Behavioural Brain Research 394 (2020) 112812

Contents lists available at ScienceDirect

Behavioural Brain Research

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

Research report

The influence of scene and object orientation on the scene consistency effect

Tim Lauer[a,*], Verena Willenbockel[a,b], Laura Maffongelli[a,c], Melissa L.-H. Võ[a]

[a] Department of Psychology, Goethe University Frankfurt, Theodor-W.-Adorno-Platz 6, Frankfurt am Main 60323, Germany
[b] Department of Psychology, University of Victoria, Victoria, BC, Canada
[c] Department of Psychology, Johannes-Gutenberg University, Mainz, Germany

ARTICLE INFO

Keywords:
Scene consistency
Object recognition
Orientation
ERPs
N300
N400

ABSTRACT

Contextual regularities help us make sense of our visual environment. In scenes, semantically consistent objects are typically better recognized than inconsistent ones (e.g., a toaster vs. printer in a kitchen). What is the role of object and scene orientation in this so-called scene consistency effect? We presented consistent and inconsistent objects either upright (Experiment 1) or inverted (rotated 180°; Experiment 2) on upright, inverted, and scrambled background scenes. In Experiment 1, on upright scenes, consistent objects were recognized with higher accuracy than inconsistent ones, and we observed N300/N400 event-related potentials (ERPs) reflecting object-scene semantic processing. No such effects were observed for inverted or scrambled scenes. In Experiment 2, on both upright and inverted scenes, consistent objects were recognized with higher accuracy than inconsistent ones. Moreover, inconsistent objects on upright scenes triggered N300/N400 responses. Interestingly, no N300 but only an N400 deflection was found for inconsistent objects on inverted scenes. No effects were observed for scrambled scenes. These data suggest that while upright scenes modulate recognition irrespective of object orientation, inverted scenes only modulate the recognition of inverted objects. In ERPs, we found further evidence that inverted scenes can affect semantic object processing, with contextual influences occurring later in time, possibly driven by delayed or impaired scene gist processing. Mere object inversion does not seem to explain the later emergence of contextual influences. Taken together, the results suggest that the orientation of objects and scenes as well as their relationship to each other can influence ongoing object identification.

1. Introduction

Visual object recognition is a crucial cognitive process that is carried out repeatedly and usually effortlessly – despite a seemingly endless number of objects, variability in exemplars, viewpoints, etc. What are the "ingredients" of efficient object recognition? The occurrence of objects in the real world is typically not random but follows certain regularities. For instance, a cheese grater is often found in a kitchen but seldom if ever in a bathroom. Knowledge of the co-variation of objects and scenes, and the spatial arrangement of objects within scenes may facilitate object processing (for reviews, see [1,2]). Behavioral experiments have shown that objects that are consistent with the scene context (e.g., a fire hydrant on a street) are detected faster and more accurately than *semantically* or *syntactically* inconsistent objects (a fire hydrant in a kitchen or a fire hydrant floating in the air above a street, respectively ([3]; see also [4–6]; but see [7,8]). In line with this, eye-tracking studies have revealed that inconsistent objects are fixated longer and more often [9–15]. Scenes thus seem to be rule-governed and composed of a "grammar" – virtually like sentences – that we have implicitly learned and may exploit in a variety of tasks [16,17]. The current study aims at understanding how and under which conditions semantic consistency of an object with its surroundings influences object recognition.

While early behavioral studies have mostly used line drawings and variants of forced-choice tasks to examine context effects on object detection, more recent studies have used naturalistic scenes and a free object-naming paradigm that is not prone to response bias. Davenport and Potter [18] briefly presented participants with color scenes, each containing a semantically consistent or inconsistent object in the foreground. Participants named consistent objects with higher accuracy than inconsistent ones (hereafter referred to as *scene consistency effect*) (see also [19,20]). Likewise, scenes were reported more accurately if they contained a semantically consistent versus inconsistent object in the foreground, suggesting that objects and scenes are processed in an interactive manner. The scene consistency effect can be more pronounced when objects are seen from an "accidental" viewpoint that impedes recognition compared to a "canonical" (easy) viewpoint [21], indicating that the magnitude of context effects can vary with the

* Corresponding author at: Department of Psychology, Goethe University Frankfurt, Theodor-W.-Adorno-Platz 6, PEG 5.G144, 60323, Frankfurt am Main, Germany.
E-mail address: tlauer@psych.uni-frankfurt.de (T. Lauer).

https://doi.org/10.1016/j.bbr.2020.112812
Received 24 September 2019; Received in revised form 6 July 2020; Accepted 14 July 2020
Available online 16 July 2020
0166-4328/

difficulty of recognizing the object itself (see also [22]).

Context effects are not only evident in behavioral outcomes or eye-tracking measures but also in neurophysiological studies using event-related potentials (ERPs) that provide an online measure of cognitive processing with high temporal precision. Specifically, the N400 ERP component, which is an index of semantic processing traditionally related to the language domain (for a review, see [23]), has more recently also been reported in the scene perception domain ([24–26]; see also [27–30]). Seeing an inconsistent object in a scene – like reading an inconsistent word in a sentence – evokes a negative deflection that peaks about 400 ms post stimulus onset and is thought to index context effects on a semantic, conceptual level. Moreover, semantic violations in scenes trigger an N300 (e.g., [25,26,29]; but see [24]), an earlier negative deflection that has been suggested to reflect context effects on a more perceptual level, before object identification is completed (e.g., [25,29]). Whether this component is distinguishable from the N400 in terms of the underlying processes or not is still debated [27,30]. In any case, both components respond to semantic manipulations in scenes – even when the scene and the critical object are presented simultaneously (i.e., without a preview of the scene), ruling out the possibility that context effects can be reduced to prior expectations or prepared responses [25].

While context effects have been studied extensively, the question of what information of the scene context modulates object processing has received less attention. Contextual influences on object processing can arise from co-occurring objects in a scene [19]. Moreover, global scene properties could play a crucial role in explaining the immediate effects of context on object processing. Possibly, the spatial layout of a scene, its global "shape" [31], conveys the gist of the scene, leading to predictions of related objects [1,32]. The gist of a scene is a first impression of global meaning that one can obtain within a single glance (e.g., [33–35]). A recent study found a scene consistency effect when priming a target object with a semantically related versus unrelated global ensemble texture (i.e., a visual representation of a scene's spatial frequency and orientation information at multiple regions), preserving global shape information but not any recognizable objects [36]. Moreover, we recently demonstrated that a scene's summary statistics can even be relevant for object processing in the absence of scene shape information [28]. Likely, the visual system utilizes a combination of different sources of scene information for object processing – at overlapping or distinct points in time.

Back in 1999, Henderson and Hollingworth [8] determined three explanations of context effects on object detection performance: The *description enhancement model* argues that activation of scene schemata results in a finer, more detailed visual description of consistent versus inconsistent objects on a perceptual level. The *criterion modulation model* proposes that the incoming visual description of an object is matched with stored memory representations and that activated scene schemata may lower the perceptual threshold for finding a successful match; less perceptual information may be necessary for identifying a consistent object. By contrast, the *functional isolation model* argues that object identification is achieved independently from scene identification and that there are no interactions on a perceptual stage; however, context effects may manifest on a later, post-perceptual stage [7,8].

To advance our understanding of how context influences object processing, in the present study we explore the effects of object-scene inversion, that is, rotating objects and scenes independently 180 degrees in the picture-plane. Inversion is a simple yet powerful experimental manipulation that has been used in a variety of paradigms, predominantly in studies on face perception, since it preserves most of the information contained in the original upright image in terms of low-level features while interfering with the semantic interpretation of the image (e.g., [37,38]). Scene and object inversion thus provides a window into higher-level scene and object cognition. It has been proposed that scene inversion can disrupt the rapid extraction of scene gist [39,40]. Inverted scenes are categorized worse than upright scenes, evident both on the behavioral and the neuronal level [41]. In change detection paradigms, high interest changes in upright scenes (i.e., changes that are important for the meaning of the scene such as a goal on a football field) are typically detected better than low interest changes (i.e., changes that are not important for scene meaning), whereas this advantage drops for inverted scenes [42,43]. However, some studies have found only marginal effects of scene inversion on behavioral measures that were weaker than face inversion effects (e.g., [38,44,45]). Animal detection in natural scenes was shown to be hardly affected by, or even unaffected by, scene inversion ([45–47]; see also [48]). Arguably, the magnitude of inversion effects may vary with task demands: superordinate-level object discrimination (e.g., animal vs. no animal) may, for instance, not be impaired as much as fine-grained basic-level discrimination (see [47,49]).

Here, we investigated the impact of object-scene inversions on object processing as a function of semantic consistency. Does seeing a scene in an unfamiliar orientation impair its contextual influences on object processing and diminish the well-known scene consistency effect? And, what is the role of the relation of object and scene orientation? To this end, we used the inversion manipulation not only as a way to look at, or control for, low-level features, but to selectively impede semantic access of object and scene perception in order to investigate the processes underlying context effects. In Experiment 1, participants were presented with semantically consistent and inconsistent upright thumbnail objects (i.e., objects on white backgrounds) superimposed on three types of scene backgrounds: upright scenes, inverted scenes, and scrambled scenes (control condition). We first looked at object naming performance as a behavioral measure of context effects on object recognition (Experiment 1A). If inverted scenes modulate object recognition like upright scenes have been shown to, briefly presented consistent objects on inverted scenes should be named with higher accuracy than inconsistent ones. Scrambled scenes should not affect naming performance differentially. To address the question as to whether there is contextual facilitation of object recognition as opposed to mere interference by the semantic violations (e.g., functional isolation model), we contrasted accuracies for consistent objects on upright and scrambled scenes since the latter should not contain accessible scene semantics that would help (or hinder) object recognition. If there is facilitation, performance for consistent upright scenes should be higher than for scrambled scenes. Conversely, if there is interference, performance for inconsistent upright scenes should be lower than for scrambled scenes. Moreover, we recorded ERPs (Experiment 1B) to specifically track the time course of these context processes by looking at the N300/N400 components as online markers of context effects on object processing. If inverted scenes modulate semantic object processing, we should see an increased N300 and/or N400 effect in response to the inconsistent objects while no such effects should be observed for scrambled scenes. However, the effects of scene inversion on object processing may vary with the orientation of the critical object. Possibly, scene context modulates object processing more strongly if the object and the scene have a coherent orientation (i.e., they share the same orientation) as in previous semantic manipulation studies. To our knowledge, previous object or animal detection studies have not disentangled scene and object inversions (but see [47], for independent rotations by 90 degrees). Therefore, in Experiment 2, we presented participants with consistent and inconsistent inverted objects on upright scenes, inverted scenes, and scrambled scenes. Again, if inverted scenes modulate the processing of inverted objects, we should see higher object naming accuracy for consistent relative to inconsistent objects (Experiment 2A) – characterized by higher performance for inverted versus scrambled scenes – as well as an N300 and/or N400 effect (Experiment 2B).

