Cerebral Cortex, July 2020;30: 4158–4168

doi: 10.1093/cercor/bhaa040
Advance Access Publication Date: 20 March 2020
Original Article

ORIGINAL ARTICLE

Automatic Attention Capture by Threatening, But Not by Semantically Incongruent Natural Scene Images

Marcin Furtak[1], Łucja Doradzińska[1], Alina Ptashynska[1], Liad Mudrik[2,3], Anna Nowicka[4] and Michał Bola[1]

[1]Laboratory of Brain Imaging, Nencki Institute of Experimental Biology, 02-093 Warsaw, Poland, [2]School of Psychological Science, Tel Aviv University, 69978 Tel Aviv, Israel, [3]Sagol School of Neuroscience, Tel Aviv University, 69978 Tel Aviv, Israel and [4]Laboratory of Language Neurobiology, Nencki Institute of Experimental Biology, 02-093 Warsaw, Poland

Address correspondence to: Michał Bola, Laboratory of Brain Imaging, Nencki Institute of Experimental Biology, 3 Pasteur Street, 02-093 Warsaw, Poland. Email: m.bola@nencki.edu.pl

Abstract

Visual objects are typically perceived as parts of an entire visual scene, and the scene's context provides information crucial in the object recognition process. Fundamental insights into the mechanisms of context-object integration have come from research on semantically incongruent objects, which are defined as objects with a very low probability of occurring in a given context. However, the role of attention in processing of the context-object mismatch remains unclear, with some studies providing evidence in favor, but other against an automatic capture of attention by incongruent objects. Therefore, in the present study, 25 subjects completed a dot-probe task, in which pairs of scenes—congruent and incongruent or neutral and threatening—were presented as task-irrelevant distractors. Importantly, threatening scenes are known to robustly capture attention and thus were included in the present study to provide a context for interpretation of results regarding incongruent scenes. Using N2 posterior-contralateral ERP component as a primary measure, we revealed that threatening images indeed capture attention automatically and rapidly, but semantically incongruent scenes do not benefit from an automatic attentional selection. Thus, our results suggest that identification of the context-object mismatch is not preattentive.

Key words: attention capture, incongruent objects, N2pc, natural scenes

Introduction

The question of which features automatically attract visual attention has been intensely debated for decades. Fundamental insights into the mechanisms of attentional selection have been provided by studies using arrays of mutually independent stimuli that are defined by simple features, like color or orientation (review: Wolfe and Horowitz 2017). In sharp contrast, objects in natural scenes are complex and defined by many features, and—importantly—they are always embedded within a general "gist" of a scene and occur in relation to other objects (Bar 2004; Peelen and Kastner 2014; Kaiser et al. 2019; Võ et al. 2019). Therefore, there is a growing interest in investigating mechanisms involved specifically in perception of such highly structured, naturalistic stimuli (e.g., Henderson and Hollingworth 1999; Epstein et al. 2003; Mack 2003; Cohen et al. 2016). Most importantly, the regularities present in natural scenes give rise to expectations regarding location and identity of objects and thus guide our exploration of the environment beyond detection of simple features (for review, see Oliva and Torralba 2007; Wolfe et al. 2011).

One way to learn about such context-based mechanisms is to investigate perception of semantically incongruent scenes, which include elements with very low probability of appearing in a given context and thus violate expectations of an observer

(Biederman et al. 1982). Indeed, because incongruent objects do not benefit from contextual facilitation, their recognition is impaired with respect to both speed and accuracy (Boyce et al. 1989; Davenport and Potter 2004; Rieger et al. 2008). Nevertheless, despite the importance of context-object interactions for vision (Bar 2004; Kaiser et al. 2019), the role of attention in this process is still a matter of debate. More specifically, two key questions have attracted substantial interest: is attention captured by incongruent objects, and is it engaged by them? A positive answer to either question will suggest that attention is preferentially allocated to incongruent objects. Yet, while the former implies that incongruencies might be processed preattentively (so to enable attentional capture), the latter does not, as only when an incongruent object is incidentally viewed it engages (holds) attention for a longer time.

Accordingly, while there is strong evidence that incongruent objects engage attention (e.g., Võ and Henderson 2009, 2011; Mudrik et al. 2011; Cornelissen and Võ 2017), a controversy evolved around the attentional capture. Supporting captures are two main sources of evidence. First, several eye-tracking studies suggest that observers direct initial saccades toward incongruent objects (Loftus and Mackworth 1978; Underwood and Foulsham 2006; Becker et al. 2007; Underwood et al. 2007, 2008; Bonitz and Gordon 2008). Second, studies using a change blindness paradigm indicate higher detection rates of a change when the key object is incongruent, even though recognition of the object is impaired (Hollingworth and Henderson 2000; LaPointe et al. 2013; Mack et al. 2017; LaPointe and Milliken 2017; Ortiz-Tudela et al. 2017, 2018).

On the other hand, several other eye-tracking studies found no evidence that incongruent objects preferentially attract early fixations, but only that they are scrutinized for a longer time once fixated, thus suggesting greater engagement (De Graef et al. 1990; Gareze and Findlay 2007; Rayner et al. 2009; Võ and Henderson 2009, 2011; Cornelissen and Võ 2017). Similarly, Mudrik et al. (2011) found that once the incongruent scene has been perceived in a binocular rivalry paradigm, subjects exhibit difficulty with disengaging attention, but no evidence for a preferential selection of such scenes. Further, incongruent scenes do not break the continuous flash suppression faster than congruent ones, which again suggests no attention capture by a context-object mismatch (i.e., Moors et al. 2016; see also Biderman and Mudrik 2018). Finally, a recent study of Mack et al. (2017) found no evidence for an automatic capture of attention by semantically incongruent scenes across three behavioral paradigms.

Given these conflicting results, here we aimed to directly test whether incongruent objects are able to attract attention automatically. To this end, we used a procedure developed by Kappenman et al. (2015), who showed an attention capture by threatening scene images included in the International Affective Picture System (IAPS) stimulus set (Lang et al. 2008). This proves the procedure is sensitive to detect attentional capture by complex, real-life scenes. The procedure involves a dot-probe task, which is a classic task to study automatic shifts of spatial attention (MacLeod et al. 1986). In this task, pairs of distractor stimuli (attentional cues) are presented laterally and followed by a target stimulus displayed on one of the sides. Subjects are instructed to maintain their gaze on the centrally presented fixation cross, ignore the distractors, and manually respond to the presentation of a target. However, if one of the distractors exhibits some attention-grabbing properties, then attention will be automatically directed laterally toward such a stimulus, and this effect can be uncovered by analysis of reaction times (RTs) of responses to targets or by the N2 posterior-contralateral (N2pc) ERP component. N2pc was the primary index of attention in the Kappenman et al. (2015) and in the present study. It is defined as greater negativity at posterior electrodes contralateral to the visual field of an attended visual stimulus relative to the voltage at corresponding ipsilateral electrodes observed in the ERP response. N2pc exhibits a posterior scalp distribution, with a maximum typically in the P7/P8 locations. The majority of researchers agree that N2pc indicates differential processing of stimuli in one visual field with respect to the other and that the mechanism involves covert shifts of spatial attention (Eimer 1996; Kiss et al. 2008). Importantly, N2pc was treated as a primary measure because ERPs provide a continuous index of stimulus processing and thus might reveal transient and covert shifts of attention (unlike behavioral measures, which provide an aggregate index of a whole chain of processing stages, e.g., perceptual, cognitive, motor). Indeed, previous analyses of the dot-probe task data, including studies by Kappenman et al. (2014, 2015), indicate that the behavioral RT effect exhibits poor internal reliability (and, by consequence, its external validity must be also poor; see also Schmukle 2005), whereas the reliability of the N2pc effect is moderate.

Therefore, using the outlined procedure, we aimed to realize three specific goals. First, to replicate the main result of Kappenman et al. (2015), namely, that threatening scenes automatically capture attention as indicated by the N2pc component. Successful replication of their finding would indicate that the paradigm established in our experimental setting is indeed sensitive enough. Interestingly, to detect a threat it is typically sufficient to identify either a gist or a key object. But to detect incongruency, it is necessary to recognize both the gist of a scene and identity of a key object and then to integrate them to establish a mismatch. From this perspective, threatening scenes provide an interesting context for the incongruent scenes, and investigating both types might help to establish which features of a scene can be identified preattentively. Second, we aimed to extend the results of Kappenman et al. (2015) by testing whether threatening scenes capture attention also when presented briefly (for 100 ms, instead of 500 ms). Previous behavioral dot-probe studies suggest that processing and attentional prioritization of such scenes is indeed rapid and efficient (Koster et al. 2005; Cooper and Langton 2006), and thus here we aimed to provide electrophysiological evidence for this claim. Finally, the main research question of this study focuses on attentional capture by incongruent scenes; considering the lack of consensus around such a capture effect, the same procedure was used to investigate whether incongruent scenes automatically capture attention or not. For that, we used a set of congruent and incongruent scenes developed by Mudrik et al. (2010), which were also displayed for a shorter (100 ms) or longer time (500 ms).

