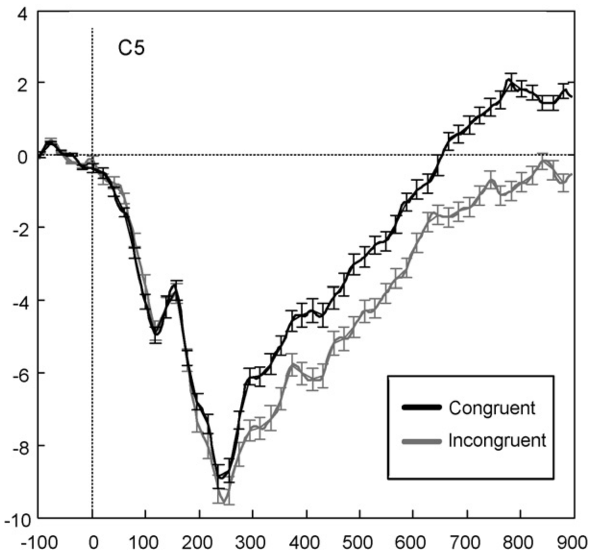

Fig. 6. Congruent and incongruent waveforms in electrode C5, with standard errors, across subjects ($n = 13$). Significant effect begins at ~270 ms.

Significant interactions were found between Time, Congruity, Region and Laterality ($F(4,48) = 5.22$, $p < 0.01$, $\varepsilon = 0.621$), between Time, Congruity and Region ($F(2,24) = 59.69$, $p < 0.001$, $\varepsilon = 0.574$) and between Time, Congruity and Laterality ($F(2,24) = 5.56$, $p < 0.05$, $\varepsilon = 0.840$). These interactions support the conclusion that the distribution of the congruency effect (as indicated by the two-way interactions between Congruity and Region and between Congruity and Laterality) differed between the 270–600 and the 650–850 ms time windows.

3.3. *Follow-up behavioral experiment*

This follow-up experiment was designed to test the possibility that despite the efforts taken in preparation of the incongruent scenes, subjects could nonetheless notice the digital ‘cut-and-paste’ operation. Whereas subjects easily discriminated between congruent and incongruent images ($d' = 2.67$, $t(7) = 17.67$, $p < 0.001$), they were not able to tell ‘pasted’ congruent images from natural congruent images above chance (mean $d' = 0.25$, not significantly different from zero, $t(7) = 0.64$, $p = 0.54$).

4. Discussion

Deciphering complex visual scenes is a formidable computational challenge. Yet, biological organisms seem to deal with this load quite well. Prior knowledge and expectations narrow the range of probable interpretations and thereby render scene analysis easier. Accordingly, scenes that conform to natural configurations are interpreted faster and more accurately, relative to scenes in which irregularities occur (e.g., Antes et al., 1981; Biederman, 1972; Biederman et al., 1982; Palmer, 1975). The present study investigated the neural correlates of context violation effects on scene processing. Scenes of objects being handled were presented when the object was congruent with the scene or incongruent with it. The event-related potentials discriminated congruent and incongruent scenes starting from 270 ms post-stimulus. The data further suggest that context violation processing is associated with two temporally distinct components with different scalp distributions, an earlier one at ~270–600 ms, and a later one at ~650–850 ms.

4.1. *Early negativity*

The early effect found in the present study replicates the negativity observed by Ganis and Kutas (2003) during the 300–500 ms time window and labeled N390. In that study, the context was presented 300 ms prior to the presentation of the object, so that observers, who were asked to actually judge the congruency of objects with their scenes, were likely to pre-form expectations even before the object had been seen. Thus, it remained possible that the effect seen at the 300–500 ms time window was due to a mismatch between perceptual predictions and actual input, rather than online interaction between scene and object processing. We presented the critical object and its context simultaneously rather than sequentially, and subjects could not form perceptual expectations with regard to the critical object, or prepare a naming response before the object appeared. Nevertheless, we replicated the N390 congruity effect. Importantly, although the context was presented 300 ms later than in the study of Ganis and Kutas, the congruity effect started at the same time, or even earlier, suggesting that the underlying process was independent of the time allowed for context processing (at least within the range tested across the experiments). Our results therefore confirm that the N390 component found by Ganis and Kutas genuinely reflects context violation, rather than a mismatch between expected vs. actual events, or between prepared vs. correct responses.

