Fixation-related Brain Potentials during Semantic Integration of Object–Scene Information

Moreno I. Coco[1,2], Antje Nuthmann[3], and Olaf Dimigen[4]

Abstract

■ In vision science, a particularly controversial topic is whether and how quickly the semantic information about objects is available outside foveal vision. Here, we aimed at contributing to this debate by coregistering eye movements and EEG while participants viewed photographs of indoor scenes that contained a semantically consistent or inconsistent target object. Linear deconvolution modeling was used to analyze the ERPs evoked by scene onset as well as the fixation-related potentials (FRPs) elicited by the fixation on the target object (t) and by the preceding fixation ($t - 1$). Object–scene consistency did not influence the probability of immediate target fixation or the ERP evoked by scene onset, which suggests that object–scene semantics was not accessed immediately. However, during the subsequent scene exploration, inconsistent objects were prioritized over consistent objects in extrafoveal vision (i.e., looked at earlier) and were more effortful to process in foveal vision (i.e., looked at longer). In FRPs, we demonstrate a fixation-related N300/N400 effect, whereby inconsistent objects elicit a larger frontocentral negativity than consistent objects. In line with the behavioral findings, this effect was already seen in FRPs aligned to the pretarget fixation $t - 1$ and persisted throughout fixation t, indicating that the extraction of object semantics can already begin in extrafoveal vision. Taken together, the results emphasize the usefulness of combined EEG/eye movement recordings for understanding the mechanisms of object–scene integration during natural viewing. ■

INTRODUCTION

In our daily activities—for example, when we search for something in a room—our attention is mostly oriented to objects. The time course of object recognition and the role of overt attention in this process are therefore topics of considerable interest in the visual sciences. In the context of real-world scene perception, the question of what constitutes an object is a more complex question than intuition would suggest (e.g., Wolfe, Alvarez, Rosenholtz, Kuzmova, & Sherman, 2011). An object is likely a hierarchical construct (e.g., Feldman, 2003), with both low-level features (e.g., visual saliency) and high-level properties (e.g., semantics) contributing to its identity. Accordingly, when a natural scene is inspected with eye movements, the observer's attentional selection is thought to be based either on objects (e.g., Nuthmann & Henderson, 2010), image features (saliency; Itti, Koch, & Niebur, 1998), or some combination of the two (e.g., Stoll, Thrun, Nuthmann, & Einhäuser, 2015).

An early and uncontroversial finding is that the recognition of objects is mediated by their semantic consistency. For example, an object that the observer would not expect to occur in a particular scene (e.g., a toothbrush in a kitchen) is recognized less accurately (e.g., Fenske, Aminoff, Gronau, & Bar, 2006; Davenport & Potter, 2004; Biederman, 1972) and looked at for longer than an expected object (e.g., Cornelissen & Võ, 2017; Henderson, Weeks, & Hollingworth, 1999; De Graef, Christiaens, & d'Ydewalle, 1990).

What is more controversial, however, is the exact time course along which the meaning of an object is processed and how this semantic processing then influences the overt allocation of visual attention (see Wu, Wick, & Pomplun, 2014, for a review). Two interrelated questions are at the core of this debate: (1) How much time is needed to access the meaning of objects after a scene is displayed, and (2) Can object semantics be extracted before the object is overtly attended, that is, while the object is still outside high-acuity foveal vision (> 1° eccentricity) or even in the periphery (> 5° eccentricity)?

Evidence that the meaning of not-yet-fixated objects can capture overt attention comes from experiments that have used sparse displays of several standalone objects (e.g., Cimminella, Della Sala, & Coco, in press; Nuthmann, de Groot, Huettig, & Olivers, 2019; Belke, Humphreys, Watson, Meyer, & Telling, 2008; Moores, Laiti, & Chelazzi, 2003). For example, across three different experiments, Nuthmann et al. found that the very first saccade in the display was directed more frequently to objects that were semantically related to a target object rather than to unrelated objects.

Whether such findings generalize to objects embedded in real-world scenes is currently an open research

[1]The University of East London, [2]CICPSI, Faculdade de Psicologia, Universidade de Lisboa, [3]Christian-Albrechts-Universität zu Kiel, [4]Humboldt-Universität zu Berlin

Journal of Cognitive Neuroscience 32:4, pp. 571–589
https://doi.org/10.1162/jocn_a_01504

question. The size of the visual span—that is, the area of the visual field from which observers can take in useful information (see Rayner, 2014, for a review)—is large in scene viewing. For object-in-scene search, it corresponded to approximately 8° in each direction from fixation (Nuthmann, 2013). This opens up the possibility that both low- and high-level object properties can be processed outside the fovea. This is clearly the case for low-level visual features: Objects that are highly salient (i.e., visually distinct) are preferentially selected for fixation (e.g., Stoll et al., 2015). If semantic processing also takes place in extrafoveal vision, then objects that are inconsistent with the scene context (which are also thought to be more informative; Antes, 1974) should be fixated earlier in time than consistent ones (Loftus & Mackworth, 1978; Mackworth & Morandi, 1967).