2. Methods

2.1. Participants

Twenty-four participants took part in each experiment (Experiment 1A: 20 females, M = 23.4 years old, SD = 4; Experiment 1B: 16 females, M = 22.3 years old, SD = 4; Experiment 2A: 17 females, M = 21.6 years old, SD = 4.3; Experiment 2B: 15 females, M = 24.9 years old, SD = 4.3). One additional participant was excluded from Experiment 1A because instructions were not followed. Moreover, eleven participants were excluded from Experiment 1B and six from Experiment 2B due to bad performance in the RDT (N = 1; see Procedure, ERP paradigm), recording problems or noisy EEG signals (N = 6), and/or excessive EOG artefacts (e.g., blinks, eye-movements, N = 6) and/or alpha activity (N = 4). Participants received course credit or payment. They had normal or corrected-to-normal vision (at least 20/25 acuity), were unfamiliar with the stimulus material, naïve regarding the purpose of the study, and gave written informed consent. All participants who took part in Experiments 1A and 2A were native German speakers. All aspects of the data collection and analysis for all experiments reported here were carried out in accordance with guidelines approved by the Human Research Ethics Committee of the Goethe University Frankfurt.

2.2. Stimuli and design

We collected 288 real-world scenes (1024 × 768 pixels) from different indoor and outdoor categories (kitchen, office, bathroom, bedroom, mountain, beach, forest, street; 36 exemplars each) using Google search and the LabelMe database [50]. Each scene was paired with a semantically consistent thumbnail object (256 × 256 pixels) [51–54]. Semantic inconsistencies were created by pairing indoor and outdoor scenes, so that each object was consistent with one scene but inconsistent with another (see Fig. 1). The majority of the object-scene pairs were identical to the ones used in a previous study [28] but note that we extended the original stimulus set and balanced the number of exemplars per scene category.

In addition, we generated an inverted version of each scene by rotating the scene 180 degrees in the picture-plane, and a scrambled version (control condition) by randomly re-arranging all pixels of the original scene. In both experiments, scenes were randomly assigned to the six conditions (consistent upright scene, inconsistent upright scene, consistent inverted scene, inconsistent inverted scene, consistent scrambled scene, inconsistent scrambled scene) and counter-balanced across participants using a 3 × 2 Latin square design (see also [28]).

In Experiments 1A and 1B, thumbnail objects were upright (see Fig. 1, top panel), whereas in Experiments 2A and 2B, they were inverted by rotating them 180 degrees in the picture-plane (Fig. 1, bottom panel). For Experiments 1A and 2A (behavioral paradigm), four perceptual masks (1024 × 768 pixels) containing random squares were generated with the Masked Priming Toolbox [55].

2.3. Apparatus

Stimuli were presented on a 24-inch monitor with a refresh rate of 144 Hz. Presentation was controlled using MATLAB and the Psychophysics Toolbox [56,57]. Participants viewed stimuli at 60 cm distance in a dimly lit room, resulting in 26.56° of visual angle horizontally and 20.03° vertically for scenes, and 6.75° both horizontally and vertically for thumbnail objects.

2.4. Procedure

2.4.1. Behavioral paradigm

The procedure was based on a previous study [28]. In Experiments 1A and 2A, participants started each of the six practice trials and 288 experimental trials (see trial sequence in Fig. 2) by pressing the space bar on a computer keyboard. First, a fixation cross was presented for 304 ms, immediately followed by a blank screen for 200 ms, a preview of the background image (either an upright scene, inverted scene, or scrambled scene) for 104 ms, the consistent or inconsistent target object (upright in Experiment 1A; inverted in Experiment 2A) superimposed on the same background image for 56 ms, and a dynamic mask (4 × 56 ms). Subsequently, an input panel appeared, where participants entered the name of the target object. They were instructed to name the object as precisely as possible, and to enter their best guess if they had missed an object or were uncertain about their response. Note that it was emphasized that background images were task-irrelevant. Responses were self-paced and validated via keypress. At the end of each trial, participants were asked how confident they were about their response on a scale from one (very unconfident) to six (very confident). Participants were instructed to fixate the center of the screen during object-scene presentations.

2.4.2. ERP paradigm

The procedure was based on a previous study [28]. In Experiments 1B and 2B, participants completed six practice trials and 336 main trials (see trial sequence in Fig. 3) out of which 288 were experimental trials and 48 were intermixed Repetition Detection Task trials (RDT, see [26] for similar procedure). The latter were included to ensure that participants attended the stimuli. Experimental and RDT trials were identical in procedure as outlined below. In the beginning of each trial, a red fixation cross was presented until the space bar was pressed. Participants were instructed to fixate the cross and to blink, if necessary, before pressing the space bar to minimize blinks during the trial. Upon keypress, the fixation cross remained on the screen for 1000 ms plus a random jitter between 0 and 300 ms. Then, the background image (either upright scene, inverted scene or scrambled scene) and the critical object (upright in Experiment 1B; inverted in Experiment 2B) were presented simultaneously for 2000 ms. Participants were instructed to fixate the critical object in the center of the screen and to avoid blinks. Then, a green fixation cross (indicating the occurrence of a task) was presented for 2000 ms. Participants were instructed to press a key during this time if they thought they had seen the same object-scene combination (as shown on the previous frame) before at any time during the experiment but not if they believed it to be a novel object-scene combination or a lure (i.e., either a certain object presented again but on a different background image or the same background presented again but paired with a different object). If the key was pressed, feedback on whether the response was a hit or a false alarm was provided. If no key was pressed, participants received feedback on misses, whereas feedback on correct rejections was not provided. All background images and objects that were included in the RDT (16 exact repetitions, 8 lures) were collected in addition to the main stimulus set and excluded from the ERP analysis. Participants did not know which object-scene combinations would be repeated or how many trials were in between repetitions. Repetitions occurred up to ten trials after initial presentation.

2.5. Preprocessing

2.5.1. Behavioral data

For Experiments 1A and 2A, the analysis was conducted as follows: object naming responses were evaluated by three independent raters (undergraduates of Psychology at the Goethe University Frankfurt) who were blind to condition (see also [28]). Raters were instructed to consider responses as correct if they matched a sample solution [28] or a synonym of it, regardless of spelling mistakes. Critically, less precise responses (e.g., “fruit” instead of “apple”) were considered as incorrect [18]. If the majority of the raters considered a response as correct it was deemed correct; otherwise it was incorrect. For display purposes, we calculated the proportion of correct responses and mean confidence ratings per participant as a function of the consistency manipulation

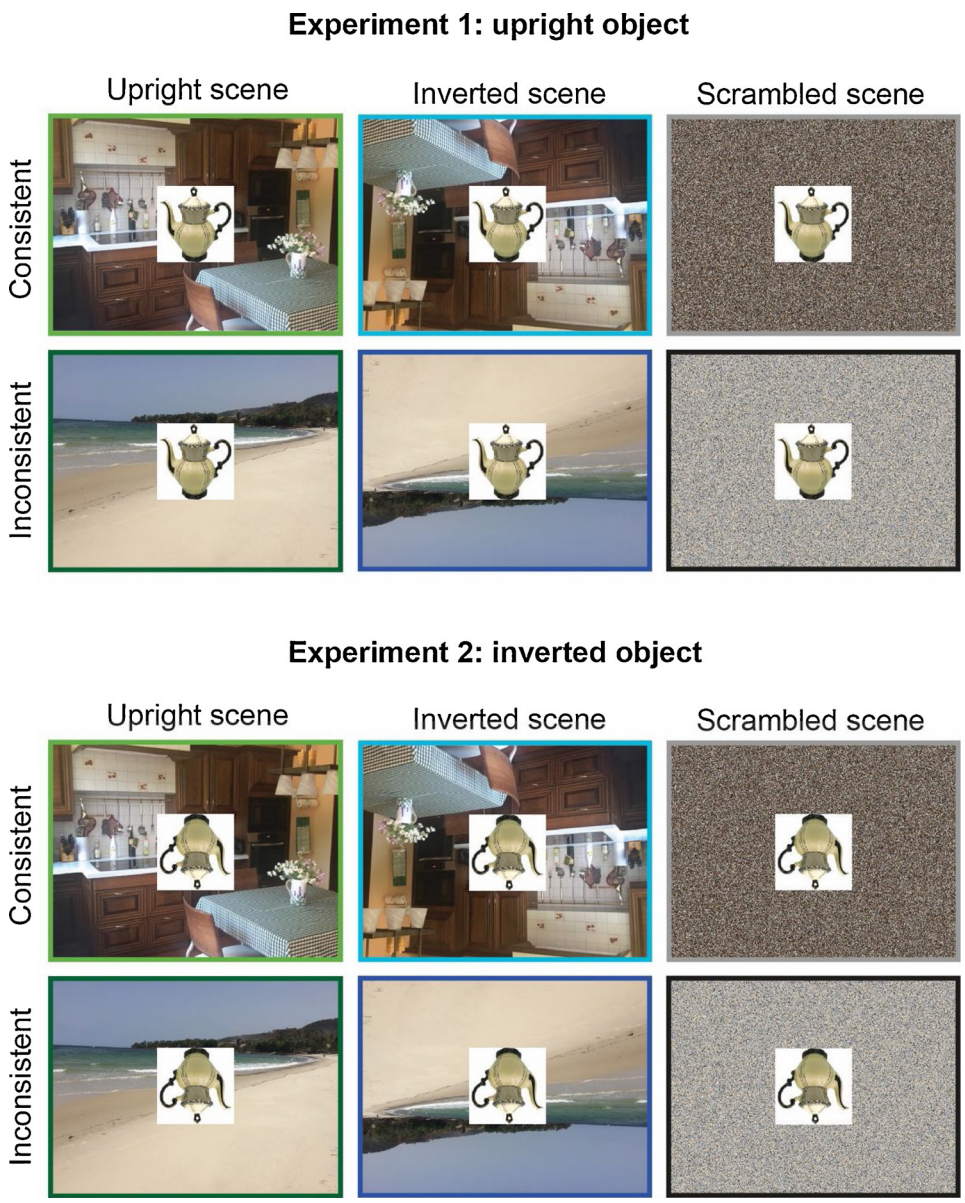

Fig. 1. Examples of object-background combinations presented in Experiment 1 (top panel) and Experiment 2 (bottom panel). Frame colors were added for illustration purposes.

and the type of background image.