However, there are a few noteworthy differences between the N390 effect reported by Ganis and Kutas (2003) and the early effect found in our study. First, the effect emerged somewhat earlier in our study: it became significant around 270 ms post-stimulus (Figs. 4–6), whereas Ganis and Kutas’ study did not find a significant congruity effect prior to 300 ms. In fact, judging from the topographical distributions of the effect, it seems that the stable topographical pattern lasting until ~650 ms is already manifested at about 250 ms post-stimulus presentation, even if not reaching statistical significance at its onset (Fig. 5). These latencies are consistent with the findings of Sitnikova, Holcomb, Kiyonaga, and Kuperberg (2008, Experiment 1) reporting an “N300” incongruity effect starting around 250 ms from the presentation of an object violating goal-related action requirements (e.g., ironing a loaf of bread), quite akin to our scenes, albeit with prior context established, akin to Ganis and Kutas’ design. Similar latencies of an N300 congruency effect were found also with simple priming designs (e.g., McPherson & Holcomb, 1999) and with pictures at sentence endings (e.g., Federmeier & Kutas, 2001).

Second, the distribution of the current effect was much more frontal than that of the N390 reported by Ganis and Kutas. The effect found here had a fronto-central maximum, and was therefore typical of the neural activity associated with pictorial context violations (e.g., Federmeier & Kutas, 2001; Holcomb & McPherson, 1994). By contrast, the effect reported by Ganis and Kutas had a more central distribution and was not reliable at frontal sites. Third, the effect we observed seemed much more lateralized to the left than their N390 effect, in line with some (Holcomb & McPherson, 1994; Willems, Ozyurek, & Hagoort, 2008), but not all (e.g., Barrett & Rugg, 1990; Federmeier & Kutas, 2001; Sitnikova et al., 2008) previous studies of object incongruity effects.

Whereas our study was designed to investigate the N390 effect, we suggest that the early fronto-central negativity we found may reflect in fact the spatio-temporal overlap between successive N300-like negativities associated with perceptual object processing (with a typical time window of 250–350 ms; e.g., Barrett & Rugg, 1990; Federmeier & Kutas, 2001; Sitnikova et al., 2008) and N400-like negativities related to post-identification processing of more conceptual, semantic knowledge (N390, with a time window of 300–500 ms; Ganis & Kutas, 2003). A similar proposal has been put forward to account for the more anterior distribution observed

Please cite this article in press as: Mudrik, L., et al. ERP evidence for context congruity effects during simultaneous object–scene processing. *Neuropsychologia* (2009), doi:10.1016/j.neuropsychologia.2009.10.011

for context violations with pictures relative to words (Schendan & Maher, 2008; Sitnikova et al., 2008). Quite likely, activation of semantic representations starts before the model selection has been fully resolved (e.g., Bar, 2004; Friedman, 1979; Ullman, 1996), resulting in process overlap.