However, results from eye-movement studies on this issue have been mixed. A number of studies have indeed reported evidence for an inconsistent object advantage (e.g., Borges, Fernandes, & Coco, 2019; LaPointe & Milliken, 2016; Bonitz & Gordon, 2008; Underwood, Templeman, Lamming, & Foulsham, 2008; Loftus & Mackworth, 1978). Among these studies, only Loftus and Mackworth (1978) have reported evidence for immediate extrafoveal attentional capture (i.e., within the first fixation) by object–scene semantics. In this study, which used relatively sparse line drawings of scenes, the mean amplitude of the saccade into the critical object was more than 7°, suggesting that viewers could process semantic information based on peripheral information obtained in a single fixation. In contrast, other studies have failed to find any advantage for inconsistent objects in attracting overt attention (e.g., Võ & Henderson, 2009, 2011; Henderson et al., 1999; De Graef et al., 1990). In these experiments, only measures of foveal processing—such as gaze duration—were influenced by object–scene consistency, with longer fixation times on inconsistent than on consistent objects.

Interestingly, a similar controversy exists in the literature on eye guidance in sentence reading. Although some degree of parafoveal processing during reading is uncontroversial, it is less clear whether semantic information is acquired from the parafovea (Andrews & Veldre, 2019, for a review). Most evidence from studies involving readers of English has been negative (e.g., Rayner, Balota, & Pollatsek, 1986), whereas results from reading German (e.g., Hohenstein & Kliegl, 2014) and Chinese (e.g., Yan, Richter, Shu, & Kliegl, 2009) suggest that parafoveal processing can advance up to the level of semantic processing.

The processing of object–scene inconsistencies and its time course have also been investigated in electrophysiological studies (e.g., Mudrik, Lamy, & Deouell, 2010; Ganis & Kutas, 2003). In ERPs, it is commonly found that scene-inconsistent objects elicit a larger negative brain response compared with consistent ones. This long-lasting negative shift typically starts as early as 200–250 msec after stimulus onset (e.g., Draschkow, Heikel, Võ, Fiebach, & Sassenhagen, 2018; Mudrik, Shalgi, Lamy, & Deouell, 2014) and has its maximum at frontocentral scalp sites, in contrast to the centroparietal N400 effect for words (e.g., Kutas & Federmeier, 2011). The effect was found for objects that appeared at a cued location after the scene background was already shown (Ganis & Kutas, 2003), for objects that were photoshopped into the scene (Coco, Araujo, & Petersson, 2017; Mudrik et al., 2010, 2014), and for objects that were part of realistic photographs (Võ & Wolfe, 2013). These ERP effects of object–scene consistency have typically been subdivided into two distinct components: N300 and N400. The earlier part of the negative response, usually referred to as N300, has been taken to reflect the context-dependent difficulty of object identification, whereas the later N400 has been linked to semantic integration processes after the object is identified (e.g., Dyck & Brodeur, 2015). The present study was not designed to differentiate between these two subcomponents, especially considering that their scalp distribution is strongly overlapping or even topographically indistinguishable (Draschkow et al., 2018). Thus, for reasons of simplicity, we will in most cases simply refer to all frontocentral negativities as "N400."

One limiting factor of existing ERP studies is that the data were gathered using steady-fixation paradigms in which the free exploration of the scene through eye movements was not permitted. Instead, the critical object was typically large and/or located relatively close to the center of the screen, and ERPs were time-locked to the onset of the image (e.g., Mudrik et al., 2010). Because of these limitations, it remains unclear whether foveation of the object is a necessary condition for processing object–scene consistencies or whether such processing can at least begin in extrafoveal vision.

In the current study, we used fixation-related potentials (FRPs), that is, EEG waveforms aligned to fixation onset, to shed new light on the controversial findings of the role of foveal versus extrafoveal vision in extracting object semantics, while providing insights into the patterns of brain activity that underlie them (for reviews about FRPs, see Nikolaev, Meghanathan, & van Leeuwen, 2016; Dimigen, Sommer, Hohlfeld, Jacobs, & Kliegl, 2011).

FRPs have been used to investigate the brain-electric correlates of natural reading, as opposed to serial word presentation, helping researchers to provide finer details about the online processing of linguistic features (such as word predictability; Kliegl, Dambacher, Dimigen, Jacobs, & Sommer, 2012; Kretzschmar, Bornkessel-Schlesewsky, & Schlesewsky, 2009) or the dynamics of the perceptual span during reading (e.g., parafovea-on-fovea effects; Niefind & Dimigen, 2016). More recently, the coregistration method has also been applied to investigate active visual search (e.g., Ušćumlić & Blankertz, 2016; Devillez, Guyader, & Guérin-Dugué, 2015; Kaunitz et al., 2014; Brouwer, Reuderink, Vincent, van Gerven, & van Erp, 2013; Kamienkowski, Ison, Quiroga, & Sigman, 2012), object

identification (Rämä & Baccino, 2010), and affective processing in natural scene viewing (Simola, Le Fevre, Torniainen, & Baccino, 2015).

In this study, we simultaneously recorded eye movements and FRPs during the viewing of real-world scenes to distinguish between three alternative hypotheses on object–scene integration that can be derived from the literature: (A) One glance of the scene is sufficient to extract object semantics from extrafoveal vision (e.g., Loftus & Mackworth, 1978), (B) extrafoveal processing of object–scene semantics is possible but takes some time to unfold (e.g., Bonitz & Gordon, 2008; Underwood et al., 2008), and (C) the processing of object semantics requires foveal vision, that is, a direct fixation of the critical object (e.g., Võ & Henderson, 2009; Henderson et al., 1999; De Graef et al., 1990). We note that these possibilities are not mutually exclusive, an issue we elaborate on in the Discussion section.