2.5.2. EEG data

The electroencephalogram (EEG) was recorded from the scalp using 64 active electrodes (actiChamp, Brain Products, Germany) with a sampling rate of 1000 Hz. Electrodes were positioned according to the common 10–20 system. Two electrodes placed on the left and right mastoids served as reference and one electrode below the left eye served as an EOG channel. EEG signals were band-pass filtered (0.1 − 45 Hz) offline. The data was segmented into 1200 ms epochs (−200 to 1000 ms), time locked to stimulus onset, and baseline corrected by subtracting the mean voltage in the 200 ms prior to stimulus onset. All epochs that were part of the RDT (N = 48 out of 336 epochs per participant) and experimental epochs in which a false alarm occurred (Experiment 1B: M = 2, min = 0, max = 11 epochs per participant; Experiment 2B: M = 1.96, min = 0, max = 13 epochs per participant) were excluded from further analysis. Noisy electrodes were removed after visual inspection. Epochs with gross artefacts were removed automatically when the signal exceeded a threshold of +/− 500 microvolt at any channel. Remaining epochs were fed into an independent component analysis (ICA) [58] to correct for EOG artefacts (e.g., blinks, eye-movements). Following removal of artefactual ICA components, the removed electrodes were interpolated. Then, experimental epochs containing extant artefacts were rejected with a semi-automatic procedure [59,60] by tailoring an absolute voltage threshold and a moving window peak-to-peak threshold to each participant's data (see Supplementary information). In Experiment 1B, on average, 45.83 experimental epochs were retained per condition (range: 36–48). In Experiment 2B, on average, 46.21 epochs were retained per condition (range: 32–48).

The influence of scene context on semantic object processing was examined as follows. We calculated the mean amplitude for the mid-central region (averaged over the electrodes FC1, FCz, FC2, C1, Cz, C2, CP1, CPz, CP2) per participant, per trial for two pre-defined time windows, that is, the N300 (250 − 350 ms) and N400 window (350 − 600 ms). The choice of the region and time windows of interest was based on previous work ([26]; see also [27,28]). EEG pre-processing and analysis was conducted in EEGLAB [59] and ERPLAB [60]. For

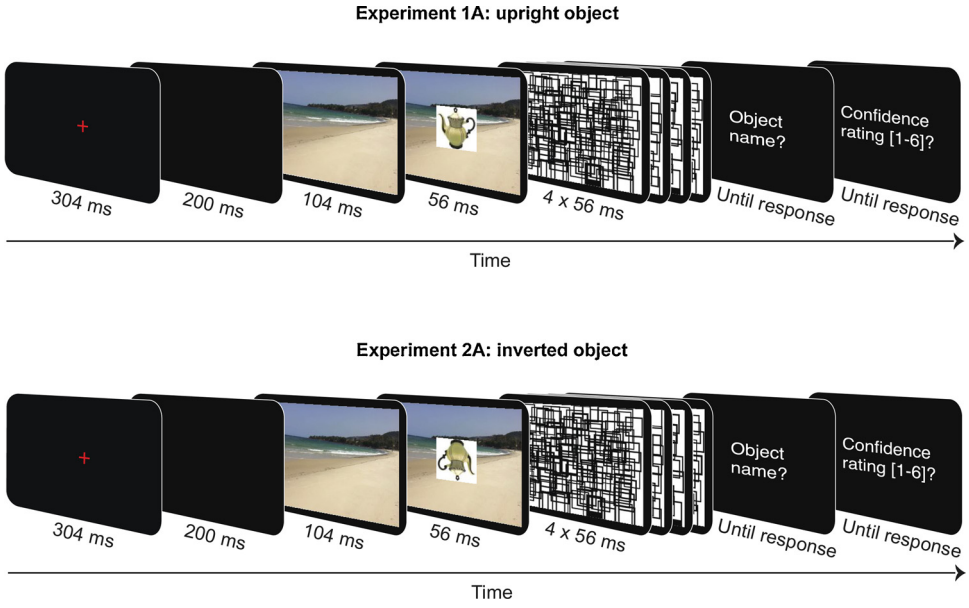

Fig. 2. Trial sequence of Experiment 1A and 2A (behavioral paradigm). Note that both sequences are identical except for the orientation of the critical object.

display purposes, a grand-average waveform per condition was calculated and low-pass filtered at 30 Hz.

2.6. Analysis

Single trial data were analyzed with generalized or linear mixed-effects models (GLMM/LMM) using lme4 [61], a package for the programming environment R [62]. For each dependent variable (behavioral data: object naming accuracy, confidence rating; EEG data: N300 amplitude, N400 amplitude), three planned contrasts were calculated (as in [28]). Data points for consistent objects were compared with those for inconsistent objects per background type using least-square means (R packages lsmeans and emmeans [63]). Moreover, we assessed if there were main effects for upright and inverted scenes compared to scrambled scenes, which served as baseline. We also assessed if there was a main effect of consistency. Further, we assessed if there was an interaction between upright and scrambled scenes given treatment contrasts for consistency (consistent vs. inconsistent). Similarly, we checked if there was an interaction between inverted and scrambled scenes. Subsequently, only for significant interactions in the behavioral paradigm (Experiments 1A and 2A), we performed post-hoc tests. In this case, we first contrasted object naming accuracy for consistent

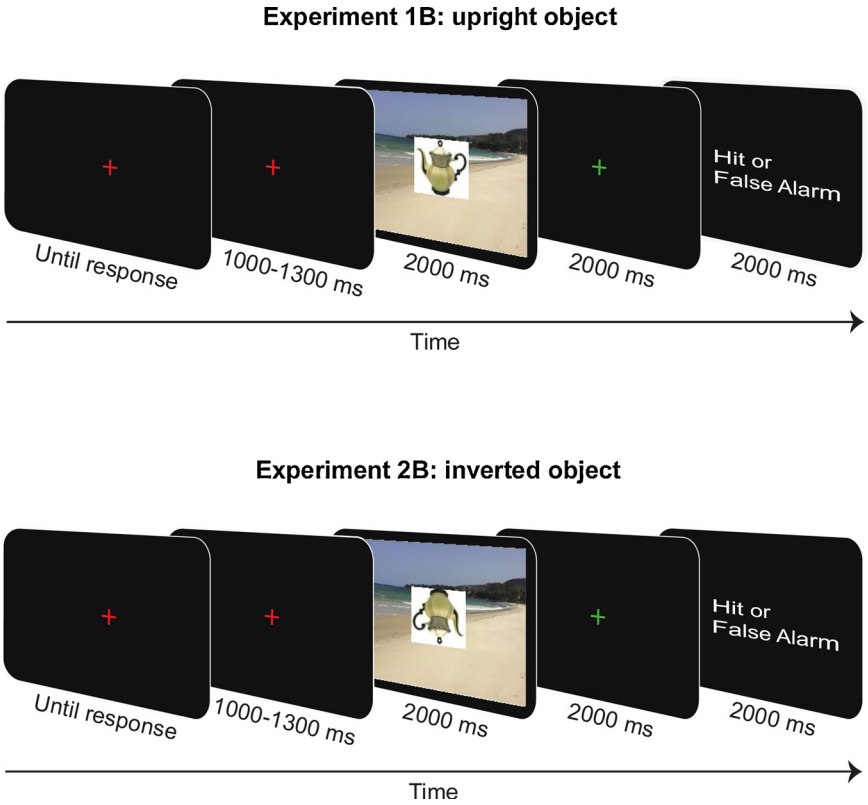

Fig. 3. Trial sequence of ERP Experiments 1B (top) and 2B (bottom). Note that both sequences are identical except for the orientation of the critical object. The last frame ("Hit or False Alarm") appeared only if participants pressed a key while the green fixation cross was presented; they were instructed to do so when spotting an exact repetition, that is, when they thought they had seen the same object-background combination before at any time during the experiment. If no key was pressed, participants received feedback on misses, whereas feedback on correct rejections was not provided.

(upright or inverted) scenes with consistent scrambled scenes to see if there was facilitation of object recognition. Conversely, we then contrasted accuracy for inconsistent (upright or inverted) scenes with inconsistent scrambled scenes. These latter two types of comparisons with the control condition were not based on our initial hypotheses, therefore considered post-hoc and were p-value adjusted using the Holm correction in the R package lsmeans [63].

As fixed effects, each model included the type of scene background (upright, inverted, scrambled) and the consistency (consistent, inconsistent) with an interaction term. The random effects structure was initially set maximal [64] with random intercepts for participants, scene categories and items (i.e., individual scenes) included, as well as random slopes for participants and scene categories. Typically, models with random intercepts and slopes for all fixed effects fail to converge or lead to overparameterization [65]. To determine models that converge on a stable solution and are properly supported by the data, we used a Principal Components Analysis (PCA) of the random-effects variance-covariance estimates for each fitted mixed-effects model to identify overparameterization [65]. Random slopes not supported by the PCA and not contributing significantly to the goodness of fit in likelihood ratio tests were removed from the models. Item-related slopes (i.e., for scene category) were removed first, and then – if necessary – participant-related slopes were removed. The resulting best fitting models' retained variance components were as follows. Note that all planned comparisons and post-hoc tests were run on these final, best fitting models, and that maximum likelihood estimation was used to fit all models.

2.6.1. Behavioral data

Intercepts for participants, scene categories, and items were included as well as a by-category random slope for consistency (consistent, inconsistent). All models were GLMMs with either a Binomial distribution (object naming accuracy) or a Poisson distribution (confidence ratings). P-values were obtained from asymptotic Wald tests.

2.6.2. EEG data

ERP amplitudes were centered and scaled for each of the two time windows of interest before fitting the LMMs. Intercepts for participants, scene categories, and items were included as well as a by-participant random slope for the type of background image (upright scene, inverted scene, scrambled scene; note that this slope was removed from the N400 model in Experiment 1B). In addition, the N400 model in Experiment 2B included a by-category random slope for the type of background image. P-values were obtained using Satterthwaite's degrees of freedom method (using R packages lmerTest, lsmeans, and emmeans; [63,66]).