N300 congruity effect, referred to also as N350 (e.g., Ganis & Kutas, 2003; Schendan & Kutas, 2003, 2007; Schendan & Maher, 2008) or Ncl (Doniger et al., 2000) is held to index object model selection via matching routines (Schendan & Kutas, 2002, 2003). By this account, scene–congruent object representations are activated after the gist of the scene has been rapidly grasped (Biederman et al., 1974; Carr et al., 1982; Hochstein & Ahissar, 2002; Intraub, 1981; Mccauley et al., 1980; Oliva & Schyns, 1997, 2000; Oliva & Torralba, 2006). Then, search for the matching object representation is affected by the discrepancy between the activated representations and the upcoming information about the incongruent object, leading to repeated and sometimes failed attempts to reach identification. Indeed, in tasks requiring the subjects to identify ambiguous objects (e.g., fragmented line drawings or objects presented from an unusual point of view), N300 is enhanced for unidentified objects, or non-objects, relative to identified ones (Doniger et al., 2000; Folstein, Van Petten, & Rose, 2008; McPherson & Holcomb, 1992; Pietrowsky et al., 1996; Schendan & Kutas, 2002, 2003; Vannucci et al., 2006). Additionally, N300 is also modulated by semantic congruity: its amplitude is greater when the second object in a pair of sequentially presented objects was unidentifiable or semantically unrelated to the first one, relative to when the two objects were related (Barrett & Rugg, 1990; Holcomb & McPherson, 1994; McPherson & Holcomb, 1999).

The exact timing of object model selection onset may change dramatically, depending on the images used. Its onset was reported as early as 175–192 ms post-stimulus (Holcomb & McPherson, 1994; McPherson & Holcomb, 1999) and as late as 316 ms (Schendan & Maher, 2008), and its duration increases with matching difficulty (Schendan & Kutas, 2002, 2003). The influence of context violations on model selection processes might explain why incongruent objects draw attention, while still being identified later than congruent ones (Bar & Ullman, 1996; Boyce & Pollatsek, 1992; Chun & Jiang, 1998; Davenport & Potter, 2004; Neider & Zelinsky, 2006; Palmer, 1975; Rieger et al., 2008), in line with the suggestion made by Underwood and his colleagues (2008). If context influences object identification processes, a non-fixated incongruent object will require further inspection, and draw attention even before its identification is complete.

While being cautious in interpreting their (null) result showing no congruity effect prior to 300 ms, Ganis and Kutas (2003) conjectured that N300 could have either been delayed in their study because of the added effort of identifying a small object in a scene relative to in isolation, or that not enough time was given between the presentation of the context and the object to affect object recognition processes (and thus the N300), or that the high variability of the objects presented in their study, prevented precise predictions that could have affected object recognition or matching processes. However, the objects in our study were also presented in complex scenes, were also highly variable, and there was even shorter time to analyze the scene, as scenes and objects were presented simultaneously. Yet a negativity conforming with previous reports of an “N300” congruity effect, including a fronto-central distribution and onset around 250 ms, was found. In line with the above interpretation of the N300, these findings suggest that object model selection stages may be affected by the surrounding scene even during simultaneous presentation.

The interpretation of the initial part of our early component as representing an N300 component, involving object identification stages, suffers from a design limitation of our study, which was not planned to test early visual effects. Objects and scenes were not fully counterbalanced in the experiment: while all backgrounds served both as congruent and as incongruent contexts, different objects appeared in congruent and in incongruent scenes. In addition, incongruent scenes were created by digitally pasting a new object into a scene, whereas congruent images were not digitally manipulated. This digital manipulation was done with special care to eliminate traces of the ‘photomontage’ (see Section 2) and incongruent objects were randomly chosen from any category with the only requirement that they should match the space occupied by the original object. Nevertheless, it may be claimed that the earliest difference in neural activity (∼270 ms post-stimulus) between congruent and incongruent scenes may be related to processing of low-level visual peculiarities of the incongruent scenes rather than to semantic incongruity.