For the behavioral data, these hypotheses translate as follows: under Hypothesis A, the probability of immediate target fixation should reveal that already the first saccade on the scene goes more often toward inconsistent than consistent objects. Under Hypothesis B, there should be no effect on the first eye movement, but the latency to first fixation on the critical object should be shorter for inconsistent than consistent objects. Under Hypothesis C, only fixation times on the critical object itself should differ as a function of object–scene consistency, with longer gaze durations on inconsistent objects.

For the electrophysiological data analysis, we used a novel regression-based analysis approach (linear deconvolution modeling; Cornelissen, Sassenhagen, & Võ, 2019; Dimigen & Ehinger, 2019; Ehinger & Dimigen, 2019; Kristensen, Rivet, & Guérin-Dugué, 2017; Smith & Kutas, 2015b; Dandekar, Privitera, Carney, & Klein, 2012), which allowed us to control for the confounding influences of overlapping potentials and oculomotor covariates on the neural responses during natural viewing. In the EEG, Hypothesis A can be tested by computing the ERP time-locked to the onset of the scene on the display, following the traditional approach. Given that the critical objects in our study were not placed directly in the center of the screen from which observers started their exploration of the scene, any effect of object–scene congruency in this ERP would suggest that object semantics is rapidly processed in extrafoveal vision, even before the first eye movement is generated, in line with Loftus and Mackworth (1978). Under Hypothesis B, we would not expect to see an effect in the scene-onset ERP. Instead, we should find a negative brain potential (N400) for inconsistent as compared with consistent objects in the FRP aligned to the fixation that precedes the one that first lands on the critical object. Finally, if Hypothesis C is correct, an N400 for inconsistent objects should only arise once the critical object is foveated, that is, in the FRP aligned to the target fixation (fixation t). In contrast, no consistency effects should appear in the scene-onset ERP or in the FRP aligned to the pretarget fixation (fixation $t - 1$). To preview the results, both the eye movement and the EEG data lend support for Hypothesis B.

METHODS

Design and Task Overview

We designed a short-term visual working memory change detection task, illustrated in Figures 1 and 2. During the study phase, participants were exposed to photographs

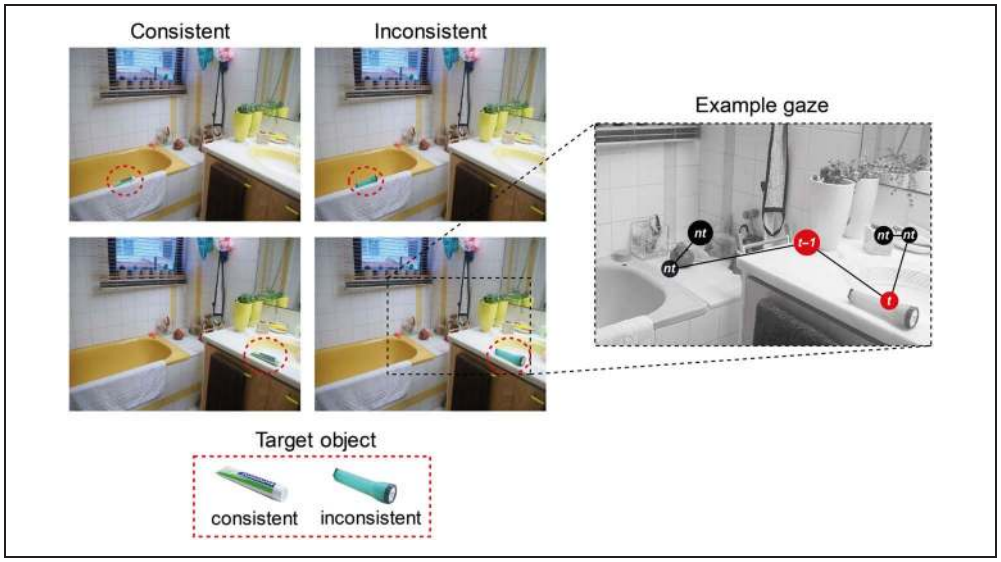

Figure 1. Example stimuli and conditions in the study. Participants viewed photographs of indoor scenes that contained a target object (highlighted with a red circle) that was either semantically consistent (here, toothpaste) or semantically inconsistent (here, flashlight) with the context of the scene. The target object could be placed at different locations within the scene, on either the left or right side. The example gaze path plotted on the right illustrates the three types of fixations analyzed in the study: (a) $t - 1$, the fixation preceding the first fixation to the target object; (b) t, the first fixation to the target; and (c) nt, all other (nontarget) fixations. Fixation duration is proportional to the diameter of the circle, which is red for the critical fixations and black for the nontarget fixations.