3. Results

3.1. Experiment 1A: Behavioral paradigm

3.1.1. Object naming accuracy

Fig. 4 (left panel) depicts the object naming accuracies for consistent versus inconsistent upright objects on upright scenes, inverted scenes, and scrambled scenes. Planned contrasts for consistent versus inconsistent objects yielded a significant difference for upright scenes, $ß = 0.73$, $SE = 0.27$, $z_{ratio} = 2.704$, $p = 0.007$, but not for inverted scenes, $ß = 0.373$, $SE = 0.269$, $z_{ratio} = 1.384$, $p = 0.166$, or scrambled scenes, $|z_{ratio}| < 1$. Compared to scrambled scenes (baseline), we found main effects for upright scenes, $ß = 0.195$, $SE = 0.097$, $z = 2.023$, $p = 0.043$, and inverted scenes, $ß = -0.2$, $SE = 0.095$, $z = -2.102$, $p = 0.036$. The main effect of consistency was not significant, $|z| < 1$. There was an interaction between scrambled scenes and upright scenes regarding the consistency manipulation, $ß = -0.594$, $SE = 0.136$, $z = -4.374$, $p < 0.001$. There was no interaction between scrambled and inverted scenes, $ß = -0.237$, $SE = 0.135$, $z = -1.758$, $p = 0.079$. Post-hoc, p-value adjusted comparisons revealed a significant difference for consistent objects on upright versus scrambled scenes, $ß = -0.195$, $SE = 0.097$, $z_{ratio} = -2.023$, $p = 0.043$, as well as for inconsistent objects on upright versus scrambled scenes, $ß = 0.398$, $SE = 0.095$, $z_{ratio} = 4.181$, $p < 0.001$.

3.1.2. Confidence ratings

Fig. 4 (right panel) illustrates the participants' confidence ratings. Planned contrasts for consistent versus inconsistent objects yielded a significant difference for upright scenes, $ß = 0.145$, $SE = 0.053$, $z_{ratio} = 2.725$, $p = 0.006$, but not for inverted scenes, $ß = 0.076$, $SE = 0.053$, $z_{ratio} = 1.417$, $p = 0.157$, or scrambled scenes, $|z_{ratio}| < 1$. Compared to scrambled scenes (baseline), we did not find a main effect for upright scenes, $|z| < 1$, but for inverted scenes, $ß = -0.071$, $SE = 0.021$, $z = -3.318$, $p = 0.001$. The main effect of consistency was not significant, $|z| < 1$. There was an interaction between scrambled scenes and upright scenes regarding the consistency manipulation, $ß = -0.117$, $SE = 0.03$, $z = -3.853$, $p < 0.001$. There was no interaction between scrambled and inverted scenes, $ß = -0.048$, $SE = 0.031$, $z = -1.558$, $p = 0.119$.

3.2. Experiment 1B: ERP paradigm

3.2.1. Behavioral results

On average, the RDT yielded 13.88 hits (i.e., exact repetitions were correctly reported as such; $min = 9$, $max = 16$) and 1.67 false alarms (i.e. images that were part of the RDT were falsely reported as repetitions; $min = 0$, $max = 7$).

3.2.2. ERP results

Fig. 5 shows the grand-averaged ERPs per condition recorded from the mid-central region and corresponding scalp topographies of the difference between consistent and inconsistent objects in the N300 and N400 time windows.

3.2.2.1. N300 time window. Planned contrasts for consistent versus inconsistent objects yielded a significant difference for upright scenes, $ß = 0.142$, $SE = 0.034$, $t_{ratio} = 4.135$, $p < 0.001$, but not for inverted scenes, $ß = 0.057$, $SE = 0.034$, $t_{ratio} = 1.664$, $p = 0.096$, or scrambled scenes, $ß = -0.038$, $SE = 0.035$, $t_{ratio} = -1.101$, $p = 0.271$. Compared to scrambled scenes (baseline), we found no main effects for upright scenes, $|t| < 1$, or inverted scenes, $ß = -0.068$, $SE = 0.043$, $t = -1.593$, $p = 0.118$. The main effect of consistency was not significant, $ß = 0.038$, $SE = 0.035$, $t = 1.101$, $p = 0.271$. There was an interaction between scrambled scenes and upright scenes regarding the consistency manipulation, $ß = 0.18$, $SE = 0.049$, $t = -3.696$, $p < 0.001$. The interaction between scrambled and inverted scenes was nearly significant, $ß = -0.095$, $SE = 0.049$, $t = -1.954$, $p = 0.051$.

3.2.2.2. N400 time window. Planned contrasts for consistent versus inconsistent objects yielded a significant difference for upright scenes, $ß = 0.113$, $SE = 0.035$, $t_{ratio} = 3.270$, $p = 0.001$, but not for inverted scenes, $|t_{ratio}| < 1$, or scrambled scenes, $ß = -0.045$, $SE = 0.0347$, $t_{ratio} = -1.297$, $p = 0.195$. Compared to scrambled scenes (baseline), we found a main effect for upright scenes, $ß = -0.083$, $SE = 0.035$, $t = -2.378$, $p = 0.017$, and inverted scenes, $ß = -0.118$, $SE = 0.035$, $t = -3.397$, $p < 0.001$. The main effect of consistency was not significant, $ß = 0.045$, $SE = 0.035$, $t = 1.297$, $p = 0.195$. There was an interaction between scrambled scenes and upright scenes regarding the consistency manipulation, $ß = -0.158$, $SE = 0.049$, $t = -3.225$, $p = 0.001$. There was no interaction between scrambled and inverted scenes, $ß = -0.075$, $SE = 0.049$, $t = -1.540$, $p = 0.124$.

In sum, in Experiment 1A, we found a consistency effect for upright objects on upright but not inverted or scrambled scenes. In line with this, upright objects on upright scenes elicited N300/N400 responses in

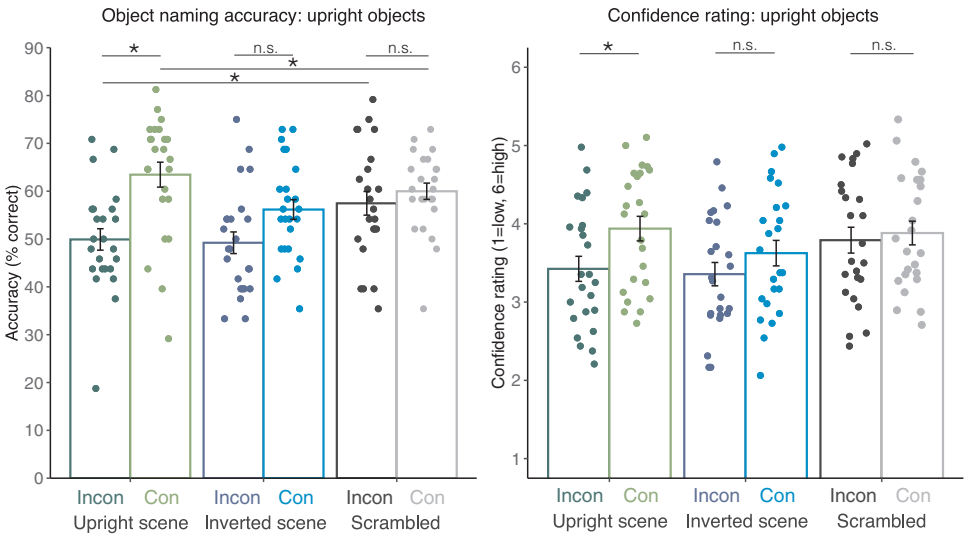

Fig. 4. Object naming accuracy (left panel) and confidence rating (right panel) for inconsistent (Incon) and consistent (Con) upright objects presented on upright scenes, inverted scenes, and scrambled scenes (control condition). Error bars show the standard error of the mean. Significant comparisons ($p < 0.05$) are marked with asterisks; non-significant comparisons are marked with "n.s.".

Experiment 1B, suggesting that object processing was modulated by upright but not by inverted or scrambled scenes. In Experiment 2, we explore the influence of object inversion, using the same paradigms.

3.3. Experiment 2A: Behavioral paradigm

3.3.1. Object naming accuracy

Fig. 6 (left panel) depicts the object naming accuracies for consistent versus inconsistent inverted objects on upright scenes, inverted scenes, and scrambled scenes. Planned contrasts for consistent versus inconsistent objects yielded a significant difference for upright scenes, $ß = 0.663$, $SE = 0.196$, $z_{ratio} = 3.390$, $p < 0.001$, and inverted scenes, $ß = 0.484$, $SE = 0.195$, $z_{ratio} = 2.480$, $p = 0.013$, but not for scrambled scenes, $|z_{ratio}| < 1$. Compared to scrambled scenes (baseline), we found main effects for upright scenes, $ß = 0.282$, $SE = 0.094$, $z = 2.999$, $p = 0.003$, and inverted scenes, $ß = 0.298$, $SE = 0.094$, $z = 3.164$, $p = 0.002$. The main effect of consistency was not significant, $|z| < 1$. There was an interaction between scrambled scenes and upright scenes regarding the consistency manipulation, $ß = -0.704$, $SE = 0.134$, $z = -5.254$, $p < 0.001$. There was also an interaction between scrambled and inverted scenes, $ß = -0.525$, $SE = 0.133$, $z = -3.935$, $p < 0.001$. Post-hoc, p-value adjusted comparisons revealed a significant difference for consistent objects on upright versus scrambled scenes, $ß = -0.282$, $SE = 0.094$, $z_{ratio} = -2.999$, $p = 0.005$, inconsistent objects on upright versus scrambled scenes, $ß = 0.422$, $SE = 0.095$, $z_{ratio} = 4.428$, $p < 0.001$, consistent objects on inverted versus scrambled scenes, $ß = -0.298$, $SE = 0.094$, $z_{ratio} = -3.164$, $p = 0.005$, as well as for inconsistent objects on inverted versus scrambled scenes, $ß = 0.228$, $SE = 0.095$, $z_{ratio} = 2.406$, $p = 0.016$.