To investigate automatic attention capture by threatening and semantically incongruent scene images, we analyzed two behavioral measures (accuracy and reaction time, RT) and a lateralized N2pc ERP component obtained in the dot-probe task. We specifically expected 1) higher accuracy when target dots followed a threatening/incongruent scene; 2) shorter RT when target dots followed a threatening/incongruent scene; and 3) lower amplitude of the contralateral ERP waveform in comparison to ipsilateral waveform (i.e., presence of N2pc) at electrodes P7/P8 in the time window 175–225 ms with respect to the stimulus onset (as reported by Kappenman et al. 2015). Further, in our study we also included an identification task to test how well subjects were able to recognize that a scene is threatening

or incongruent. We thus analyzed sensitivity index (d′) and confidence ratings and expected better identification (higher d′) and higher subjective confidence for threatening images (in contrast to incongruent) and for images presented for a longer time (irrespective of the type).

Methods

Subjects

We collected data of 25 subjects (16 females, mean age = 23.7 years, standard deviation [SD] = 3.8 years, range: 19–34 years, 3 left-handed). They all declared normal or corrected-to-normal vision and no history of mental or neurological disorders. The sample size was defined based on the Bayes factor (BF) calculated for a critical N2pc comparison (details described in the Statistical Analysis section). Specifically, we collected data until BF for these comparisons exceeded 0.1 or 10 in all four conditions, providing strong evidence in favor of, respectively, null or alternative hypothesis.

Data of 7 additional subjects were collected but excluded from the analysis: data of 2 subjects has not been properly saved, electrooculographic (EOG) signals of 2 subjects were too noisy to use it in the analysis, and 3 subjects were excluded due to insufficient number of epochs remaining after EEG signal preprocessing (detailed criteria are described in the EEG Recording and Analysis section). Further, data of one subject had to be excluded from the identification task (but not the dot-probe task) analyses due to a technical problem with the procedure (thus, the final sample for the identification task is 24 subjects), but his/her data were included in the dot-probe analyses.

All experimental procedures were approved by the local Research Ethics Committee at the Faculty of Psychology, University of Warsaw. All subjects provided written informed consent and received monetary compensation for their time (100 PLN = c.a. 25 EUR).

Stimuli

Two sets of stimuli were used. The first was a subset of the International Affective Picture System stimulus set (Lang et al. 2008), comprising of 50 neutral and 50 threatening images used in the study of Kappenman et al. (2015). Neutral images presented scenes without an emotional content (e.g., a man with a newspaper), while threatening images showed unpleasant or disturbing scenes, conveying negative emotional content (e.g., cockroach, attacking dog, mutilated bodies). The identification numbers of the images used are provided in Kappenman et al.'s (2015) publication.

The second set was a subset of 50 pairs of scenes developed by Mudrik et al. (2010). One version of each scene presents a person performing an action with an object, which is highly probable in a given context (congruent version, e.g., a man playing a violin), whereas the other version the key object has a very low probability of occurring in a given context (incongruent version, e.g., a man "playing" a broomstick). In both versions, the critical object has been pasted onto the scene.

The low-level (physical) properties of the images were calculated using the Python Imaging Library (http://www.pythonware.com/products/pil/) and compared statistically using Bayesian independent samples *t*-tests. We found moderate evidence that threatening IAPS images did not differ from neutral IAPS images in terms of luminance (BF = 0.214), contrast (BF = 0.211), and entropy (BF = 0.218). Similarly, we observed moderate evidence that incongruent scenes did not differ from congruent scenes, neither in terms of luminance (BF = 0.222) or contrast (BF = 0.258) nor entropy (BF = 0.213; as reported by Mudrik et al. 2010). This indicates that neutral and congruent images constitute a good control for, respectively, threatening and incongruent images and that attention shifts cannot be ascribed to differences in physical properties within presented pairs. However, comparisons between the stimuli sets (i.e., all neutral/threatening vs. all congruent/incongruent) indicate that the neutral/threatening set exhibits lower luminance (BF = 734, extreme evidence), higher contrast (BF = 79, very strong evidence), and lower entropy (BF = 18, strong evidence) than the congruent/incongruent images. However, our main interest lies in comparing the two categories within each set, and thus this difference does not affect the comparisons of interest.

Procedure

The experimental procedure was written in the Presentation software (Neurobehavioral Systems, Albany, CA) and presented on a FlexScan EV-2450 computer monitor. The viewing distance was 60 cm and it was maintained by a chinrest.

Dot-Probe Task

The dot-probe task consisted of 4 blocks (Fig. 1). Each block was specified by a combination of 2 factors: "stimulus type" (neutral/emotional IAPS images or congruent/incongruent scene images; i.e., IAPS images were not presented with congruent/incongruent images in the same block) and "presentation time" (100 or 500 ms). Each block included 400 trials (thus, each stimulus was presented 8 times) and was further subdivided into 4 series (100 trials each), between which subjects had a self-paced break. Blocks were presented in a semifixed order, that is, the initial block was chosen randomly, and it was followed by a second block, in which the same stimulus type was used but with a different presentation time. Next, two blocks with the other stimulus type were presented, where the order of blocks in terms of a presentation time was the same as in the first two blocks. The order of trials within each block was fully random.

The dot-probe procedure was the same as the one used by Kappenman et al. (2015). First, a fixation cross (subtending 0.4° × 0.4° of the visual angle) appeared at the center of the screen for 500 ms. Then, the stimuli were presented for either 100 or 500 ms. In each trial, the stimuli were either a congruent and an incongruent image or a neutral and a threatening image (depending on the block). In the former case, two images from the same pair were never presented together (i.e., the congruent image was always from a different pair than the incongruent image). The side in which the threatening/incongruent scene was presented was balanced for all blocks (i.e., in each block they were presented 200 times on the left, and 200 times on the right side, randomly intermixed). The center of each image was always located 4.8° laterally from the center of the screen, but the two sets of images differed in orientation: the images in the IAPS set are horizontal (images size, 6.7° horizontal × 4.8° vertical of the visual angle), while the congruent/incongruent scenes are vertical (images size, 4.8° horizontal × 6.7° vertical). The two images were followed by two target asterisks (0.3° of the visual angle away from each other, with each subtending 0.3° × 0.3° of the visual angle), presented with their center of gravity 4.8° laterally (i.e., in the location of the center of a scene) for 400 ms. In half the trials, the dots were in a horizontal orientation and,

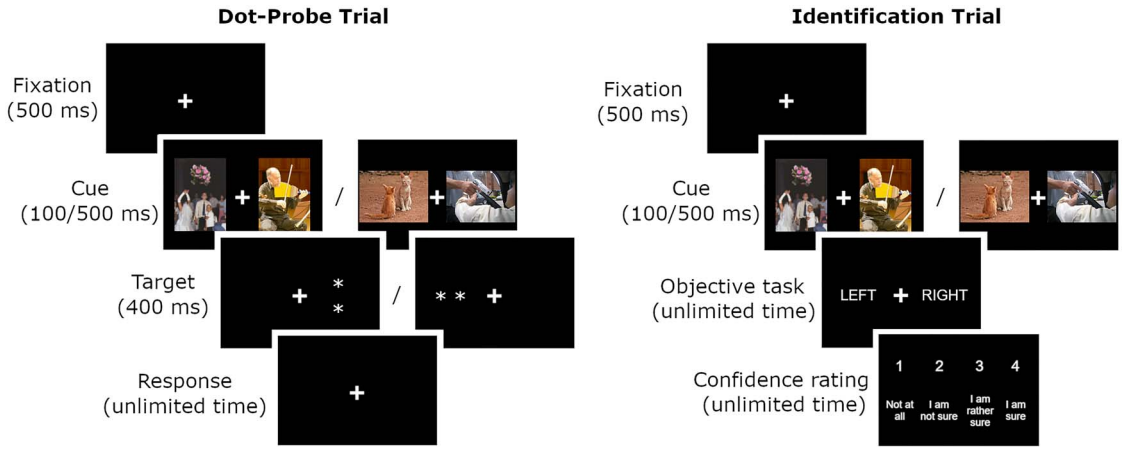

Figure 1. Experimental procedure. Subjects performed a dot-probe procedure in which their task was to indicate with a button press an orientation of two target dots. Dots were preceded by pairs of scenes, which subjects were supposed to ignore. A threatening/incongruent scene was presented on one side and a neutral/congruent scene on the other side. Subjects also completed an identification task, in which they had to first indicate the side of a threatening/incongruent scene presentation and then to rate confidence they had in their decision.

in the other, in a vertical one. For each orientation in half of the trials, dots were presented on the threatening/incongruent image side and in the other half on the neutral/congruent image side. Subjects were asked to respond by pressing one of the two buttons using index fingers of their left or right hand—whether the dots were in horizontal or vertical orientation. Subjects were asked to respond as quickly and accurately as possible. The response time was unlimited and the next trial started after the manual response.