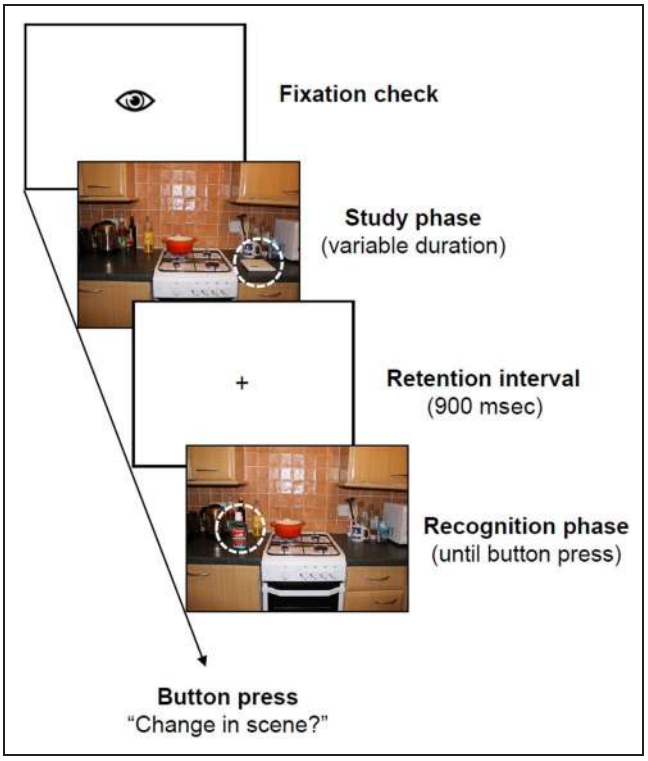

Figure 2. Trial scheme. After a drift correction, the study scene appeared. The display duration of the scene was controlled by a gaze-contingent mechanism, and it disappeared, on average, 2000 msec after the target object was fixated. In the following retention interval, only a fixation cross was presented. During the recognition phase, the scene was presented again until participants pressed a button to indicate whether or not a change had occurred within the scene. All analyses in the present article focus on eye-movement and EEG data collected during the study phase.

of indoor scenes (e.g., a bathroom), each of which contained a target object that was either semantically consistent (e.g., toothpaste) or inconsistent (e.g., a flashlight) with the scene context. In the following recognition phase, after a short retention interval of 900 msec, the same scene was shown again, but in half of the trials, either the identity, the location, or both the identity and location of the target object had changed relative to the study phase.

The participants' task was to indicate with a keyboard press whether or not a change had happened to the scene (see also LaPointe & Milliken, 2016). All eye-movement and EEG analyses in the present article focus on the semantic consistency manipulation of the target object during the study phase.

Participants

Twenty-four participants (nine men) between the ages of 18 and 33 years (M = 25.0 years) took part in the experiment after providing written informed consent. They were compensated with £7 per hour. All participants had normal or corrected-to-normal vision. Data from an additional two participants were recorded but removed from the analysis because of excessive scalp muscle (EMG) activity or skin potentials in the raw EEG. Ethics approval was obtained from the Psychology Research Ethics Committee of the University of Edinburgh.

Apparatus and Recording

Scenes were presented on a 19-in. CRT monitor (Iiyama Vision Master Pro 454) at a vertical refresh rate of 75 Hz. At the viewing distance of 60 cm, each scene subtended 35.8° × 26.9° (width × height). Eye movements were recorded monocularly from the dominant eye using an SR Research EyeLink 1000 desktop-mounted system at a sampling rate of 1000 Hz. Eye dominance for each participant was determined with a parallax test. A chin-and-forehead rest was used to stabilize the participant's head. Nine-point calibrations were run at the beginning of each session and whenever the participant's fixation deviated by > 0.5° horizontally or > 1° vertically from a drift correction point presented at trial onset.

The EEG was recorded from 64 active electrodes at a sampling rate of 512 Hz using BioSemi ActiveTwo amplifiers. Four electrodes, located near the left and right canthus and above and below the right eye, recorded the EOG. All channels were referenced against the BioSemi common mode sense (active electrode) and grounded to a passive electrode. The BioSemi hardware is DC coupled and applies digital low-pass filtering through the A/D-converter's decimation filter, which has a fifth-order sinc response with a −3 dB point at one fifth of the sample rate (corresponding approximately to a 100-Hz low-pass filter).

Offline, the EEG was rereferenced to the average of all scalp electrodes and filtered using EEGLAB's (Delorme & Makeig, 2004) Hamming-windowed sinc finite impulse response filter (pop_eegfiltnew.m) with default settings. The lower edge of the filter's passband was set to 0.2 Hz (with −6 dB attenuation at 0.1 Hz); and the upper edge, to 30 Hz (with −6 dB attenuation at 33.75 Hz). Eye tracking and EEG data were synchronized using shared triggers sent via the parallel port of the stimulus presentation PC to the two recording computers. Synchronization was performed offline using the EYE-EEG extension (v0.8) for EEGLAB (Dimigen et al., 2011). All data sets were aligned with a mean synchronization error ≤ 2 msec as computed based on trigger alignment after synchronization.

Materials and Rating

Stimuli consisted of 192 color photographs of indoor scenes (e.g., bedrooms, bathrooms, offices). Real target objects were placed in the physical scene, before each picture was taken with a tripod under controlled lighting conditions and with a fixed aperture (i.e., there was no photo-editing). One scene is shown in Figure 1; miniature versions of all stimuli used in this study are found

online at https://osf.io/sjprh/. Of the 192 scenes, 96 were conceived as change items and 96 were conceived as no-change items. Each one of the 96 change scenes was created in four versions. In particular, the scene (e.g., a bathroom) was photographed with two alternative target objects in it, one that was consistent with the scene context (e.g., a toothbrush) and one that was not (e.g., a flashlight). Moreover, each of these two objects was placed at two alternative locations (left or right side) within the scene (e.g., either on the sink or on the bathtub). Accordingly, three types of change were implemented during the recognition phase (Congruency, Location, and Both; see Procedure section below).