3.3.2. Confidence ratings

The confidence ratings depicted a similar pattern of results (see Fig. 6, right panel). Planned contrasts for consistent versus inconsistent objects yielded a significant difference for upright scenes, $ß = 0.127$, $SE = 0.051$, $z_{ratio} = 2.502$, $p = 0.012$, but not inverted scenes, $ß = 0.092$, $SE = 0.051$, $z_{ratio} = 1.814$, $p = 0.07$, or scrambled scenes, $|z_{ratio}| < 1$. Compared to scrambled scenes (baseline), we did not find main effects for upright scenes, $ß = 0.03$, $SE = 0.023$, $z = 1.283$, $p = 0.2$, or inverted scenes, $ß = 0.028$, $SE = 0.023$, $z = 1.205$, $p = 0.228$. The main effect of consistency was not significant, $|z| < 1$. There was an interaction between scrambled scenes and upright scenes regarding the consistency manipulation, $ß = -0.133$, $SE = 0.033$, $z = -3.994$, $p < 0.001$. There was also an interaction between scrambled and inverted scenes, $ß = -0.098$, $SE = 0.033$, $z = -2.957$, $p = 0.003$.

3.4. Experiment 2B: ERP paradigm

3.4.1. Behavioral results

On average, the RDT yielded 13.13 hits (i.e., exact repetitions were correctly reported as such; $min = 9$, $max = 16$) and 2.29 false alarms (i.e., images that were part of the RDT were falsely reported as repetitions; $min = 0$, $max = 6$).

3.4.2. ERP results

Fig. 7 shows the grand-averaged ERPs per condition recorded from the mid-central region and corresponding scalp topographies of the difference between consistent and inconsistent objects in the N300 and N400 time windows.

3.4.2.1. N300 time window. Planned contrasts for consistent and inconsistent objects yielded a significant difference for upright scenes, $ß = 0.12$, $SE = 0.04$, $t_{ratio} = 3.044$, $p = 0.002$, but not for inverted scenes, $ß = 0.04$, $SE = 0.04$, $t_{ratio} = 1.004$, $p = 0.315$, or scrambled scenes, $|t_{ratio}| < 1$. Compared to scrambled scenes (baseline), we found no main effects for upright scenes, $|t| < 1$, or inverted scenes, $|t| < 1$. The main effect of consistency was not significant, $|t| < 1$. There was no interaction between scrambled scenes and upright scenes regarding the consistency manipulation, $ß = -0.093$, $SE = 0.056$, $t = -1.659$, $p = 0.097$. There was no interaction between scrambled and inverted scenes, $|t| < 1$.

3.4.2.2. N400 time window. Planned contrasts for consistent and inconsistent objects yielded a significant difference for upright scenes, $ß = 0.111$, $SE = 0.04$, $t_{ratio} = 2.771$, $p = 0.006$, and inverted scenes, $ß = 0.127$, $SE = 0.04$, $t_{ratio} = 3.149$, $p = 0.002$, but not for scrambled scenes, $|t_{ratio}| < 1$. Compared to scrambled scenes (baseline), we found no main effects for upright scenes, $ß = -0.076$, $SE = 0.061$, $t = -1.233$, $p = 0.231$, or inverted scenes, $ß = -0.103$, $SE = 0.057$, $t = -1.83$, $p = 0.078$. The main effect of consistency was not significant, $|t| < 1$. There was no interaction between scrambled scenes and upright scenes regarding the consistency manipulation, $ß = -0.108$, $SE = 0.057$, $t = -1.894$, $p = 0.058$. There was an interaction between scrambled and inverted scenes, $ß = -0.123$, $SE = 0.057$, $t = -2.164$, $p = 0.031$.

4. Discussion

The current study sheds light on the question whether scene context effects on object processing are orientation-dependent. Specifically, we investigated if inverted scenes (i.e., rotated 180 degrees) modulate

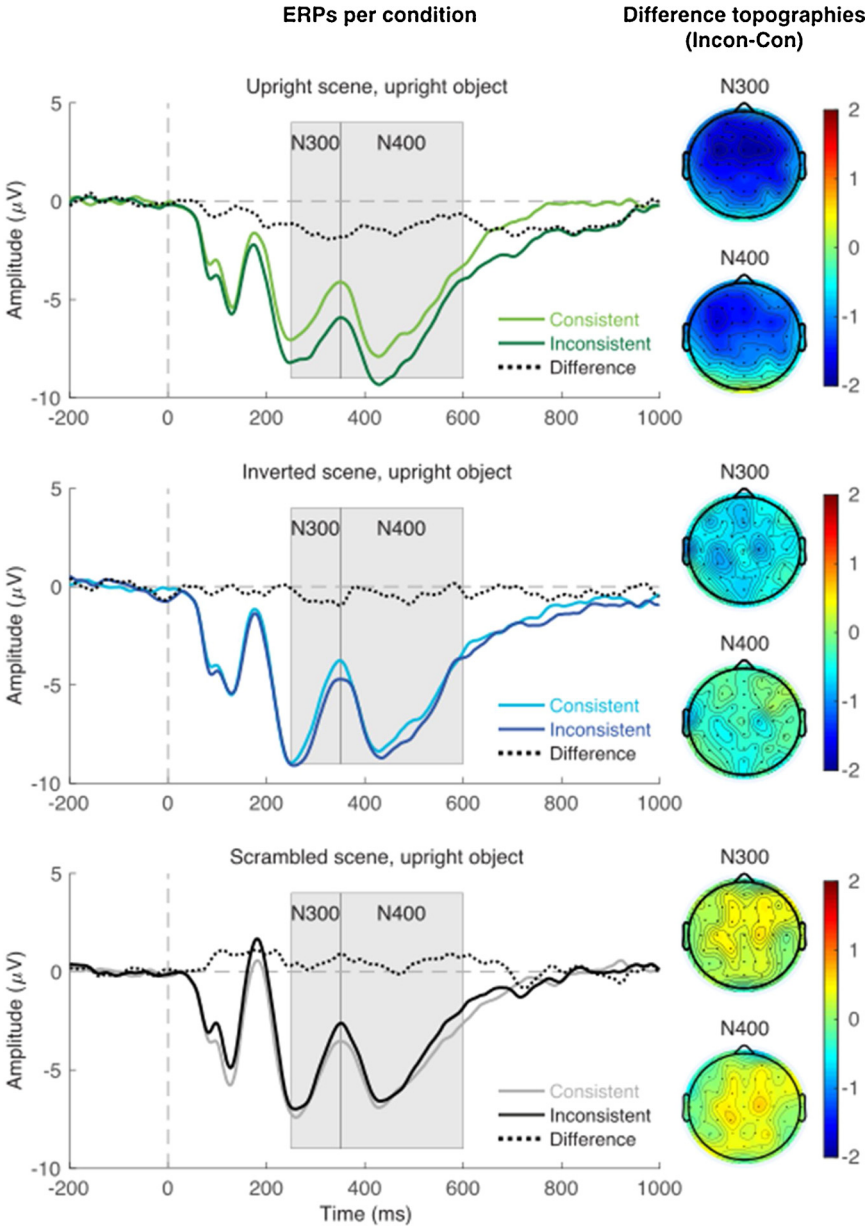

Fig. 5. Grand-average ERPs recorded from the mid-central region (electrodes FC1, FCz, FC2, C1, Cz, C2, CP1, CPz, CP2) for semantically consistent vs. inconsistent upright objects on upright scenes (top panel; in green), inverted scenes (middle; in blue), and scrambled scenes (bottom; in black), and corresponding topographies of the difference between consistent and inconsistent objects in the N300 and N400 time windows.

semantic object processing like upright scenes have been shown to (e.g., [18,24,26]), and if this modulation depends on the orientation of the critical object. Accordingly, the orientation manipulation was used to selectively render semantic access in object and scene perception more difficult in order to examine the processes underlying context effects. We presented consistent and inconsistent objects either upright (Experiment 1) or inverted (Experiment 2) on three types of scene backgrounds: upright, inverted, and scrambled (control condition).

In Experiment 1A (behavioral paradigm), we found that on upright scenes, consistent upright objects were named with higher accuracy and confidence than inconsistent ones. This finding is in line with previous reports of scene consistency effects for thumbnail objects superimposed on upright scenes [28], object cutouts embedded in scenes [18–21], as well as some earlier work using line drawings (e.g., [3,5]; but see [7,8]). For upright objects on inverted scenes, no such effects were observed, suggesting that object recognition was not modulated by the inverted scene context. To our knowledge, the current study is the first to investigate the effect of scene inversion on object identification as a function of semantic consistency. Some previous studies have reported minor or no effects as to the influence of scene inversion on object or animal detection [45–48]. However, it is important to consider task demands: superordinate-level object detection may result in weaker inversion effects than basic-level object naming (see [47,49]). For upright objects on scrambled scenes, we found no effects of the consistency manipulation.

In Experiment 1B (ERP paradigm), we recorded ERPs while participants were engaged in a Repetition Detection Task. Specifically, we looked at the N300/N400 components as online markers of object-scene semantic processing (e.g., [24–26,29]). On upright scenes, inconsistent relative to consistent upright objects elicited N300/N400 responses which is in line with previous studies that used thumbnail objects superimposed on scenes [28], object cutouts embedded in scenes [24,25,29,30], or natural photographs [26,27]. For upright objects on inverted scenes, no N300/N400 effects were found. In line with the

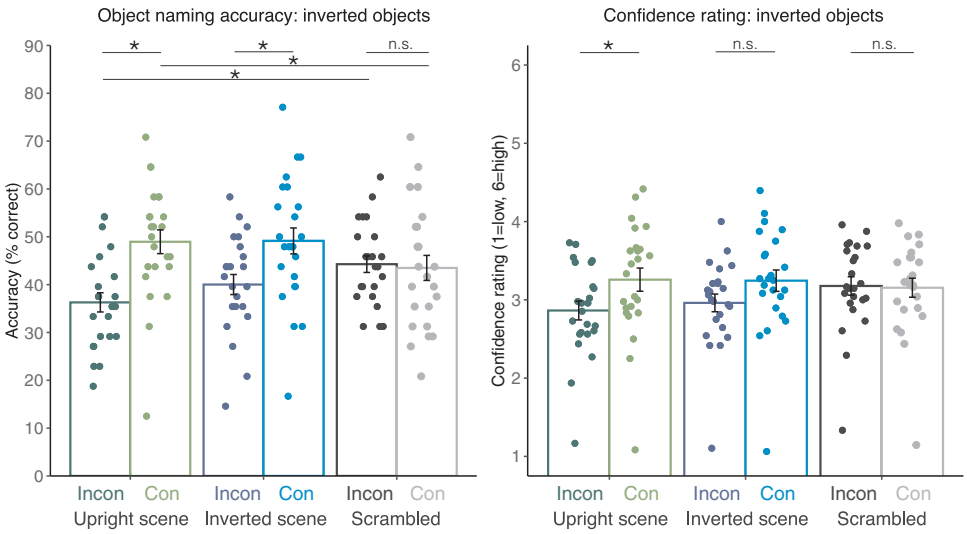

Fig. 6. Object naming accuracy (left panel) and confidence rating (right panel) for inconsistent (Incon) and consistent (Con) inverted objects presented on upright scenes, inverted scenes, and scrambled scenes (control condition). Error bars show the standard error of the mean. Significant comparisons ($p < 0.05$) are marked with asterisks; non-significant comparisons are marked with "n.s.".

behavioral results, this would imply that inverted scenes did not modulate the processing of upright thumbnail objects with respect to the consistency manipulation. Possibly, the lack of an ERP modulation can be accounted for by the incoherent, upright orientation of the thumbnail objects. In previous semantic violation studies, the scenes and the critical objects shared the same orientation (e.g., [18,24,26]). Therefore, we inverted all objects in Experiment 2 on the same types of background images as in Experiment 1.