Identification Task

To test the ability to identify the threatening and incongruent scenes, each subject completed an identification task (which was not used in the study of Kappenman et al. 2015). The identification procedure was always performed before the dot-probe task, as we did not want the recognition rates to be biased (i.e., improved) by eight presentations of each image in the dot probe. The identification part also consisted of four blocks, and the order of which was the same as in the dot-probe task. However, each block included 100 trials (thus, each stimulus was presented 2 times). The order of trial presentation within each block was fully random.

The trial structure was the same as in the dot-probe task, except target dots were not presented (Fig. 1). Immediately after the two scene images subjects were asked to indicate on which side the "emotional" (threatening) or "weird" (incongruent) scene was presented by pressing a button either on the left or on the right side of the response pad. Next subjects were asked to rate how sure they were that their answer was correct using a scale: 1—"not sure at all" 2—"not sure" 3—"rather sure" 4—"absolutely sure" and pressing one of the four response pad buttons. Before the procedure was started, subjects were shown printed versions of a few "emotional" and "weird" (incongruent) scenes, which were not part of the subsets used in the experiment.

Analysis of Behavioral Data

All analyses of behavioral data were conducted using custom-made Python scripts. Analysis of the dot-probe task data focused on establishing whether accuracy and RTs of a manual response to the target dots differ between two types of trials: those in which dots followed the potentially attention-grabbing stimulus (threatening or incongruent scene) and those in which dots followed the neutral stimulus (neutral or congruent scene). Only trials in which RT did not exceed 1500 ms were used in the analysis. Accuracy was calculated as a percentage of correct responses to the dot orientation. Accuracy of one subject was close to 10%, likely as a result of wrong orientation to response mapping; thus, the mapping was inverted during the analysis. RTs were calculated only for the correct responses.

For the identification task, d′ (sensitivity index) and subjective confidence ratings were analyzed. Subjects were informed that response time is unlimited and thus trials were not excluded based on the response time. d′ was calculated to evaluate whether subjects were able to distinguish between the target (incongruent/threatening image, which they were supposed to recognize) and the "noise" stimuli (congruent/neutral). Hits and false alarms equal to 0 or 1 for each subject were replaced using the log-linear rule, the least biased method of correcting extreme values (Stanislaw and Todorov 1999). Concerning confidence ratings, we first calculated the percentage of trials in which a given level of confidence was reported. In the statistical analysis, we compared the percentage of trials in which subjects rated their confidence as high ("I was rather sure" or "I was sure" responses) between conditions.

EEG Recording and Analysis

During the experiment, EEG signal was recorded with 64 Ag–AgCl electrically shielded electrodes mounted on an elastic cap (actiCAP, Munich, Germany) and positioned according to the extended 10–20 system. Vertical electrooculogram (VEOG) and horizontal electrooculogram (HEOG) were recorded using bipolar electrodes placed at the supra- and suborbit of the right eye and at the external canthi. Electrode impedances were kept below 10 kΩ. The data were amplified using a 128-channel amplifier (QuickAmp, Brain Products, Enschede, Netherlands) and digitized with BrainVisionRecorder software (Brain Products, Munich, Germany) at a 500-Hz sampling rate. The EEG signal was recorded against an average of all channels calculated by the amplifier hardware.

EEG and EOG data were analyzed using EEGLAB 14 functions and MATLAB 2016b. First, all signals were filtered using high-pass (0.5 Hz) and low-pass (45 Hz) Butterworth IIR filter (filter order = 2; MATLAB functions, *butter* and *filtfilt*). Then, data were rereferenced to the average of signals recorded from left and right earlobes and downsampled to 250 Hz. All data were divided into 1600 epochs (400 epochs per condition [−200 to 1000] ms with respect to the scene image onset), and the epochs were baseline-corrected by subtracting the mean of the prestimulus period (i.e., −200 to 0 ms). Further, epochs were rejected based on the following criteria: 1) when there was no manual response to the target dots until 1.5 s after the onset (144 ± 10 epochs per subject); 2) when activity of the HEOG electrode in the time window (−200 to 600) ms exceeded −40 or 40 μV (298 ± 38 epochs); and 3) when activity of the P7 or P8 electrode in the time window (−200 to 600) ms exceeded −80 or 80 μV (very few epochs rejected, only 0.28 ± 0.10 per subject). Thus, after applying the described criteria, the average number of analyzed epochs per subject was 1157 ± 46 (range: [603 1492]).

A subject was excluded if the number of epochs in any condition was < 100. This criterion resulted in excluding 3 subjects out of 32 (but additional 4 subjects were excluded due to other criteria, as described in the "Subjects" section). The numbers of epochs provided above were calculated based on the final sample of 25 subjects.

Next, each EEG–EOG data set was decomposed into 50 components using an independent component analysis as implemented in the EEGLAB *pop_runica* function. To remove residual oculographic artifacts from the data, the following procedures were used: time course of each component was correlated with time courses of HEOG and VEOG electrodes, and in case the Spearman correlation coefficient exceeded −0.3 or 0.3, a component was subtracted from the data. Using this procedure 2.83 ± 0.17 components (range [2–5]) per subject were removed.

After applying the described preprocessing steps, data were divided with respect to the condition and presentation side of the threatening/incongruent scene. To calculate the N2pc component, we used the P8 and P7 electrodes, similarly to Kappenman et al. (2015). Specifically, when threatening/emotional scene was presented on the left side, P8 was the contralateral electrode and P7 was the ipsilateral electrode. When threatening/emotional scene was presented on the right side, P7 was the contralateral electrode and P8 was the ipsilateral electrode. For each condition contralateral and ipsilateral signals were first concatenated and then averaged to obtain contralateral and ipsilateral waveforms. Then for each subject and condition, the difference between the contralateral and ipsilateral waveforms was calculated. All statistical analyses were conducted on the difference waveforms averaged within the defined time windows.

Statistical Analysis

All statistical analyses were conducted in the JASP software and cross-checked with Statcheck (http://statcheck.io/index.php). The values are reported as mean ± standard error of the mean, unless stated otherwise.

In the present study, the BF was used as the primary statistical measure. The main reason for choosing BF was that, unlike the classic frequentist statistics, BF evaluates how strongly both alternative and null hypotheses are supported by data. Specifically, BF is a ratio of the probability (or likelihood) of observing the data given the alternative hypothesis is true to the probability of observing the data given the null hypothesis is true. Thus, in our particular case, BF allows providing further evidence either in favor of or against attention capture by threatening or semantically incongruent scenes.

In all Bayesian tests, the medium prior scale (Cauchy scale 0.707) was used. In the Results section, we provide interpretations of the BF according to Wagenmakers et al. (2018), with $0.33 < BF < 3$ indicating inconclusive (anecdotal) evidence. Additionally, for each comparison, we also provide results of a frequentist test to complement BF. Data distribution was first tested with the Shapiro–Wilk test, and a *t*-test was used when the distribution was Gaussian or a nonparametric Wilcoxon test when it deviated from normality.

To test for the presence of the dot-probe task effects (accuracy, RT, and N2pc), one-tailed (directional) *t*-tests were used within each condition, to define whether an effect was present (BF indicating evidence for alternative hypothesis) or absent (BF indicating evidence for null hypothesis). Further, based on the inspection of obtained ERP waveforms, two more time windows (300–400 ms and 600–1000 ms) were analyzed in an exploratory way. In these exploratory analyses, two-tailed tests were used. Finally, when analyzing behavioral data, we aimed to compare two measures from the identification task—d′ and confidence ratings—between conditions. Thus, we used repeated measures ANOVA with two factors: stimulus type (threat, congruency) and presentation time (100 ms, 500 ms).