Each of the 96 no-change scenes was also a real photograph with either a consistent or an inconsistent object in it, which was again located in either the left or right half of the scene. Across the 96 no-change scenes, the factors consistency (consistent vs. inconsistent objects) and location (left and right) were also balanced. However, each no-change scene was unique; that is, we did not create four different versions of each no-change scene. The data of the 96 no-change scenes, which were originally conceived to be filler trials, were included to improve the signal-to-noise ratio of the EEG analyses, as these scenes also had a balanced distribution of inconsistent and consistent objects.

As explained above, scenes contained a critical object that was either consistent or inconsistent with the scene context. Object consistency was assessed in a pretest rating study by eight naive participants who were not involved in any other aspect of the study. Each participant rated all of the no-change scenes as well as one of the four versions of each change-scene (counterbalanced across raters). Together with the scene, raters saw a box with a cropped image of the critical object. They were asked (a) to write down the name for the displayed object and (b) to respond to the question "How likely is it that this object would be found in this room?" using a 6-point Likert scale (1–6). For the object naming, a mean naming agreement of 96.35% was obtained. Furthermore, consistent objects were judged as significantly more likely (M = 5.78, SD = 0.57) to appear in the scene than inconsistent objects (M = 1.88, SD = 1.11), as confirmed by an independent-samples Kruskal–Wallis H test, $\chi^2(1)$ = 616.09, $p < .001$.

In addition, we ensured that there was no difference between consistent and inconsistent objects on three important low-level variables: object size (pixels square), distance from the center of the scene (degrees of visual angle), and mean visual saliency of the object as computed using the Adaptive Whitening Saliency model (Garcia-Diaz, Fdez-Vidal, Pardo, & Dosil, 2012). Table 1 provides additional information about the target object. Independent t tests showed no significant difference between inconsistent and consistent objects in size, $t(476) = -1.27$, p = .2; visual saliency, $t(476) = 0.82$, $p = .41$; and distance from the center, $t(476) = -1.75$, $p = .08$.

The position of each target object was marked with an invisible rectangular bounding box, which was used to implement the gaze contingency mechanism (described in the Procedure section below) and to determine whether a fixation was inside the target object. The average width of the bounding box was 6.1° ± 2.0° for consistent objects and 6.1° ± 2.1° for inconsistent objects (see Table 1); the average height was 5.1° ± 1.8° and/or 5.4° ± 2.2°, respectively. The average distance of the object centroid from the center of the scene was 12.1° (± 2.8°) for consistent and 11.7° (± 3.0°) for inconsistent objects.

Procedure

A schematic representation of the task is shown in Figure 2. Each trial started with a drift correction of the eye tracker. Afterward, the study scene was presented (e.g., a bathroom). The display duration of the study scene was controlled by a gaze-contingent mechanism that ensured that participants fixated the target object (e.g., toothbrush or flashlight) at least once during the trial. Specifically, the study scene disappeared, on average, 2000 msec (with a random jitter of ± 200 msec, drawn from a uniform distribution) after the participant's eyes left the invisible bounding box of the target object (and provided that the target had been fixated for at least 150 msec). The jittered delay of about 2000 msec was implemented to prevent participants from learning to associate the last fixated object during the study phase with the changed object during the recognition phase. If the participant did not fixate the target object within 10 sec, the study scene disappeared from the screen and the retention interval was triggered, which lasted for 900 msec.

In the following recognition phase (data not analyzed here), the scene was presented again, either with (50% of trials) or without (50% of trials) a change to an object in the scene. Three types of object changes occurred with equal probability: Location, Consistency, or Both. In the (a) Location condition, the target object changed its position and moved either from left to right or from right to left to another plausible location within the scene (e.g., a toothbrush was placed elsewhere within the bathroom scene). In the (b) Consistency condition, the object remained in the same location but was replaced with another object of opposite semantic consistency (e.g., the toothbrush was replaced by a flashlight). Finally, in the (c) Both condition, the object was both replaced and moved within the scene (e.g., a toothbrush was replaced by a flashlight at a different location).

During the recognition phase, participants had to indicate whether they noticed any kind of change within the scene by pressing the arrow keys on the keyboard. Afterward, the scene disappeared, and the next trial began. If participants did not respond within 10 sec, a missing response was recorded.

The type of change between trials was fully counterbalanced using a Latin Square rotation. Specifically, the 96

Table 1. Eye Movement Behavior in the Task and Properties of the Target Object

		Consistent	*Inconsistent*
		Mean ± SD	*Mean ± SD*
Eye movement behavior	Ordinal fixation number of first target fixation	6.7 ± 6.0	5.2 ± 5.3
	Fixation duration ($t - 2$), in msec	220.7 ± 105	212.9 ± 95
	Fixation duration ($t - 1$), in msec	207.6 ± 96	197 ± 91
	Fixation duration (t), in msec	261.6 ± 146	263.3 ± 136
	Gaze duration on target, in msec	408.5 ± 367.1	519.1 ± 373.6
	Number of refixations on target	1.7 ± 2	2.2 ± 2.1
	Duration of refixations on target, in msec	238.9 ± 121.8	250.2 ± 135.7
	Fixation duration ($t + 1$), in msec	245.3 ± 148	243.7 ± 146
	Incoming saccade amplitude to $t - 1$ (°)	6.1 ± 5.2	6 ± 4.8
	Incoming saccade amplitude to t (°)	8.5 ± 5.2	8.3 ± 4.8
	Incoming saccade amplitude to $t + 1$ (°)	9.5 ± 5.9	10.2 ± 5.8
	Distance of fixation $t - 1$ from the closest edge of target (°)	6.8 ± 5.8	6.3 ± 5.3
	Number of fixations after first encountering target object until end of study phase	7.3 ± 2.1	7.3 ± 1.7
	Duration of fixations after first encountering target object (until end of study phase)	254.6 ± 120.4	251.7 ± 118.8
Target object properties	Distance of target object center from screen center (°)	12.1 ± 2.8	11.7 ± 3
	Mean visual saliency (AWS model)	0.36 ± 0.16	0.37 ± 0.16
	Width (°)	6.1 ± 2	6.1 ± 2.1
	Height (°)	5.1 ± 1.8	5.4 ± 2.2
	Area (degrees of visual angle squared)	16.1 ± 8.7	17.3 ± 11.4