In Experiment 2A (behavioral paradigm), we found that, on both upright and inverted scenes, consistent inverted objects were named with higher accuracy than inconsistent ones. This suggests that object recognition was modulated by both types of scene backgrounds and that the scene consistency effect does not rely on a coherent object-scene orientation. An effect on confidence ratings was only present for upright scenes. We found no consistency effect for scrambled scenes. So far, scene and object inversion have not been disentangled but note that, in one study, the orientations of objects and scenes were selectively manipulated by 90 degrees (yet without inversion or semantic manipulation); no dominance of scene orientation over object orientation, or vice versa, was found [47].

In Experiment 2B (ERP paradigm), for inverted objects on upright scenes, we found N300/N400 responses. Interestingly, while no N300 deflection was present for inverted objects on inverted scenes, we still found an N400 response. No effects were observed for scrambled scenes. These findings suggest that inverted scenes can modulate semantic object processing if the critical object is inverted as well. However, contextual influences seem to occur later in time. The later emergence of contextual influences does not seem to be explained by mere object inversion. The ERP effects hint at a dissociation of the contributions of object and scene inversion to interactive object-scene processing: on upright scenes, both upright and inverted objects evoked an N300 (see Experiments 1B, 2B), whereas for inverted objects on inverted scenes, no N300 but only an N400 was found (see Experiment 2B).

4.1. Contextual modulation by upright scenes

What are the processes underlying the context effects found for the upright objects on upright scenes in Experiment 1? The gist of the upright scenes may have been extracted very quickly [35,67] and, through sufficient activation of scene schemata, utilized to generate context-based predictions of related objects [1,32] that in turn may have eased access to the consistent target objects. Specifically, the activation of scene schemata and schema-congruent predictions may have lowered the perceptual threshold for the target objects [8]. One possible interpretation of our findings is that there was contextual facilitation of object identification as proposed in the criterion modulation model ([8]; see also the more recent framework of contextual facilitation [1]).

However, an important alternative interpretation should be considered: In line with the functional isolation model [8], the consistency manipulation might not have modulated object identification but interfered with participants' performance at a later, post-perceptual stage of processing, possibly through a semantic mismatch detection. We addressed this question by contrasting accuracies for consistent objects on upright and scrambled scenes (Experiment 1A, 2A). Indeed, we found a significant difference – irrespective of object orientation – indicating that there was some contextual facilitation on object processing in the behavioral paradigm. Accordingly, this finding conflicts with the view of functional isolation of scene and object identification claiming that object identification occurs independently from scene identification [7,8]. Davenport and Potter [18] found that object recognition was better for isolated objects (control condition) than for consistent objects in scenes but noted that isolated objects benefitted from a clear contour which makes figure ground segmentation unnecessary. Since, in the current study, both consistent objects in scenes and objects in scrambled scenes (control condition) were presented isolated on a white frame, the comparison with the control condition may be more meaningful in our study; it does not give the control condition a segmentation advantage and may more likely reveal indication of contextual facilitation. In addition, when contrasting accuracies for inconsistent upright objects on upright and scrambled scenes, we found strong evidence that semantic violations interfered with performance – irrespective of object orientation – possibly at a perceptual and/or post-perceptual stage.

Recording ERPs allowed us to look at the time course of context effects and the underlying processes more closely. It has been suggested that the N300 ERP component reflects identification difficulties as a result of matching routines [25,29]. That is, on inconsistent trials, the scene context may have yielded "misleading" predictions of likely objects that were matched with incoming information of the object prior to full object identification. Consequently, mismatches and failed attempts to identify the target object may have occurred. It has been suggested that the later N400 response reflects context effects on a semantic, conceptual level [24,25,29]. In line with the strong interference effect that we found in the behavioral data, this could indicate that there was interference on a post-perceptual stage, such as integration difficulties of the inconsistent object with the scene context. However, it should be noted that it is still debated whether the N300 component is functionally distinct from the N400 component [27,30]. A recent study from our group found shared neuronal activity patterns

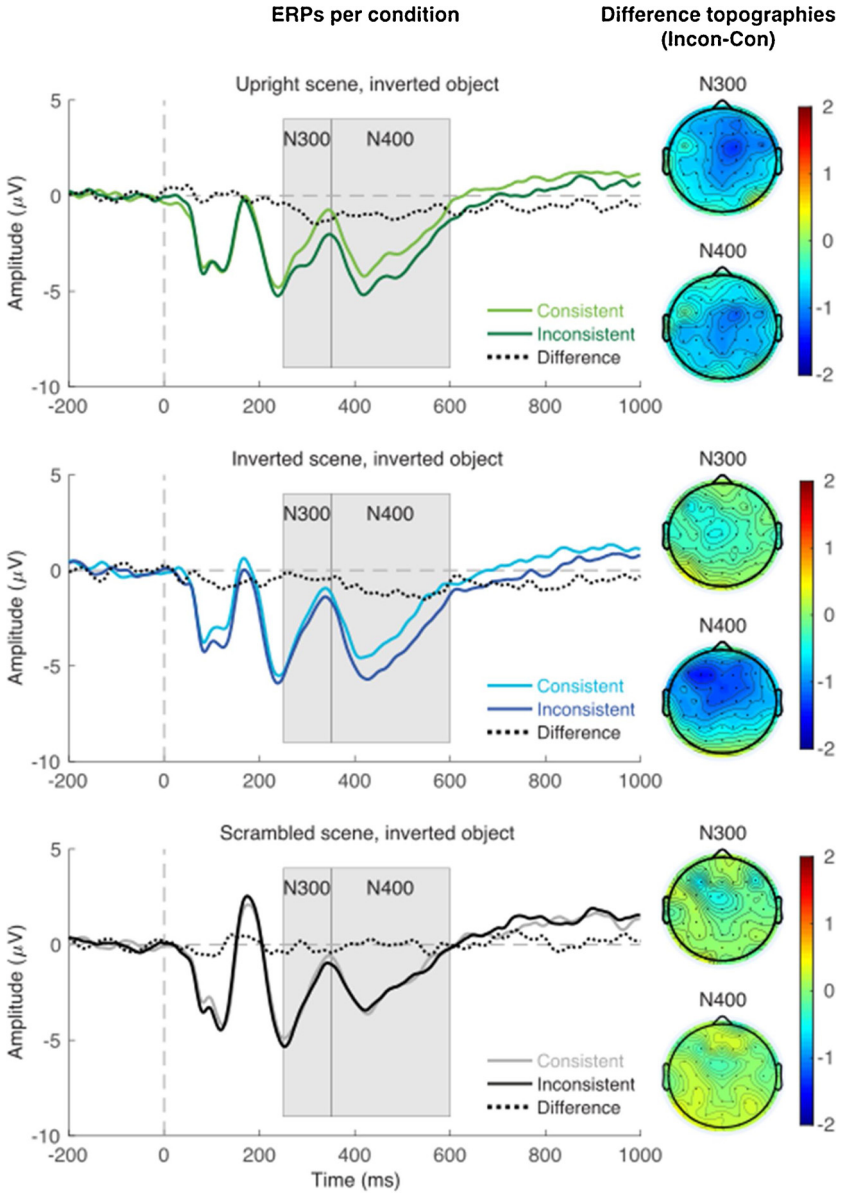

Fig. 7. Grand-average ERPs recorded from the mid-central region (electrodes FC1, FCz, FC2, C1, Cz, C2, CP1, CPz, CP2) for semantically consistent vs. inconsistent inverted objects on upright scenes (top panel; in green), inverted scenes (middle; in blue), and scrambled scenes (bottom; in black), and corresponding topographies of the difference between consistent and inconsistent objects in the N300 and N400 time windows.

across the two components using a time-generalized decoding approach [27]. While there was no evidence for a functional dissociation of the ERP components, this account does not rule out the possibility that there are some distinct processes that remained undetected. Notably, the N300 incongruency effect typically peaks more frontally on the scalp compared to the more central N400 component which may hint at some distinct underlying processes. Recently, Truman and Mudrik [30] manipulated both object-to-scene congruency and object identifiability: ERPs for congruent, identifiable objects diverged from ERPs for unidentifiable objects earlier than ERPs for incongruent, identifiable objects did. This implies that scene context influences object identification early on in the N300 time window. Further, the observation that ERPs for incongruent, identifiable objects diverged from ERPs for both identifiable and unidentifiable objects in the later N400 time window has been regarded as support for the functional distinction view of the N300/N400 components.

Together, our data for upright scenes suggest that scene context indeed affects object identification before it is completed, and thus not merely exerts influence on a post-perceptual stage. Possibly, object inversion may have rendered object identification slightly more difficult: when visually inspecting the ERPs, the N300 for inverted objects (Experiment 2B) appears to onset slightly later than for upright objects (Experiment 1B). Still, the rapid extraction of the upright scene's gist could have especially aided the recognition of the more difficult to perceive inverted objects.

4.2. Contextual modulation by inverted scenes

Scene inversion on the other hand may have resulted in slower or impaired processing of scene gist [39–41]. In turn, context-based predictions of the critical object may have been slowed down or impaired; identification of the upright object could have been achieved without such top-down predictions in Experiment 1B, where we found neither an N300 nor an N400 response for inverted scenes. That is, ambiguity, if any, could have been resolved by incoming perceptual information of the target object before context-based predictions could have

contributed. Similarly, in the behavioral paradigm (Experiment 1A), the gist of the inverted scene may not have been processed in time to aid (or hinder) upright object recognition – even though participants saw a brief preview of the scene before the critical object was superimposed.