Although we cannot categorically overrule this concern in the present design, it is mitigated by several arguments. We analyzed the contrast, luminance, chromaticity and spatial frequency of the congruent and incongruent images using physiologically based models (Neumann & Gegenfurtner, 2006) and found no difference between the two groups of pictures. We further examined the low-level perceptual saliency of objects using a recent computational model (Itti & Koch, 2000) and found no difference. Finally, we conducted a behavioral follow-up experiment to investigate the possibility that subjects may perceive the digital processing (pasting) of objects based on parameters not captured by either of these analytic approaches. Naïve subjects were asked to explicitly discriminate natural congruent scenes and scenes in which a new object was pasted; this object was congruent with the scene but copied from another picture. The pasting procedure was similar to that used to create the incongruent images. Only congruent images were used in this follow-up experiment to dissociate the semantic incongruence from the visual manipulation procedure. The results showed that under the same conditions that allowed subjects to confidently decide whether the scene was congruent or not, and the same conditions under which the ERP effects were observed, subjects could not detect the digital manipulation above chance. Thus, although this follow-up experiment did not test the original pasted images (as they were all incongruent, see above), it seems safe to conclude that our procedures did not allow observers to detect the graphical manipulation. Future studies designed to address this early effect should fully counterbalance the conditions to obviate these post-hoc considerations. Nevertheless, it seems that low-level factors were not likely to elicit the N300 effect in our study.

4.2. *Late negativity*

Between 650 and 850 ms post-stimulus, incongruent images yielded a widespread negativity. A similar late negativity has been previously reported in response to stimuli that violated a previously formed context (Barrett & Rugg, 1990; Ganis & Kutas, 2003; Holcomb & McPherson, 1994; Sitnikova et al., 2008; Sitnikova et al., 2003), to unidentified objects (Schendan & Kutas, 2002; Schendan & Maher, 2008), to unnamable pseudo-objects (Holcomb & McPherson, 1994; McPherson & Holcomb, 1999), or to unstudied objects (Schendan & Kutas, 2003, 2007). This late effect is held to reflect secondary processes that are influenced by context violations, such as inferential procedures that allocate attention, the activation of semantic knowledge and names, or the internal evaluation of the success of a category decision (Schendan & Kutas, 2002; Schendan & Maher, 2008). It was also suggested that the late anterior negativity seen ∼700 ms post-stimulus presentation indexes working memory related to selecting, preparing, and monitoring a verbal report or response preparation processes that are more challenging when response selection conflict is greater (Schendan & Maher, 2008). In our study, it was more difficult to answer the

Please cite this article in press as: Mudrik, L., et al. ERP evidence for context congruity effects during simultaneous object–scene processing. *Neuropsychologia* (2009), doi:10.1016/j.neuropsychologia.2009.10.011

question about the number of hands used in the incongruent relative to the congruent condition probably because it was harder to recognize the action taken with an incongruous object.

Taken together, the results of the current study confirmed that context violations in a scene yield an N400-like component (N390), arguably related to semantic activation by recognized objects, followed by a late negativity related to response conflict. Our study showed conclusively that such components appear not only when expectations are set prior to object presentation, as previously shown, but also when the object and its context are presented simultaneously, suggesting online interaction between scene and object processing. The results also imply that online context violations might affect not only post-identification processes at the semantic level, indexed by N400-like effects, but also earlier routines of matching between possible scene-congruent activated schemas and the upcoming information about the actual object, indexed by N300. The latter conclusion awaits verification by studies designed to conclusively rule out visual differences between congruent and incongruent images.

Acknowledgement

The study was supported by a grant from the National Institute for Psychobiology in Israel, founded by the Charles E. Smith family, to L.Y.D.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.neuropsychologia.2009.10.011.

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.

Auckland, M. E., Cave, K. R., & Donnelly, N. (2007). Nontarget objects can influence perceptual processes during object recognition. *Psychonomic Bulletin & Review*, *14*(2), 332–337.

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.

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

Biederman, I. (1972). Perceiving real-world scenes. *Science*, *177*(43), 77–80.

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 sciences. *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.

Biederman, I., Teitelbaum, R. C., & Mezzanotte, R. J. (1983). Scene perception—A failure to find a benefit from prior expectancy or familiarity. *Journal of Experimental Psychology-Learning Memory and Cognition*, *9*(3), 411–429.

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.

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

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.

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.