Target object size and distance to target are based on the bounding box around the object. The fixation $t + 1$ is the first fixation after leaving the bounding box of the target object.

change trials were distributed across 12 different lists, implementing the different types of change. This implies that each participant was exposed to an equal number of consistent and inconsistent change trials. The 96 no-change trials also were composed of an equal number of consistent and inconsistent scenes and were the same for each participant. During the experiment, all 192 trials were presented in a randomized order. They were preceded by four practice trials at the start of the session. Written instructions were given to explain the task, which took 20–40 min to complete. The experiment was implemented using the SR Research Experiment Builder software.

Data Preprocessing

Eye-movement Events and Data Exclusion

Fixations and saccade events were extracted from the raw gaze data using the SR Research Data Viewer software, which performs saccade detection based on velocity and acceleration thresholds of 30° sec^{-1} and 9500° sec^{-2}, respectively. To provide directly comparable results for eye-movement behavior and FRP analyses, we discarded all trials on which we did not have clean data from both recordings. Specifically, from 4608 trials (24 participants × 192 trials), we excluded 10 trials (0.2%) because of machine error (i.e., no data were recorded for those trials), 689 trials (15.0%) because the participant responded incorrectly after the recognition phase, and 494 trials (10.7%) because the target object was not fixated during the study phase. Finally, we removed an additional 97 trials (2.1%) for which the target fixation overlapped with intervals of the EEG that contained nonocular artifacts (see below). The final data set for the behavioral and FRP analyses therefore was composed of 3318 unique trials: 1567 for the consistent condition and 1751 for the inconsistent condition. Per participant, this corresponded to an average of 65.3 trials (± 6.9, range = 48–78) for consistent and 73.0 trials (± 6.9, range = 59–82) for inconsistent

items. Because of the fixation check, participants were always fixating at the screen center when the scene appeared on the display. This ongoing central fixation was removed from all analyses.

EEG Ocular Artifact Correction

EEG recordings during free viewing are contaminated by three types of ocular artifacts (Plöchl, Ossandón, & König, 2012) that need to be removed to get at the genuine brain activity. Here, we applied an optimized variant (Dimigen, 2020) of independent component analysis (ICA; Jung et al., 1998), which uses the information provided by the eye tracker to objectively identify ocular ICA components (Plöchl et al., 2012).

In a first step, we created optimized ICA training data by high-pass filtering a copy of the EEG at 2 Hz (Dimigen, 2020; Winkler, Debener, Müller, & Tangermann, 2015) and segmenting it into epochs lasting from scene onset until 3 sec thereafter. These high-pass-filtered training data were entered into an extended Infomax ICA using EEGLAB, and the resulting unmixing weights were then transferred to the original (i.e., less strictly filtered) recording (Debener, Thorne, Schneider, & Viola, 2010). From this original EEG data set, we then removed all independent components whose time course varied more strongly during saccade intervals (defined as lasting from −20 msec before saccade onset until 20 msec after saccade offset) than during fixations, with the threshold for the variance ratio (saccade/fixation; see Plöchl et al., 2012) set to 1.3. Finally, the artifact-corrected continuous EEG was back-projected to the sensor space. For a validation of the ICA procedure, please refer to Supplementary Figure S1.

In a next step, intervals with residual nonocular artifacts (e.g., EMG bursts) were detected by shifting a 2000-msec moving window in steps of 100 msec across the continuous recording. Whenever the voltages within the window exceeded a peak-to-peak threshold of 100 μV in at least one of the channels, all data within the window were marked as "bad" and subsequently excluded from analysis. Within the linear deconvolution framework (see below), this can easily be done by setting all predictors to zero during these bad EEG intervals (Smith & Kutas, 2015b), meaning that the data in these intervals will not affect the computation.

Analysis

Eye-movement Data

Dependent measures. Behavioral analyses focused on four eye-movement measures commonly reported in the semantic consistency literature: (a) the cumulative probability of having fixated the target object as a function of the ordinal fixation number, (b) the probability of immediate object fixation, (c) the latency to first fixation on the target object, and (d) the gaze duration on the target object (cf. Võ & Henderson, 2009).

Linear mixed-effects modeling. Eye-movement data were analyzed using linear mixed-effects models (LMMs) and generalized LMMs (GLMM) as implemented in the lme4 package in R (Bates, Mächler, Bolker, & Walker, 2015). The only exception was the cumulative probability of first fixations on the target for which a generalized linear model (GLM) was used. One advantage of (G)LMM modeling is that it allows one to simultaneously model the intrinsic variability of both participants and scenes (e.g., Nuthmann & Einhäuser, 2015).