Data Availability

Data used in the statistical analyses are freely available from https://osf.io/zpagb/. These include behavioral data from the dot-probe task (accuracy and RT) and from the identification task (d′ and confidence ratings) and mean amplitudes of the analyzed ERP components. Raw EEG data and analysis scripts are available from the authors per request.

Results

Identification Task

The sensitivity index (d′) was calculated to evaluate subjects' ability to recognize threatening and incongruent scenes when they were displayed for longer (500 ms) or shorter (100 ms) time (Fig. 1). Analysis of variance provides extremely strong evidence that d′ is affected by both "scene type" ($BF > 1000$; $F(1, 23) = 122.07, P < 0.001$) and presentation time ($BF > 1000$; $F(1, 23) = 34.21$, $P < 0.001$; Fig. 2A). Specifically, subjects were better at recognizing threatening than incongruent scenes (mean d′ difference: 1.51 ± 0.13), and they performed better when scenes were displayed for a longer time (mean d′ difference: 0.90 ± 0.15). However, evidence for an interaction was inconclusive ($BF = 1.18$; $F(1, 23) = 5.10$, $P = 0.034$).

Further, we analyzed confidence ratings from the identification task. For each subject and condition, we calculated a percentage of trials in which subjects rated their confidence as high ("I was rather sure" or "I was sure" responses; Fig. 2B). Again, we found extremely strong evidence that confidence ratings are affected by both scene type ($BF > 1000$; $F(1, 23) = 57.12$, $P < 0.001$) and presentation time ($BF > 1000$; $F(1, 23) = 28.24$, $P < 0.001$). Subjects reported higher confidence on trials where threatening scenes were presented (mean difference: 16.4 ± 2.2% of trials) and when scenes were displayed for a longer time (mean difference: 14.4 ± 2.8%). Evidence for interaction was again

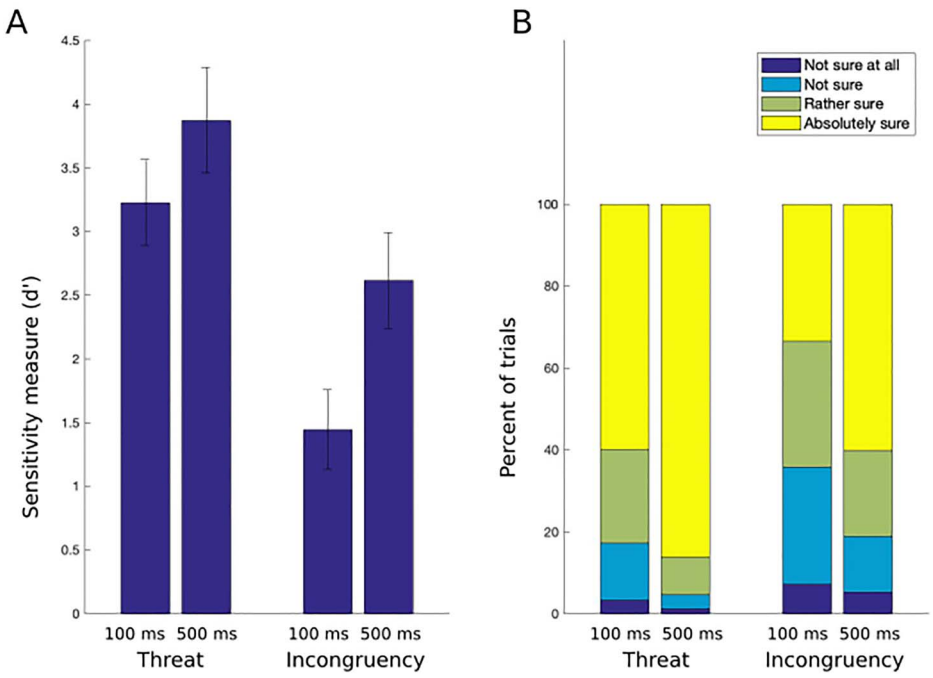

Figure 2. Results of the identification task, in which subjects were asked to decide whether a threatening/incongruent scene was presented on the left or on the right side and then rate their confidence in their choice. d′, used as an index of objective identification performance, is presented in (*A*). Percentage of trials with a given confidence rating per condition is presented in (*B*).

inconclusive (BF = 0.44; $F(1, 23) = 3.08$, $P < 0.09$). Thus, both objective (d′) and subjective (confidence ratings) measures provide consistent results.

Dot-Probe Task—Behavioral Results

Based on the dot-probe data, we analyzed accuracy and RTs with respect to the target dots. Specifically, we tested whether in any condition 1) accuracy was higher and 2) RT was shorter in trials when targets followed threatening/incongruent scenes, in comparison to trials when dots followed neutral/congruent scenes. For accuracy, we found moderate evidence for the null hypothesis (i.e., no accuracy effect) when incongruent scenes were presented for 100 ms (95.0 ± 0.7% vs. 95.2 ± 0.8%; BF = 0.24; null hypothesis 4.2 times more likely; $t(24) = -0.55$, $P = 0.58$) and inconclusive data in the 500-ms presentation time condition (95.2 ± 0.8% vs. 96.1 ± 0.7%; BF = 1.11; $Z = 84$, $P = 0.10$). For threatening scenes moderate evidence favoring the null hypothesis was observed for both 100 ms (95.4 ± 1.0% vs. 95.3 ± 0. 9%; BF = 0.22; null hypothesis 4.5 times more likely; $t(24) = 0.28$, $P = 0.78$) and 500-ms presentation time (95.5 ± 0.9% vs. 96.1 ± 0.7%; BF = 0.37; null hypothesis 2.7 times more likely; $t(24) = -1.129$, $P = 0.27$). Of note, in all conditions accuracy exhibited very high (ceiling) values (greater than 95%), which is not surprising considering the task was relatively simple.

Similarly, for RTs we found moderate evidence for the null hypothesis (i.e., no RT effect) for incongruent scenes for both presentations time of 100 ms (629.71 ± 15.16 vs. 629.24 ± 15.11 ms; BF = 0.21; null hypothesis 4.8 times more likely; $t(24) = 0.15$, $P = 0.88$) and 500 ms (616.56 ± 11.15 vs. 615.65 ± 10.84 ms; BF = 0.22; null hypothesis 4.5 times more likely; $Z = 194$, $P = 0.41$). As for threatening scenes, moderate evidence in favor of the null hypothesis was found for 100-ms presentations (651.61 ± 13.08 vs. 652.22 ± 13.62 ms; BF = 0.22; null hypothesis 4.5 times more likely; $Z = 175$, $P = 0.75$); however, in the 500-ms presentation time condition, we found moderate evidence in favor of "alternative" hypothesis (650.33 ± 11.93 vs. 644.09 ± 11.62 ms; BF = 3.5; alternative hypothesis 3.5 times more likely; $t(24) = 2.64$, $P = 0.014$). Thus, no behavioral effects of attention capture were found for semantically incongruent scenes (in line with Mudrik et al. 2011 and Mack et al. 2017), whereas for threatening scenes such an effect was found for 500-ms presentation time only.

Dot-Probe Task—Electrophysiological Results

Next, we analyzed the N2 posterior-contralateral (N2pc) ERP component, which is a robust index of covert attention shifts (Eimer 1996; Kiss et al. 2008). For consistency with Kappenman et al. (2015), we analyzed signals from the P7/P8 electrodes and used the 175–225 ms time window (Fig. 3).