For inverted objects on inverted scenes, we did find an N400 effect (Experiment 2B). Here, context-based predictions may have been slowed down or impaired as well but still may have influenced the identification of the *inverted* target object. That is, identification of the inverted object may have been more difficult and not completed before context-based predictions were available. Once available, such predictions may have eased access to the identity of the consistent inverted object, possibly by lowering the perceptual threshold for recognition (criterion modulation model, [8]), or hindered access to the inconsistent object. The absence of an N300 effect, that is typically present for upright objects on upright scenes, suggests that contextual modulation occurred at a later point in time. Again, we note that it is still debated if the N300 ERP component is functionally distinct from the N400 component in the case of upright stimuli [27,30]. Our ERP data do not resolve the question what processes underlie the N400 response observed for inverted scenes and to what extent they are similar to the N300 found for upright scenes (see also 4.3). The behavioral data corroborate that inverted scenes can facilitate object recognition given that performance for consistent inverted objects on inverted scenes was higher than for objects on scrambled scenes (Experiment 2A). This again conflicts with the view of functional isolation of scene and object identification. In addition, we found strong evidence that semantic violations interfered with performance – possibly at a perceptual and/or post-perceptual stage – given that performance for inconsistent inverted objects on inverted scenes was higher than for scrambled scenes.

4.3. Limitations and possible future directions

Although the current findings can be interpreted with the criterion modulation model ([8]; see also [1]) such that inverted scenes modulate the semantic processing of objects at later points in time, it remains unclear whether the inversions resulted in *qualitatively* or *quantitatively* different processes. In the face processing literature, one can find a long-standing debate about the cause of the face inversion effect that is potentially also informative for scene perception: It has been argued that face inversion causes a disruption of certain processes underlying face perception, such as holistic processing routines, and thus qualitative differences in the processing of upright versus inverted faces (e.g., [68]). By contrast, it has been postulated that due to our life-long experience with upright faces, the processing of inverted faces may simply be less efficient – quantitatively different – but does not require qualitatively different routines (e.g. [69]). Likewise, our current findings do not resolve the question whether scene inversion resulted in qualitatively different processes (e.g., impaired processing of scene gist via spatial layout information) or quantitative differences (e.g., slowed down processing of scene gist). In favor of the qualitative account, Epstein and colleagues [44] found less neuronal activity in the parahippocampal place area and enhanced activity in the lateral occipital object area to be associated with scene inversion, proposing a shift from "specialized processing streams towards generic object-processing mechanisms".

A possible limitation of this study is that the objects were superimposed on the background images within a white bounding box. Segmentation demands that naturally occur in object recognition were therefore not present here. However, we chose this approach to avoid variability in segmentation demands across different types of background images which might affect the time course of context effects on semantic object processing differentially (see also [28]). Specifically, when comparing accuracies for scenes and scrambled scenes (control condition), we found indication of facilitation which might have been obscured if the objects had been embedded in the scenes.

Future studies could replicate these findings in a more naturalistic setup, and/or extend it by using other stimuli such as co-occurring objects or words. Moreover, another interesting way to follow-up on this line of work could be to use magnetoencephalography or functional magnetic resonance imaging – possibly using a decoding approach – to better understand the "when", "where", and "how" of context effects.

5. Conclusion

In sum, the present behavioral findings suggest that while upright scenes modulate object recognition irrespective of object orientation, inverted scenes only modulate the recognition of inverted objects. In line with these findings, ERPs showed that inverted scenes can modulate semantic object processing if the object is inverted too, as seen in an N400 effect known to reflect object-scene semantic processing for upright stimuli. The lack of an N300 effect for inverted objects on inverted scenes provides first evidence that object-scene inversion causes contextual influences to occur later in time, possibly driven by delayed or impaired scene gist processing. The later emergence of contextual influences does not seem to be explained by mere object inversion; rather, the ERP effects hint at a dissociation of the contributions of object and scene inversion to interactive object-scene processing. Taken together, these results show that the orientation of both objects and scenes as well as their relationship to each other modulate ongoing object identification.

Data availability statement

Data and analysis scripts are available for download under this link: https://osf.io/2x4wz/?view_only=b3578cb175ab4498a2971c6def5afede.

Funding

This work was supported by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) – project number 222641018 – SFB/TRR 135TP C7.

Author contributions

T.L., V.W., and M.L.V. conceived the experiments. T.L.,V.W, and L.M. collected the data. T.L., V.W., L.M., and M.L.V conducted the analyses. T.L. drafted the first version of the manuscript. T.L., V.W., L.M., and M.L.V revised the manuscript and approved its final version.

Declarations of Competing Interest

None.

Acknowledgements

We wish to thank Julia Kunz, Aylin Kallmayer, Melvin Kallmayer, Lotte Kirschbaum, and Leila Zacharias for assisting in the experiments and evaluating the behavioral data. Further, we thank Benjamin Gagl, Naomi Vingron, and Dejan Draschkow for valuable discussions on the modelling.

References

[1] M. Bar, Visual objects in context, Nat. Rev. Neurosci. 5 (2004) 617–629, https://doi.org/10.1038/nrn1476.
[2] A. Oliva, A. Torralba, The role of context in object recognition, Trends Cogn. Sci. 11 (2007) 520–527, https://doi.org/10.1016/j.tics.2007.09.009.
[3] I. Biederman, R.J. Mezzanotte, J.C. Rabinowitz, Scene perception: detecting and judging objects undergoing relational violations, Cogn. Psychol. 14 (1982) 143–177, https://doi.org/10.1016/0010-0285(82)90007-X.
[4] S.J. Boyce, A. Pollatsek, Identification of objects in scenes: the role of scene background in object naming, J. Exp. Psychol. Learn. Mem. Cogn. 18 (1992) 531–543,