In all analyses, the main predictor was the consistency of the critical object (contrast coding: consistent = −0.5, inconsistent = 0.5) in the study scene. In the (G)LMMs, Participant (24) and Scene (192) were included as random intercepts.[1] The cumulative probability of object fixation was analyzed using a GLM with a binomial (probit) link. This model included the Ordinal Number of Fixation on the scene as a predictor; it was entered as a continuous variable ranging from 1 to a maximum of 28 (the 99th quantile).

In the tables of results, we report the beta coefficients, t values (LMM), z values (GLMM), and p values for each model. For LMMs, the level of significance was calculated from an F test based on the Satterthwaite approximation to the effective degrees of freedom (Satterthwaite, 1946), whereas p values in GLMMs are based on asymptotic Wald tests.

Electrophysiological Data

Linear deconvolution modeling (first level of analysis). EEG measurements during active vision are associated with two major methodological problems: overlapping potentials and low-level signal variability (Dimigen & Ehinger, 2019). Overlapping potentials arise from the rapid pace of active information sampling through eye movements, which causes the neural responses that are evoked by subsequent fixations on the stimulus to overlap with each other. Because the average fixation duration usually varies between conditions, this changing overlap can easily confound the measured waveforms. A related issue is the mutual overlap between the ERP elicited by the initial presentation of the stimulus and the FRPs evoked by the subsequent fixations on it. This second type of overlap is especially important in experiments like ours, in which the critical fixations occurred at different latencies after scene onset in the two experimental conditions.

The problem of signal variability refers to the fact that low-level visual and oculomotor variables can also influence the morphology of the predominantly visually evoked fixation-related neural responses (e.g., Kristensen et al., 2017; Nikolaev et al., 2016; Dimigen et al., 2011). The most relevant of these variables, which is known to modulate the entire FRP waveform, is the amplitude of the saccade

that precedes fixation onset (e.g., Dandekar et al., 2012; Thickbroom, Knezevič, Carroll, & Mastaglia, 1991). One option for controlling the effect of saccade amplitude is to include it as a continuous covariate in a massive univariate regression model (Smith & Kutas, 2015a, 2015b), in which a separate regression model is computed for each EEG time point and channel (Weiss, Knakker, & Vidnyánszky, 2016; Hauk, Davis, Ford, Pulvermüller, & Marslen-Wilson, 2006). However, this method does not account for overlapping potentials.

An approach that allows one to simultaneously control for overlapping potentials and low-level covariates is deconvolution within the linear model (for tutorial reviews, see Dimigen & Ehinger, 2019; Smith & Kutas, 2015a, 2015b), sometimes also called "continuous-time regression" (Smith & Kutas, 2015b). Initially developed to separate overlapping BOLD responses (e.g., Serences, 2004), linear deconvolution has also been applied to separate overlapping potentials in ERP (Smith & Kutas, 2015b) and FRP (Cornelissen et al., 2019; Ehinger & Dimigen, 2019; Kristensen et al., 2017; Dandekar et al., 2012) paradigms. Another elegant property of this approach is that the ERPs elicited by scene onset and the FRPs elicited by fixations on the scene can be disentangled and simultaneously estimated in the same regression model. The benefits of deconvolution are illustrated in more detail in Supplementary Figures S2 and S3.

Here, we applied this technique by using the new *unfold* toolbox (Ehinger & Dimigen, 2019), which represents the first-level analysis and provides us with the partial effects (i.e., the beta coefficients or "regression ERPs"; Smith & Kutas, 2015a, 2015b) for each predictor of interest. In a first step, both stimulus onset events and fixation onset events were included as stick functions (also called "finite impulse responses") in the design matrix of the regression model. To account for overlapping activity from adjacent experimental events, the design matrix was then time-expanded in a time window between −300 and +800 msec around each stimulus and fixation onset event. Time expansion means that the time points within this window are added as predictors to the regression model. Because the temporal distance between subsequent events in the experiment is variable, it is possible to disentangle their overlapping responses. Time expansion with stick functions is explained in Serences (2004) and Ehinger and Dimigen (2019; see their Figure 2). The model was run on EEG data sampled at the original 512 Hz; that is, no down-sampling was performed.

Using Wilkinson notation, the model formula for scene onset events was defined as

$$\text{ERP} \sim 1 + \text{Consistency}$$

In this formula, the beta coefficients for the intercept (1) capture the shape of the overall waveform of the stimulus ERP in the consistent condition, which was used as the reference level, whereas those for Consistency capture the differential effect of presenting an inconsistent object in the scene (relative to a consistent object) on the ERP. The coefficients for the predictor Consistency are therefore analogous to a difference waveform in a traditional ERP analysis (Smith & Kutas, 2015a, 2015b) and would reveal if semantic processing already occurs immediately after the initial presentation of the scene.

In the same regression model, we also included the onsets of all fixations made on the scene. Fixation onsets were modeled with the formula

$$\text{FRP} \sim 1 + \text{Consistency} * \text{Type} + \text{Sacc\_Amplitude}$$

Thus, we predicted the FRP for each time point as a function of the semantic Consistency of the target object (consistent vs. inconsistent; consistent as the reference level) in interaction with the Type of fixation (critical fixation vs. nontarget fixation; nontarget fixation as the reference level). In this model, any FRP consistency effects elicited by the pretarget or target fixation would appear as an interaction between Consistency and Fixation Type. In addition, we included the incoming Saccade Amplitude (in degrees of visual angle) as a continuous linear covariate to control for the effect of saccade size on the FRP waveform.[2] Thus, the full model was as follows:

$$\{\text{ERP} \sim 1 + \text{Consistency},\ \text{FRP} \sim 1 + \text{Consistency} * \text{Type} + \text{Sacc\_Amplitude}\}$$

This regression model was then solved for the betas using the LSMR algorithm in MATLAB (without regularization).