We observed extreme evidence that the N2pc component occurs when threatening scenes are presented for 500 ms ($M = -0.83 \pm 0.17$; BF = 868; $t(24) = 4.86$, $P < 0.001$). Importantly, extreme evidence in favor of N2pc was obtained also when threatening scenes were presented for 100 ms only ($M = -0.88 \pm 0.17$; BF = 1310; $Z = 11$, $P < 0.001$). However, when semantically incongruent scenes were presented, strong evidence favoring the null hypothesis was found (i.e., absence of N2p), both for 100 ms ($M = 0.16 \pm 0.07$; BF = 0.076, null hypothesis 13 times more likely; $t(24) = 0.14$, $P = 0.97$) and for 500-ms presentations ($M = 0.20 \pm 0.07$; BF = 0,065, null hypothesis 15 times more likely; $Z = 266$, $P = 0.99$). Thus, our results strongly suggest that threatening scenes automatically capture attention and that they are recognized rapidly and efficiently even when presented for a brief duration (i.e., 100 ms). In sharp contrast,

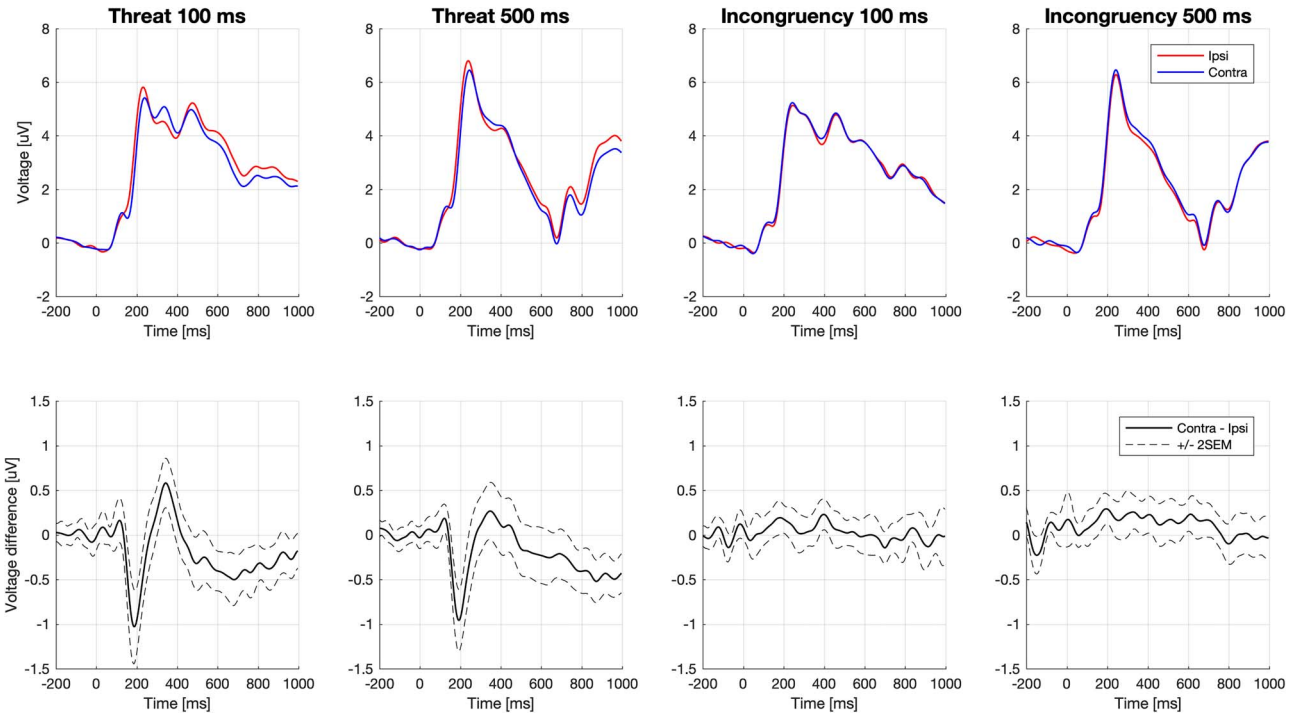

Figure 3. Event-related potentials in the dot-probe task. Electrodes P7 and P8 were chosen for the analysis. Waveforms from electrodes ipsilateral and contralateral to the threatening/incongruent images are presented in the top row. Difference waveforms (i.e., contralateral/ipsilateral) are presented in the bottom row.

semantically incongruent scenes do not attract attention, even when presented for a relatively long time.

We further conducted two exploratory analyses, motivated by inspection of the obtained ERP waveforms. First, we noticed that the N2pc evoked by threatening scenes is followed by a contralateral positivity (Fig. 3). According to Gaspelin and Luck (2018), such a positive component might indicate attention disengagement or suppression of a stimulus that initially captured attention. To investigate this effect, a 300–400 ms time window was used. We indeed found strong evidence that contralateral positivity occurs when threatening scenes are presented for 100 ms ($M = 0.40 \pm 0.12$; BF = 13.3; $t(24) = 3.30$, $P = 0.003$), but when presented for 500 ms, the evidence is inconclusive ($M = 0.19 \pm 0.15$; BF = 0.41; $t(24) = 1.23$, $P = 0.23$). Inconclusive evidence was also observed when semantically incongruent scenes were displayed for 100 ms ($M = 0.11 \pm 0.08$; BF = 0.45; $Z = 191$, $P = 0.45$) and for 500 ms ($M = 0.18 \pm 0.09$; BF = 1.14; $t(24) = 1.99$, $P = 0.058$).

Second, we analyzed the negative component which started around 600 ms after presentation of threatening scenes (time window 600–1000 ms). This component might be interpreted as sustained posterior-contralateral negativity (SPCN), which is proposed to reflect selection and maintenance of information in visual short-term memory (e.g., Eimer and Kiss 2010). We found very strong evidence for the presence of a late negative component when threatening scenes were displayed both for 100 ms ($M = -0.38 \pm 0.09$; BF = 77.74; $t(24) = 4.10$, $P < 0.001$) and for 500 ms ($M = -0.39 \pm 0.11$; BF = 25.46; $t(24) = 3.60$, $P = 0.001$). However, there was moderate evidence favoring lack of such a late negative component when incongruent scenes were displayed for 100 ms ($M = -0.02 \pm 0.08$; BF = 0.217; null hypothesis 4.6 more likely; $t(24) = 0.25$, $P = 0.80$) and for 500 ms ($M = 0 \pm 0.09$; BF = 0.211; null hypothesis 4.6 more likely; $t(24) = 0$, $P = 0.99$).

Discussion

In the present study, we investigated the role of attention in processing of semantically incongruent objects, defined as objects that are not likely to appear in a given context. Previous studies have consistently found that incongruent objects preferentially engage (i.e., hold or maintain) attention once they have been perceived (e.g., Võ and Henderson 2009, 2011; Mudrik et al. 2011; Cornelissen and Võ 2017). But whether such objects also automatically "capture" attention remains controversial, with some studies providing evidence in favor (e.g., Loftus and Mackworth 1978; Hollingworth and Henderson 2000; Underwood et al. 2007, 2008; Ortiz-Tudela et al. 2017, 2018), but other against such a possibility (e.g., Võ and Henderson 2009, 2011; Mudrik et al. 2011; Moors et al. 2016; Mack et al. 2017). Revealing automatic allocation of attention to incongruent objects would have important implications, as it would mean that semantic inconsistency is identified preattentively and results in a "semantic pop-out" (see Wu et al. 2014). Therefore, to address this question, we used a procedure—involving a dot-probe task and N2pc component analysis—that has already been shown sensitive to detect attention shifts to complex stimuli, specifically threatening scene images (Kappenman et al. 2014, 2015). Investigating attentional selection of both threatening and incongruent scenes in the same experiment, we found that while threatening images indeed capture attention automatically and rapidly, semantically incongruent images do not benefit from an automatic attentional prioritization.

Semantically Incongruent Scenes

Our conclusion that attention is not captured by semantic inconsistencies is based on the lack of both behavioral (dot-probe accuracy and RT) and electrophysiological (N2pc) evidence. Such lack of effects was further attested by Bayesian analysis, which

suggested conclusive evidence in favor of the null hypothesis, mitigating the concern that effects were not observed due to lack of power. Three further alternative explanations are also excluded by the results. First, one might claim that the experimental procedure is not sensitive enough; however, since attention shifts were caused by threatening scenes, both here and in the Kappenman et al. (2015) study, this explanation is less plausible (as well as similar claims for some other factor in our experimental setting that affected subjects' performance). Thus, the positive finding for the threatening stimuli provides a stronger context for the negative finding for the semantic incongruencies. Second, one might claim that attention shifts to incongruent objects are transient and covert and, for this reason, not manifested in behavior (see Schmukle 2005 ; Kappenman et al. 2014, 2015). Here, ERP results serve as an answer, as they provide a continuous measure of neuronal engagement. Nevertheless, ERP analysis did not suggest such transient attention shifts, nor any other form of preferential processing of incongruent scenes. Finally, the semantic incongruencies used in the present study might be considered not strong enough to evoke attentional shifts (i.e., the manipulation of images was too subtle). Indeed, images used by the previous studies varied greatly in terms of general complexity and how central and diagnostic the key object was for the scene understanding. But the scenes we used were always defined as a subject performing an action with an object, which was always central and played a key role in the scene. Importantly, previous studies using the same stimuli did find that congruent and incongruent scenes were processed differentially (e.g., Mudrik et al. 2010, 2011, 2014; Moors et al. 2016; Mack et al. 2017; Truman and Mudrik 2018; Faivre et al. 2019). However, what most effectively mitigates the abovementioned concern is subjects' performance in the identification task. There, both objective d′ and subjective confidence ratings indicate that recognition of incongruent scenes was above chance level (see also the high accuracy rates in previous studies using these stimuli, e.g., Mudrik et al. 2010, 2014).