Delorme, A., Richard, G., & Fabre-Thorpe, M. (2000). Ultra-rapid categorisation of natural scenes does not rely on colour cues: A study in monkeys and humans. *Vision Research*, *40*(16), 2187–2200.

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

Fabre-Thorpe, M., Delorme, A., Marlot, C., & Thorpe, S. (2001). A limit to the speed of processing in ultra-rapid visual categorization of novel natural scenes. *Journal of Cognitive Neuroscience*, *13*(2), 171–180.

Fabre-Thorpe, M., Richard, G., & Thorpe, S. J. (1998). Rapid categorization of natural images by rhesus monkeys. *Neuroreport*, *9*(2), 303–308.

Federmeier, K. D., & Kutas, M. (2001). Meaning and modality: Influences of context, semantic memory organization, and perceptual predictability on picture processing. *Journal of Experimental Psychology-Learning Memory and Cognition*, *27*(1), 202–224.

Federmeier, K. D., & Kutas, M. (2002). Picture the difference: Electrophysiological investigations of picture processing in the two cerebral hemispheres. *Neuropsychologia*, *40*(7), 730–747.

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.

Ganis, G., Kutas, M., & Sereno, M. I. (1996). The search for “common sense”: An electrophysiological study of the comprehension of words and pictures in reading. *Journal of Cognitive Neuroscience*, *8*(2), 89–106.

Guthrie, D., & Buchwald, J. S. (1991). Significance testing of difference potentials. *Psychophysiology*, *28*(2), 240–244.

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., & Hollingworth, A. (1999). High-level scene perception. *Annual Review of Psychology*, *50*, 243–271.

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.

Hochstein, S., & Ahissar, M. (2002). View from the top: Hierarchies and reverse hierarchies in the visual system. *Neuron*, *36*(5), 791–804.

Holcomb, P. J. (1993). Semantic priming and stimulus degradation—Implications for the role of the N400 in language processing. *Psychophysiology*, *30*(1), 47–61.

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

Hollingworth, A., & Henderson, J. M. (2000). Semantic informativeness mediates the detection of changes in natural scenes. *Visual Cognition*, *7*, 213–235.

Intraub, H. (1981). Rapid conceptual identification of sequentially presented pictures. *Journal of Experimental Psychology-Human Perception and Performance*, *7*(3), 604–610.

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.

Joubert, O. R., Rousselet, G. A., Fize, D., & Fabre-Thorpe, M. (2007). Processing scene context: Fast categorization and object interference. *Vision Research*, *47*(26), 3286–3297.

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.

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 between the lines: Event-related brain potentials during natural sentence processing. *Brain and Language*, *11*(2), 354–373.

Kutas, M., & Hillyard, S. A. (1980c). Reading senseless sentences—Brain potentials reflect semantic incongruity. *Science*, *207*(4427), 203–205.

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.

Loftus, G. R., Nelson, W. W., & Kallman, H. J. (1983). Differential acquisition rates for different types of information from pictures. *Quarterly Journal of Experimental Psychology Section A-Human Experimental Psychology*, *35*(FEB), 187–198.

Please cite this article in press as: Mudrik, L., et al. ERP evidence for context congruity effects during simultaneous object–scene processing. *Neuropsychologia* (2009), doi:10.1016/j.neuropsychologia.2009.10.011

Luce, R. D. (1986). *Response times*. Oxford: Oxford University Press.

McCarthy, G., & Wood, C. C. (1985). Scalp distributions of event-related potentials—An ambiguity associated with analysis of variance models. *Electroencephalography and Clinical Neurophysiology*, *62*(3), 203–208.

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

McPherson, W. B., & Holcomb, P. J. (1992). Semantic priming with pictures and the N400 component. *Psychophysiology*, *29*, S51.

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

Metzger, R. L., & Antes, J. R. (1983). The nature of processing early in picture perception. *Psychological Research-Psychologische Forschung*, *45*(3), 267–274.