https://doi.org/10.1037/0278-7393.18.3.531.
[5] S.J. Boyce, A. Pollatsek, K. Rayner, Effect of background information on object identification, J. Exp. Psychol. Hum. Percept. Perform. 15 (1989) 556–566, https://doi.org/10.1037/0096-1523.15.3.556.
[6] T.E. Palmer, The effects of contextual scenes on the identification of objects, Mem. Cognit. 3 (1975) 519–526, https://doi.org/10.3758/BF03197524.
[7] A. Hollingworth, J.M. Henderson, Does consistent scene context facilitate object perception? J. Exp. Psychol. Gen. 127 (1998) 398–415, https://doi.org/10.1037/0096-3445.127.4.398.
[8] A. Hollingworth, J.M. Henderson, Object identification is isolated from scene semantic constraint: evidence from object type and token discrimination, Acta Psychol. (Amst.) 102 (1999) 319–343, https://doi.org/10.1016/S0001-6918(98)00053-5.
[9] T.H.W. Cornelissen, M.L.-H. Võ, Stuck on semantics: processing of irrelevant object-scene inconsistencies modulates ongoing gaze behavior, Attention Percept. Psychophys. 79 (2017) 154–168, https://doi.org/10.3758/s13414-016-1203-7.
[10] P. De Graef, D. Christiaens, G. Ydewalle, Perceptual effect of scene context on object identification, Psychol. Res. (1990) 317–329, https://doi.org/10.1007/BF00868064.
[11] A. Friedman, Framing pictures: the role of knowledge in automatized encoding and memory for gist, J. Exp. Psychol. Gen. 108 (1979) 316–355, https://doi.org/10.1037/0096-3445.108.3.316.
[12] J.M. Henderson, P.A.J. Weeks, A. Hollingworth, The effects of semantic consistency on eye movements during complex scene viewing, J. Exp. Psychol. Hum. Percept. Perform. 25 (1999) 210–228, https://doi.org/10.1037/0096-1523.25.1.210.
[13] G.R. Loftus, N.H. Mackworth, Cognitive determinants of fixation location during picture viewing, J. Exp. Psychol. Hum. Percept. Perform. 4 (1978) 565–572, https://doi.org/10.1037/0096-1523.4.4.565.
[14] M.L.-H, J.M. Henderson, Does gravity matter? Effects of semantic and syntactic inconsistencies on the allocation of attention during scene perception, J. Vis. 9 (2009) 1–15, https://doi.org/10.1167/9.3.24.
[15] M.L.-H, J.M. Henderson, Object–scene inconsistencies do not capture gaze: evidence from the flash-preview moving-window paradigm, Attention Percept. Psychophys. 73 (2011) 1742–1753, https://doi.org/10.3758/s13414-011-0150-6.
[16] M.L.-H. Võ, S.E. Boettcher, D. Draschkow, Reading scenes: how scene grammar guides attention and aids perception in real-world environments, Curr. Opin. Psychol. 29 (2019) 205–210, https://doi.org/10.1016/j.copsyc.2019.03.009.
[17] M.L.-H. Võ, J.M. Wolfe, The role of memory for visual search in scenes, Ann. N. Y. Acad. Sci. 1339 (2015) 72–81, https://doi.org/10.1111/nyas.12667.
[18] J.L. Davenport, M.C. Potter, Scene consistency in object and background perception, Psychol. Sci. 15 (2004) 559–564, https://doi.org/10.1111/j.0956-7976.2004.00719.x.
[19] J.L. Davenport, Consistency effects between objects in scenes, Mem. Cognit. 35 (2007) 393–401, https://doi.org/10.3758/BF03193280.
[20] J. Munneke, V. Brentari, M.V. Peelen, The influence of scene context on object recognition is independent of attentional focus, Front. Psychol. 4 (2013) 1–10, https://doi.org/10.3389/fpsyg.2013.00552.
[21] G. Sastyin, R. Niimi, K. Yokosawa, Does object view influence the scene consistency effect? Attention Percept. Psychophys. 77 (2015) 856–866, https://doi.org/10.3758/s13414-014-0817-x.
[22] T. Brandman, M.V. Peelen, Interaction between scene and object processing revealed by human fMRI and MEG decoding, J. Neurosci. 37 (2017) 7700–7710, https://doi.org/10.1523/jneurosci.0582-17.2017.
[23] M. Kutas, K.D. Federmeier, Thirty years and counting: finding meaning in the N400 component of the event-related brain potential (ERP), Annu. Rev. Psychol. 62 (2011) 621–647, https://doi.org/10.1146/annurev.psych.093008.131123.
[24] G. Ganis, M. Kutas, An electrophysiological study of scene effects on object identification, Cogn. Brain Res. 16 (2003) 123–144, https://doi.org/10.1016/S0926-6410(02)00244-6.
[25] L. Mudrik, D. Lamy, L.Y. Deouell, ERP evidence for context congruity effects during simultaneous object-scene processing, Neuropsychologia 48 (2010) 507–517, https://doi.org/10.1016/j.neuropsychologia.2009.10.011.
[26] M.L.-H. Võ, J.M. Wolfe, Differential electrophysiological signatures of semantic and syntactic scene processing, Psychol. Sci. 24 (2013) 1816–1823, https://doi.org/10.1177/0956797613476955.
[27] D. Draschkow, E. Heikel, M.L.-H, C.J. Fiebach, J. Sassenhagen, No evidence from MVPA for different processes underlying the N300 and N400 incongruity effects in object-scene processing, Neuropsychologia 120 (2018) 9–17, https://doi.org/10.1016/j.neuropsychologia.2018.09.016.
[28] T. Lauer, T.H.W. Cornelissen, D. Draschkow, V. Willenbockel, M.L.H. Võ, The role of scene summary statistics in object recognition, Sci. Rep. 8 (2018) 1–12, https://doi.org/10.1038/s41598-018-32991-1.
[29] L. Mudrik, S. Shalgi, D. Lamy, L.Y. Deouell, Synchronous contextual irregularities affect early scene processing: Replication and extension, Neuropsychologia 56 (2014) 447–458, https://doi.org/10.1016/j.neuropsychologia.2014.02.020.
[30] A. Truman, L. Mudrik, Are incongruent objects harder to identify? The functional significance of the N300 component, Neuropsychologia 117 (2018) 222–232, https://doi.org/10.1016/j.neuropsychologia.2018.06.004.
[31] A. Oliva, A. Torralba, Modeling the shape of the scene: a holistic representation of the spatial envelope, Int. J. Comput. Vis. 42 (2001) 145–175, https://doi.org/10.1023/A:1011139631724.
[32] S. Trapp, M. Bar, Prediction, context, and competition in visual recognition, Ann. N. Y. Acad. Sci. 1339 (2015) 190–198, https://doi.org/10.1111/nyas.12680.
[33] M.S. Castelhano, J.M. Henderson, The influence of color on the perception of scene gist, J. Exp. Psychol. Hum. Percept. Perform. 34 (2008) 660–675, https://doi.org/10.1037/0096-1523.34.3.660.
[34] A. Oliva, A. Torralba, Building the gist of a scene: the role of global image features in recognition, Prog. Brain Res. 155 (2006) 23–36, https://doi.org/10.1016/S0079-6123(06)55002-2.
[35] G.A. Rousselet, O.R. Joubert, M. Fabre-Thorpe, How long to get to the "gist" of real-world natural scenes? Vis. Cogn. 12 (2005) 852–877, https://doi.org/10.1080/13506280444000553.
[36] T.F. Brady, A. Shafer-Skelton, G.A. Alvarez, Global ensemble texture representations are critical to rapid scene perception, J. Exp. Psychol. Hum. Percept. Perform. 43 (2017) 1160–1176, https://doi.org/10.1037/xhp0000399.
[37] P. Neri, Semantic control of feature extraction from natural scenes, J. Neurosci. 34 (2014) 2374–2388, https://doi.org/10.1523/JNEUROSCI.1755-13.2014.
[38] R.K. Yin, Looking at upside-down faces, J. Exp. Psychol. 81 (1969) 141–145, https://doi.org/10.1037/h0027474.
[39] J.R. Brockmole, J.M. Henderson, Using real-world scenes as contextual cues for search, Vis. Cogn. 13 (2006) 99–108, https://doi.org/10.1080/13506280500165188.
[40] K. Koehler, M.P. Eckstein, Scene inversion slows the rejection of false positives through saccade exploration during search, Proc. 37th Annu. Meet. Cogn. Sci. Soc. (2016) 1141–1146.
[41] D.B. Walther, E. Caddigan, L. Fei-Fei, D.M. Beck, Natural scene categories revealed in distributed patterns of activity in the human brain, J. Neurosci. 29 (2009) 10573–10581, https://doi.org/10.1523/JNEUROSCI.0559-09.2009.
[42] T.A. Kelley, M.M. Chun, K.-P. Chua, Effects of scene inversion on change detection of targets matched for visual salience, J. Vis. 3 (2003) 1–5, https://doi.org/10.1167/3.1.1.
[43] D.I. Shore, R.M. Klein, The effects of scene inversion on change blindness, J. Gen. Psychol. 127 (2000) 27–43, https://doi.org/10.1080/00221300009598569.
[44] R.A. Epstein, J.S. Higgins, W. Parker, G.K. Aguirre, S. Cooperman, Cortical correlates of face and scene inversion: a comparison, Neuropsychologia 44 (2006) 1145–1158, https://doi.org/10.1016/j.neuropsychologia.2005.10.009.
[45] G.A. Rousselet, M.J. Macé, M. Fabre-Thorpe, Is it an animal? Is it a human face? Fast processing in upright and inverted natural scenes, J. Vis. (2003) 440–455, https://doi.org/10.1167/3.6.5.
[46] R. Guyonneau, H. Kirchner, S.J. Thorpe, Animals roll around the clock: the rotation invariance of ultrarapid visual processing, J. Vis. 6 (2006) 1008–1017, https://doi.org/10.1167/6.10.1.
[47] J.W. Rieger, N. Köchy, F. Schalk, M. Grüschow, H.-J. Heinze, Speed limits: orientation and semantic context interactions constrain natural scene discrimination dynamics, J. Exp. Psychol. Hum. Percept. Perform. 34 (2008) 56–76, https://doi.org/10.1037/0096-1523.34.1.56.
[48] Q.C. Vuong, A.F. Hof, H.H. Bülthoff, I.M. Thornton, An advantage for detecting dynamic targets in natural scenes, J. Vis. 6 (2006) 87–96, https://doi.org/10.1167/6.1.2.
[49] J. Dickerson, G.W. Humphreys, On the identification of misoriented objects: Effects of task and level of stimulus description, Eur. J. Cogn. Psychol. 11 (1999) 145–166, https://doi.org/10.1080/713752310.
[50] B.C. Russell, A. Torralba, K.P. Murphy, W.T. Freeman, LabelMe: A database and web-based tool for image annotation, Int. J. Comput. Vis. 77 (2008) 157–173, https://doi.org/10.1007/s11263-007-0090-8.
[51] T.F. Brady, T. Konkle, G.A. Alvarez, A. Oliva, Visual long-term memory has a massive storage capacity for object details, Proc. Natl. Acad. Sci. U. S. A. 105 (2008) 14325–14329, https://doi.org/10.1073/pnas.0803390105.
[52] T.F. Brady, T. Konkle, A. Oliva, G.A. Alvarez, Detecting changes in real-world objects, Commun. Integr. Biol. 2 (2009) 1–3, https://doi.org/10.4161/cib.2.1.7297.
[53] T. Konkle, T.F. Brady, G.A. Alvarez, A. Oliva, Scene memory is more detailed than you think: the role of categories in visual long-term memory, Psychol. Sci. 21 (2010) 1551–1556, https://doi.org/10.1177/0956797610385359.
[54] T. Konkle, T.F. Brady, G.A. Alvarez, A. Oliva, Conceptual distinctiveness supports detailed visual long-term memory for real-world objects, J. Exp. Psychol. Gen. 139 (2010) 558–578, https://doi.org/10.1037/a0019165.
[55] A.D. Wilson, J. Tresilian, F. Schlaghecken, The masked priming toolbox: an open-source MATLAB toolbox for masked priming researchers, Behav. Res. Methods 43 (2011) 210–214, https://doi.org/10.3758/s13428-010-0034-z.
[56] D.H. Brainard, The psychophysics toolbox, Spat. Vis. 10 (1997) 433–436, https://doi.org/10.1163/156856897X00357.
[57] D.G. Pelli, The VideoToolbox software for visual psychophysics: transforming numbers into movies, Spat. Vis. 10 (1997) 437–442, https://doi.org/10.1163/156856897X00366.
[58] T.-P. Jung, S. Makeig, C. Humphries, T.-W. Lee, M.J. Mckeown, V. Iragui, T.J. Sejnowski, Removing electroencephalographic artifacts by blind source separation, Psychophysiology 37 (2000) 163–178, https://doi.org/10.1111/1469-8986.3720163.
[59] A. Delorme, S. Makeig, EEGLAB: An open source toolbox for analysis of single-trial EEG dynamics including independent component analysis, J. Neurosci. Methods 134 (2004) 9–21, https://doi.org/10.1016/j.jneumeth.2003.10.009.
[60] J. Lopez-Calderon, S.J. Luck, ERPLAB: an open-source toolbox for the analysis of event-related potentials, Front. Hum. Neurosci. 8 (2014) 1–14, https://doi.org/10.3389/fnhum.2014.00213.
[61] D. Bates, M. Mächler, B. Bolker, S. Walker, Fitting linear mixed-effects models using lme4, J. Staistical Softw. 67 (2014), https://doi.org/10.18637/jss.v067.i01.
[62] R Development Core Team, R: A Language and Environment for Statistical Computing, (2012).
[63] R.V. Lenth, Least-squares means: the R package lsmeans, J. Stat. Softw. 69 (2016) 1–33, https://doi.org/10.18637/jss.v069.i01.
[64] D.J. Barr, R. Levy, C. Scheepers, H.J. Tily, Random effects structure for confirmatory hypothesis testing: keep it maximal, J. Mem. Lang. 68 (2013) 255–278,

https://doi.org/10.1016/j.jml.2012.11.001.
[65] D. Bates, R. Kliegl, S. Vasishth, H. Baayen, Parsimonious mixed models, J. Mem. Lang. 27 (2015), http://arxiv.org/abs/1506.04967.
[66] A. Kuznetsova, P.B. Brockhoff, R.H.B. Christensen, lmerTest Package: tests in linear mixed effects models, J. Stat. Softw. 82 (2017) 1–26, https://doi.org/10.18637/jss.v082.i13.
[67] M.R. Greene, A. Oliva, The briefest of glances: the time course of natural scene understanding, Psychol. Sci. 20 (2009) 464–472, https://doi.org/10.1111/j.1467-9280.2009.02316.x.
[68] B. Rossion, Picture-plane inversion leads to qualitative changes of face perception, Acta Psychol. (Amst.) 128 (2008) 274–289, https://doi.org/10.1016/j.actpsy.2008.02.003.
[69] A.B. Sekuler, C.M. Gaspar, J.M. Gold, P.J. Bennett, Inversion leads to quantitative, not qualitative, changes in face processing, Curr. Biol. 14 (2004) 391–396, https://doi.org/10.1016/j.cub.2004.02.028.