Thus, our findings challenge results of several previous studies. First, incongruent objects were shown to cue attention to their location, as indicated by a faster reaction to a subsequent probe (Gordon 2004). If this is the case, then the dot-probe task should have uncovered such an attentional cueing. But we did not find any supporting evidence for this claim. Second, observers are faster to detect a change in the change blindness experiments, when the change involves an incongruent object (e.g., Hollingworth and Henderson 2000; LaPointe et al. 2013; Mack et al. 2017; LaPointe and Milliken 2017; Ortiz-Tudela et al. 2017, 2018). Third, several eye-tracking studies reported that incongruent objects were fixated earlier than congruent ones (Loftus and Mackworth 1978; Underwood and Foulsham 2006; Becker et al. 2007; Underwood et al. 2007, 2008; Bonitz and Gordon 2008). But while the change blindness and eye-tracking studies indicate attention might be preferentially allocated to incongruent objects, our results suggest that this process is not automatic but rather requires longer inspection and exploration of the scene. Specifically, in change blindness experiments, several reversals of the stimulus are typically needed to detect a change, even if the target object is incongruent (e.g., in the Mack et al. 2017), an average time needed to detect a change was 11.7 s). Similarly, incongruent objects typically did not attract the first saccade but rather were fixated only after several saccades have been executed.

Our results are in line with previous ERP studies, which found that differences between congruent and incongruent scenes begin around 250–300 ms after the scene onset and are most pronounced over anterior brain regions (the N300 component; Ganis and Kutas 2003; Mudrik et al. 2010, 2014; Võ and Wolfe 2013; Truman and Mudrik 2018; Draschkow et al. 2019). This suggests that identification of a semantic incongruity occurs at processing stages later than attentional selection investigated here. Further, the pattern of fMRI activations caused by incongruent scenes—involving lateral occipital, inferior temporal, parahippocampal, and prefrontal cortex—again indicates that identification of incongruency involves mainly postperceptual processing stages (Rémy et al. 2014, 2020; Faivre et al. 2019). This is also suggested by the fact that conscious perception of a scene is required for detection of incongruency, as it is not performed when scenes are presented subliminally (Moors et al. 2016; Biderman and Mudrik 2018; Faivre et al. 2019).

Finally, in the previous ERP/fMRI studies, the presented scenes were typically task-relevant, but here we presented them as irrelevant distractors. Therefore, it still might be argued that presenting incongruent objects as task-relevant, hereby increasing subjects' motivation to inspect them, might result in a preferential attention allocation. However, first, our aim was to test if the capture is indeed automatic and, second, both a passive task and an active task were used in an eye-tracking study by Võ and Henderson (2009) (i.e., subjects viewed scenes for a later recall or searched a predefined object, respectively), and in neither task inconsistent objects attracted initial saccades more frequently than congruent ones.

Threatening Scenes

In two studies, Kappenman et al. (2014, 2015) found electrophysiological evidence for an automatic attention capture by threatening natural scenes. Here, we used the same dot-probe procedure and the subset of IAPS stimuli as Kappenman et al. (2015) and successfully replicated their N2pc effect. While in the present study threatening scenes were introduced as a "control" condition, we nevertheless extended the observations of Kappenman and colleagues in three ways.

First, in Kappenman et al. (2014, 2015), threatening scenes were presented for a relatively long time (between 400 and 600 ms). Here, alongside replicating the result for the 500-ms presentation time, we also show that a 100-ms display time is sufficient to yield a robust N2pc component. Thus, detection of emotional scenes is indeed rapid and efficient. Such a result regarding scenes complements previous studies showing an automatic detection of emotional expression of faces (review Vuilleumier 2005; but also identity, Wójcik et al. 2018, 2019). Second, we introduced an identification task, which was not included by Kappenman et al. (2014, 2015). We were thus able to show that subjects not only covertly shift attention toward a threatening scene but also recognize the side on which it was presented with high accuracy and confidence, even in the 100-ms display time condition. This again points toward a very efficient, possibly parallel processing of both scenes displayed in the experiment, in line with reports that threatening IAPS images might even be processed when presented subliminally (i.e., without awareness; e.g., Gläscher and Adolphs 2003; Tooley et al. 2017; meta-analysis: Hedger et al. 2016). A third way in which we extended the Kappenman et al. (2015) study is that we observed not only the N2pc component but also later lateralized ERP components. Specifically, perception of threatening scenes led to a late (600–1000 ms) sustained posterior-contralateral negativity (SPCN). SPCN is typically interpreted as a marker

of a stimulus access to visual working memory, which might in fact occur even when the task does not explicitly require working memory encoding (Jolicœur et al. 2008; Luria et al. 2010; Sessa et al. 2011). Under this interpretation, it seems that the threatening scenes were actively processed in working memory up to 1 s after their onset, even though they were supposed to be ignored as task-irrelevant distractors. This component was not reported by Kappenman et al. (2014, 2015), as they analyzed EEG signal only up to 600 ms from scene's onset. Further, in the short presentation time condition, we also observed a contralateral positivity (300–400 ms), which can be interpreted to index attentional disengagement and/or suppression of a stimulus that automatically captures attention (Gaspelin and Luck 2018). Nevertheless, in this condition the later SPCN was still observed, indicating that this possible suppression/disengagement was not fully successful. Because these components were found in exploratory analyses, their interpretation requires caution, but warrants future investigation.

Analyzing behavioral data from the dot-probe task, we found faster RT when target dots followed the threatening scene, but only in the 500-ms presentation time condition. This is in contrast to Kappenman et al. (2014, 2015) who did not find any RT effect. The reason for this discrepancy remains unclear, but is in line with the fact that previous dot-probe studies that used an RT index to investigate attention shifts to threatening stimuli reported mixed results (for meta-analysis, see Bar-Haim et al. 2007). The most plausible reason is that RT of a manual response measures an outcome of a whole chain of processes (perceptual, cognitive, motor) occurring between stimulus presentation and response and thus might not be sensitive enough to indicate transient attention capture. Indeed, in their additional analyses, Kappenman et al. (2015) demonstrated poor internal reliability of the RT index (see also Schmukle 2005). Further, based on the fact that the RT effect was found only in the 500-ms display time condition, we argue that, rather than reflecting a rapid attention capture, it might result from a prolonged attentional engagement with the threatening scene. Thus, it would be important for future studies to define how the N2pc component relates to behavior. Importantly, even though N2pc is considered the gold standard for detecting attentional shifts, some recent studies suggest it might actually index other downstream processes, like feature integration (Zivony et al. 2018).

Conclusions and Limitations

A comprehensive theoretical model describing how our visual system deals with perception of natural scenes is still missing. There is strong evidence that human observers are very fast and efficient at recognizing a gist of a scene, defined as a general and high-level representation of what the scene depicts (review: Oliva and Torralba 2006). However, there is still controversy around the effects of the scene's gist or context on object recognition processes. One line of research suggests that the gist is identified first, and by preactivating representations of objects most likely encountered in a given context, gist perception affects the subsequent object recognition (Biederman et al. 1982; Brandman and Peelen 2017; Truman and Mudrik 2018; review Bar 2004). However, others show that objects can be categorized as rapidly as the gist, with earliest manual responses in an object classification task as short as 250 ms, claiming that object perception is basically cost-free (i.e., does not require attentional resources; Thorpe et al. 1996; Joubert et al. 2007, 2008). These results are more in line with a model in which gist and objects are processed in parallel with an ongoing interaction between the two streams (review Fabre-Thorpe 2011).