Neider, M. B., & Zelinsky, G. J. (2006). Scene context guides eye movements during visual search. *Vision Research*, *46*(5), 614–621.

Neumann, D., & Gegenfurtner, K. (2006). Image retrieval and perceptual similarity. *ACM Transactions on Applied Perception (TAP)*, *3*(1), 31–47.

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

Oliva, A., & Schyns, P. G. (1997). Coarse blobs or fine edges? Evidence that information diagnosticity changes the perception of complex visual stimuli. *Cognitive Psychology*, *34*(1), 72–107.

Oliva, A., & Schyns, P. G. (2000). Diagnostic colors mediate scene recognition. *Cognitive Psychology*, *41*(2), 176–210.

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.

Palmer, S. E. (1975). Effects of contextual scenes on identification of objects. *Memory & Cognition*, *3*(5), 519–526.

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.

Pietrowsky, R., Kuhmann, W., Krug, R., Molle, M., Fehm, H. L., & Born, J. (1996). Event-related brain potentials during identification of tachistoscopically presented pictures. *Brain and Cognition*, *32*(3), 416–428.

Rayner, K., & Pollatsek, A. (1992). Eye movements and scene perception. *Canadian Journal of Psychology*, *46*(3), 342–376.

Rieger, J. W., Braun, C., Bulthoff, H. H., & Gegenfurtner, K. R. (2005). The dynamics of visual pattern masking in natural scene processing: A magnetoencephalography study. *Journal of Vision*, *5*(3), 275–286.

Rieger, J. W., Kochy, N., Schalk, F., Gruschow, M., & Heinze, H. J. (2008). Speed limits: Orientation and semantic context interactions constrain natural scene discrimination dynamics. *Journal of Experimental Psychology-Human Perception and Performance*, *34*(1), 56–76.

Rousselet, G. A., Joubert, O. R., & Fabre-Thorpe, M. (2005). How long to get to the "gist" of real-world natural scenes? *Visual Cognition*, *12*(6), 852–877.

Rousselet, G. A., Mace, M. J., & Fabre-Thorpe, M. (2003). Is it an animal? Is it a human face? Fast processing in upright and inverted natural scenes. *Journal of Vision*, *3*(6), 440–455.

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.

Schyns, P. G., & Oliva, A. (1994). From blobs to boundary edges—Evidence for time-scale-dependent and spatial-scale-dependent scene recognition. *Psychological Science*, *5*(4), 195–200.

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.

Sitnikova, T., Kuperberg, G., & Holcomb, P. J. (2003). Semantic integration in videos of real-world events: An electrophysiological investigation. *Psychophysiology*, *40*(1), 160–164.

Ullman, S. (1996). *High-level vision: Object recognition and visual cognition*. Cambridge, MA: MIT Press.

Underwood, G. (2005). Eye fixations on pictures of natural scenes: Getting the gist and identifying the components. In G. Underwood (Ed.), *Cognitive Processes in Eye Guidance* (pp. 163–187). Oxford: Oxford University Press.

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.

Vannucci, M., Grunwald, T., Pezer, N., Dietl, T., Helmstaedter, C., Schaller, C., et al. (2006). Hippocampus proper distinguishes between identified and unidentified real-life visual objects: An intracranial ERP study. *Neuroscience Letters*, *401*(1–2), 165–170.

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.

West, W. C., & Holcomb, P. J. (2002). Event-related potentials during discourse-level semantic integration of complex pictures. *Cognitive Brain Research*, *13*(3), 363–375.

Willems, R. M., Ozyurek, A., & Hagoort, P. (2008). Seeing and hearing meaning: ERP and fMRI evidence of word versus picture integration into a sentence context. *Journal of Cognitive Neuroscience*, *20*(7), 1235–1249.

Please cite this article in press as: Mudrik, L., et al. ERP evidence for context congruity effects during simultaneous object–scene processing. *Neuropsychologia* (2009), doi:10.1016/j.neuropsychologia.2009.10.011