We argue that to recognize a threat in the scenes presented in our study, it is sufficient to either process the gist of a scene or just recognize a key object (e.g., a mutilated body, an attacking animal). In contrast, to detect an incongruency, both the gist of a scene and the key object must be recognized, and the mismatch has to be established in the process. Thus, detecting a threat indeed relies on processing that is rapid and efficient and involves automatic attentional selection (i.e., might be preattentive and cost-free), whereas detecting semantic incongruency likely requires more time and can be performed only after attention has been allocated to a given object. Involvement of different mechanisms is also suggested by the fact that the recognition level of threatening scenes presented for 100 ms and incongruent scenes presented for 500 ms was similar, but we observed robust attention capture in the former condition, but no sign of such an effect in the latter. Thus, our data suggest that gist and/or objects can be perceived in a rapid, automatic, and possibly cost-free manner, but the detection of gist–object mismatch is performed later and requires attentional resources. We argue that rapid analysis of low spatial frequencies performed by the magnocellular pathway might be sufficient to detect a threat, whereas identification of a gist–object mismatch is likely based on a more precise analysis of high spatial frequencies (in line with Bar et al. 2006; Alorda et al. 2007; Lauer et al. 2018; see also Bar 2004).

Further, interpreting the differences between the incongruent and threatening scenes should be done with caution, as some low-level features were not matched between sets (e.g., congruent/incongruent scenes exhibit higher luminance, lower contrast, and higher entropy; see the Methods section). Thus, conducting a direct comparison between the two types of stimuli is limited. Notably however, the results for each type of stimuli stand on their own; that is, in line with our experimental goals, we were able to, first, replicate the result of Kappenman et al. (2015) regarding threatening scenes and examine the effect of stimulus duration on the effect. And second, we investigated the involvement of attention in processing semantically incongruent scenes that have been already used in multiple studies (e.g., Mudrik et al. 2010, 2011; Mack et al. 2017; Faivre et al. 2019). Given that the two classes of stimuli banks were previously validated, and as modifying their low-level features might affect their processing, we decided to neither match the physical properties between sets nor modify the stimuli in any other way. Note that such a modification might have also complicated the interpretation of our results in relation to these previous studies (particularly when obtaining a null result).

In conclusion, a growing body of evidence indicates that representation of a scene context interacts with the object recognition process as early as 200 ms after the scene onset (Bar et al. 2006; Joubert et al. 2007, 2008; Brandman and Peelen 2017; Truman and Mudrik 2018; review Bar 2004; Fabre-Thorpe 2011). Our results place an important limitation on the future theoretical models by showing that the context-object integration does not take place preattentively but only after attention has been allocated to the object.

Funding

National Science Center Poland (grant number 2018/29/B/HS6/02152); COST (European Cooperation in Science and Technology,

COST Action CA18106); Polish Ministry of Science and Higher Education (555/STYP/11/2016 to M.B.). *Conflict of Interest*: The authors declare no competing interests.

References

Alorda C, Serrano-Pedraza I, Campos-Bueno JJ, Sierra-Vázquez V, Montoya P. 2007. Low spatial frequency filtering modulates early brain processing of affective complex pictures. *Neuropsychologia*. 45:3223–3233.

Bar-Haim Y, Lamy D, Pergamin L, Bakermans-Kranenburg MJ, Van Ijzendoorn MH. 2007. Threat-related attentional bias in anxious and nonanxious individuals: a meta-analytic study. *Psychol Bull*. 133:1–24.

Bar M. 2004. Visual objects in context. *Nat Rev Neurosci*. 5:617–629.

Bar M, Kassam KS, Ghuman AS, Boshyan J, Schmid AM, Dale AM, Hämäläinen MS, Marinkovic K, Schacter DL, Rosen BR, et al. 2006. Top-down facilitation of visual recognition. *Proc Natl Acad Sci U S A*. 103:449–454.

Becker MW, Pashler H, Lubin J. 2007. Object-intrinsic oddities draw early saccades. *J Exp Psychol Hum Percept Perform*. 33:20–30.

Biderman N, Mudrik L. 2018. Evidence for implicit—but not unconscious—processing of object-scene relations. *Psychol Sci*. 29:266–277.

Biederman I, Mezzanotte RJ, Rabinowitz JC. 1982. Scene perception: detecting and judging objects undergoing relational violations. *Cogn Psychol*. 14:143–177.

Bonitz VS, Gordon RD. 2008. Attention to smoking-related and incongruous objects during scene viewing. *Acta Psychol (Amst)*. 129:255–263.

Boyce SJ, Pollatsek A, Rayner K. 1989. Effect of background information on object identification. *J Exp Psychol Hum Percept Perform*. 15:556–566.

Brandman T, Peelen MV. 2017. Interaction between scene and object processing revealed by human fMRI and MEG decoding. *J Neurosci*. 37:7700–7710.

Cohen MA, Dennett DC, Kanwisher N. 2016. What is the bandwidth of perceptual experience? *Trends Cogn Sci*. 20: 324–335.

Cooper RM, Langton SR. 2006. Attentional bias to angry faces using the dot-probe task? It depends when you look for it. *Behav Res Ther*. 44:1321–1329.

Cornelissen TH, Võ MLH. 2017. Stuck on semantics: processing of irrelevant object-scene inconsistencies modulates ongoing gaze behavior. *Atten Percept Psychophys*. 79:154–168.

Davenport JL, Potter MC. 2004. Scene Consistency in Object and Background Perception. *Psychol Sci*. 15:559–564.

De Graef P, Christiaens D, d'Ydewalle G. 1990. Perceptual effects of scene context on object identification. *Psychol Res*. 52:317–329.

Draschkow D, Reinecke S, Cunningham CA, Võ MLH. 2019. The lower bounds of massive memory: Investigating memory for object details after incidental encoding. *Q J Exp Psychol*. 72:1176–1182.

Eimer M. 1996. The N2pc component as an indicator of attentional selectivity. *Electroencephalogr Clin Neurophysiol*. 99:225–234.

Eimer M, Kiss M. 2010. An electrophysiological measure of access to representations in visual working memory. *Psychophysiology*. 47:197–200.

Epstein R, Graham KS, Downing PE. 2003. Specific scene representations in human parahippocampal cortex. *Neuron*. 37:865–876.

Fabre-Thorpe M. 2011. The characteristics and limits of rapid visual categorization. *Front Psychol*. 2:243.

Faivre N, Dubois J, Schwartz N, Mudrik L. 2019. Imaging object-scene relations processing in visible and invisible natural scenes. *Sci Rep*. 9:4567.

Ganis G, Kutas M. 2003. An electrophysiological study of scene effects on object identification. *Cogn Brain Res*. 16:123–144.

Gareze L, Findlay JM. 2007. Absence of scene context effects in object detection and eye gaze capture. In: van Gompel R, Fischer M, Murray W, Hill RW, editors. *Eye movements: a window on mind and brain*. Elsevier: Amsterdam (Netherlands), pp. 537–562.

Gaspelin N, Luck SJ. 2018. Inhibition as a potential resolution to the attentional capture debate. *Curr Opin Psychol*. 29: 12–18.

Gläscher J, Adolphs R. 2003. Processing of the arousal of subliminal and supraliminal emotional stimuli by the human amygdala. *J Neurosci*. 23:10274–10282.

Gordon RD. 2004. Attentional allocation during the perception of scenes. *J Exp Psychol Hum Percept Perform*. 30:760–777.

Hedger N, Gray KL, Garner M, Adams WJ. 2016. Are visual threats prioritized without awareness? A critical review and meta-analysis involving 3 behavioral paradigms and 2696 observers. *Psychol Bull*. 142:934–968.

Henderson JM, Hollingworth A. 1999. High-level scene perception. *Annu Rev Psychol*. 50:243–271.

Hollingworth A, Henderson JM. 2000. Semantic Informativeness mediates the detection of changes in natural scenes. *Vis Cogn*. 7:213–235.

Jolicœur P, Brisson B, Robitaille N. 2008. Dissociation of the N2pc and sustained posterior contralateral negativity in a choice response task. *Brain Res*. 1215:160–172.

Joubert OR, Rousselet GA, Fize D, Fabre-Thorpe M. 2007. Processing scene context: fast categorization and object interference. *Vision Res*. 47:3286–3297.

Joubert OR, Fize D, Rousselet GA, Fabre-Thorpe M. 2008. Early interference of context congruence on object processing in rapid visual categorization of natural scenes. *J Vis*. 8(13):1–18.

Kaiser D, Quek GL, Cichy RM, Peelen MV. 2019. Object vision in a structured world. *Trends Cogn Sci*. 23:672–685.

Kappenman ES, Farrens JL, Luck SJ, Proudfit GH. 2014. Behavioral and ERP measures of attentional bias to threat in the dot-probe task: poor reliability and lack of correlation with anxiety. *Front Psychol*. 5:1368.

Kappenman ES, MacNamara A, Proudfit GH. 2015. Electrocortical evidence for rapid allocation of attention to threat in the dot-probe task. *Soc Cogn Affect Neurosci*. 10:577–583.

Kiss M, Van Velzen J, Eimer M. 2008. The N2pc component and its links to attention shifts and spatially selective visual processing. *Psychophysiology*. 45:240–249.

Koster EH, Verschuere B, Crombez G, Van Damme S. 2005. Time-course of attention for threatening pictures in high and low trait anxiety. *Behav Res Ther*. 43:1087–1098.

Lang, P.J., Bradley, M.M., Cuthbert, B.N. 2008. International affective picture system (IAPS): affective ratings of pictures and instruction manual. Gainesville (FL): University of Florida. Technical Report A-8.

LaPointe MR, Lupianez J, Milliken B. 2013. Context congruency effects in change detection: opposing effects on detection and identification. *Vis Cogn*. 21:99–122.

LaPointe MRP, Milliken B. 2017. Conflicting effects of context in change detection and visual search: A dual process account. *Can J Exp Psychol* 71:40–51.

Lauer T, Cornelissen TH, Draschkow D, Willenbockel V, Võ MLH. 2018. The role of scene summary statistics in object recognition. *Sci Rep*. 8:14666.

Loftus GR, Mackworth NH. 1978. Cognitive determinants of fixation location during picture viewing. *J Exp Psychol Hum Percept Perform*. 4:565–572.

Luria R, Sessa P, Gotler A, Jolicœur P, Dell'Acqua R. 2010. Visual short-term memory capacity for simple and complex objects. *J Cogn Neurosci*. 22:496–512.

Mack A. 2003. Inattentional blindness: looking without seeing. *Curr Dir Psychol Sci*. 12:180–184.

Mack A, Clarke J, Erol M, Bert J. 2017. Scene incongruity and attention. *Conscious Cogn*. 48:87–103.

MacLeod C, Mathews A, Tata P. 1986. Attentional bias in emotional disorders. *J Abnorm Psychol*. 95:15–20.

Moors P, Boelens D, Van Overwalle J, Wagemans J. 2016. Scene integration without awareness: no conclusive evidence for processing scene congruency during continuous flash suppression. *Psychol Sci*. 27:945–956.

Mudrik L, Lamy D, Deouell LY. 2010. ERP evidence for context congruity effects during simultaneous object–scene processing. *Neuropsychologia*. 48:507–517.

Mudrik L, Deouell LY, Lamy D. 2011. Scene congruency biases binocular rivalry. *Conscious Cogn*. 20:756–767.

Mudrik L, Shalgi S, Lamy D, Deouell LY. 2014. Synchronous contextual irregularities affect early scene processing: replication and extension. *Neuropsychologia*. 56:447–458.

Oliva A, Torralba A. 2006. Building the gist of a scene: the role of global image features in recognition. *Prog Brain Res*. 155:23–36.

Oliva A, Torralba A. 2007. The role of context in object recognition. *Trends Cogn Sci*. 11:520–527.

Ortiz-Tudela J, Milliken B, Botta F, LaPointe M, Lupiañez J. 2017. A cow on the prairie vs. a cow on the street: long-term consequences of semantic conflict on episodic encoding. *Psychol Res*. 81:1264–1275.

Ortiz-Tudela J, Martin-Arevalo E, Chica AB, Lupianez J. 2018. Semantic incongruity attracts attention at a pre-conscious level: evidence from a TMS study. *Cortex*. 102:96–106.

Peelen MV, Kastner S. 2014. Attention in the real world: toward understanding its neural basis. *Trends Cogn Sci*. 18:242–250.

Rayner K, Castelhano MS, Yang J. 2009. Viewing task influences eye movements during active scene perception. *J Exp Psychol Learn Mem Cogn*. 35:254–259.

Rémy F, Vayssière N, Pins D, Boucart M, Fabre-Thorpe M. 2014. Incongruent object/context relationships in visual scenes: where are they processed in the brain? *Brain Cogn*. 84:34–43.

Rémy F, Vayssière N, Saint-Aubert L, Bacon-Macé N, Pariente J, Barbeau E, Fabre-Thorpe M. 2020. Age effects on the neural processing of object-context associations in briefly flashed natural scenes. *Neuropsychologia*. 136:107264.

Rieger JW, Köchy N, Schalk F, Grüschow M, Heinze HJ. 2008. Speed limits: orientation and semantic context interactions constrain natural scene discrimination dynamics. *J Exp Psychol Hum Percept Perform*. 34:56–76.

Schmukle SC. 2005. Unreliability of the dot probe task. *Eur J Pers*. 19:595–605.

Sessa P, Luria R, Gotler A, Jolicœur P, Dell'Acqua R. 2011. Interhemispheric ERP asymmetries over inferior parietal cortex reveal differential visual working memory maintenance for fearful versus neutral facial identities. *Psychophysiology*. 48:187–197.

Stanislaw H, Todorov N. 1999. Calculation of signal detection theory measures. *Behav Res Meth Instrum Comput*. 31:137–149.

Thorpe S, Fize D, Marlot C. 1996. Speed of processing in the human visual system. *Nature*. 381:520–522.

Tooley MD, Carmel D, Chapman A, Grimshaw GM. 2017. Dissociating the physiological components of unconscious emotional responses. *Neurosci Conscious*. 2017(1):nix021.

Truman A, Mudrik L. 2018. Are incongruent objects harder to identify? The functional significance of the N300 component. *Neuropsychologia*. 117:222–232.

Underwood G, Foulsham T. 2006. Visual saliency and semantic incongruency influence eye movements when inspecting pictures. *Q J Exp Psychol*. 59:1931–1949.

Underwood, G., Humphreys, L., Cross, E. 2007. Congruency, saliency, and gist in the inspection of objects in natural scenes. In: van Gompel, R. P. G., Fischer, M. H., Murray, W. S., Hill, R. L. *Eye movements: a window on mind and brain*. Amsterdam (Netherlands): Elsevier. p. 564–579.

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. *Conscious Cogn*. 17:159–170.

Võ MLH, Henderson JM. 2009. Does gravity matter? Effects of semantic and syntactic inconsistencies on the allocation of attention during scene perception. *J Vis*. 9:1–15.

Võ MLH, Henderson JM. 2011. Object–scene inconsistencies do not capture gaze: evidence from the flash-preview moving-window paradigm. *Atten Percept Psycho*. 73:1742–1753.

Võ MLH, Wolfe JM. 2013. Differential electrophysiological signatures of semantic and syntactic scene processing. *Psychol Sci*. 24:1816–1823.

Võ MLH, Boettcher SE, Draschkow D. 2019. Reading scenes: how scene grammar guides attention and aids perception in real-world environments. *Curr Opin Psychol*. 29:205–210.

Vuilleumier P. 2005. How brains beware: neural mechanisms of emotional attention. *Trends Cogn Sci*. 9:585–594.

Wagenmakers EJ, Love J, Marsman M, Jamil T, Ly A, Verhagen J, Selker R, Gronau QF, Dropmann D, Boutin B, et al. 2018. Bayesian inference for psychology. Part II: example applications with JASP. *Psychon Bull Rev*. 25:58–76.

Wójcik MJ, Nowicka MM, Kotlewska I, Nowicka A. 2018. Self-face captures, holds, and biases attention. *Front Psychol*. 8:2371.

Wójcik MJ, Nowicka MM, Bola M, Nowicka A. 2019. Unconscious detection of one's own image. *Psychol Sci*. 30:471–480.

Wolfe JM, Võ MLH, Evans KK, Greene MR. 2011. Visual search in scenes involves selective and nonselective pathways. *Trends Cogn Sci*. 15:77–84.

Wolfe JM, Horowitz TS. 2017. Five factors that guide attention in visual search. *Nat Hum Behav*. 1:0058.

Wu CC, Wick FA, Pomplun M. 2014. Guidance of visual attention by semantic information in real-world scenes. *Front Psychol*. 5:54.

Zivony A, Allon AS, Luria R, Lamy D. 2018. Dissociating between the N2pc and attentional shifting: an attentional blink study. *Neuropsychologia*. 121:153–163.