The deconvolution model specified by the formula above was run twice: In one version, we treated the pretarget fixation ($t - 1$) as the critical fixation; in the other version, we treated the target fixation (t) as the critical fixation. In a given model, all fixations but the critical ones were defined as nontarget fixations. FRPs for fixation $t - 1$ and for fixation t were estimated in two separate runs of the model, rather than simultaneously within the same model, because the estimation of overlapping activity was much more stable in this case. In other words, although the deconvolution method allowed us to control for much of the overlapping brain activity from other fixations, we were not able to use the model to directly separate the (two) N400 consistency effects elicited by the fixations $t - 1$ and t.[3]

Both runs of the model (the one for $t - 1$ and t) also yield an estimate for the scene-onset ERP, but because the results for the scene-onset ERP were virtually identical, we present the betas from the first run of the model.

The average number of events entering the model per participant was 65.3 and 73.0 for scene onsets (consistent and inconsistent conditions, respectively), 883.5 and 912.4 for nontarget fixations (*nt*), 59.8 and 61.8 for pretarget fixations ($t - 1$), and 65.3 and 73.0 for target fixations (t).

Baseline placement for FRPs. Another challenging issue for free-viewing EEG experiments is the choice of an appropriate neutral baseline interval for the FRP waveforms (Nikolaev et al., 2016). Baseline placement is particularly relevant for experiments on extrafoveal processing where we do not know in advance when EEG differences will arise and whether they may already develop before fixation onset.

For the pretarget fixation $t - 1$ and nontarget fixations nt, we used a standard baseline interval by subtracting the mean channel voltages between −200 and 0 msec before the event (note that the saccadic spike potential ramping up at the end of this interval was almost completely removed by our ICA procedure; see Supplementary Figure S1). For fixation t, we cannot use such a baseline because semantic processing may already be ongoing by the time the target object is fixated. Thus, to apply a neutral baseline to fixation t, we subtracted the mean channel voltages in the 200-msec interval before the preceding fixation $t - 1$ also from the FRP aligned to the target fixations t (see Nikolaev et al., 2016, for similar procedures). The scene-onset ERP was corrected with a standard prestimulus baseline (−200 to 0 msec).

Group statistics for EEG (second level of analysis). To perform second-level group statistics, averaged EEG waveforms at the single-participant level ("regression ERPs") were reconstructed from the beta coefficients of the linear deconvolution model. These regression-based ERPs are directly analogous to participant-level averages in a traditional ERP analysis (Smith & Kutas, 2015a). We then used two complementary statistical approaches to examine consistency effect in the EEG at the group level: linear mixed models and a cluster-based permutation test.

LMM in a priori defined time windows. LMMs were used to provide hypothesis-based testing motivated by existing literature. Specifically, we adopted the spatio-temporal definitions by Võ and Wolfe (2013) and compared the consistent and inconsistent conditions in the time windows from 250 to 350 msec (early effect) and 350 to 600 msec (late effect) at a midcentral ROI of nine electrodes (comprising FC1, FCz, FC2, C1, Cz, C2, CP1, CPz, and CP2). Because the outputs provided by the linear deconvolution model (the first-level analysis) are already aggregated at the level of participant averages, the only predictor included in these LMMs was the Consistency of the object. Furthermore, to minimize the risk of Type I error (Barr, Levy, Scheepers, & Tily, 2013), we started with a random effects structure with Participant as random intercept and slope for the Consistency predictor. This random effects structure was then evaluated and backwards-reduced using the step function of the lmerTest package (Kuznetsova, Brockhoff, & Christensen, 2017) to retain the model that was justified by the data; that is, it converged, and it was parsimonious in the number of parameters (Matuschek, Kliegl, Vasishth, Baayen, & Bates, 2017).

Cluster permutation tests. It is still largely unknown to what extent the topography of traditional ERP effects translates to natural viewing. Therefore, to test for consistency effects across all channels and time points, we additionally applied the Threshold-Free Cluster Enhancement (TFCE) procedure developed by Smith and Nichols (2009) and adapted to EEG data by Mensen and Khatami (2013; http://github.com/Mensen/ept_TFCE-matlab). In a nutshell, TFCE is a nonparametric permutation test that controls for multiple comparisons across time and space, while maintaining relatively high sensitivity (e.g., compared with a Bonferroni correction). Its advantage over previous cluster permutation tests (e.g., Maris & Oostenveld, 2007) is that it does not require the experimenter to set an arbitrary cluster-forming threshold. In the first stage of the TFCE procedure, a raw statistical measure (here, t values) is weighted according to the support provided by clusters of similar values at surrounding electrodes and time points. In the second stage, these cluster-enhanced t values are then compared with the maximum cluster-enhanced values observed under the null hypotheses (based on $n = 2000$ random permutations of the data). In the present article (Figures 4 and 5), we not only report the global result of the test but also plot the spatio-temporal extent of the first-stage clusters, because they provide some indication about which time points and electrodes likely contributed to the overall significant effect established by the test.

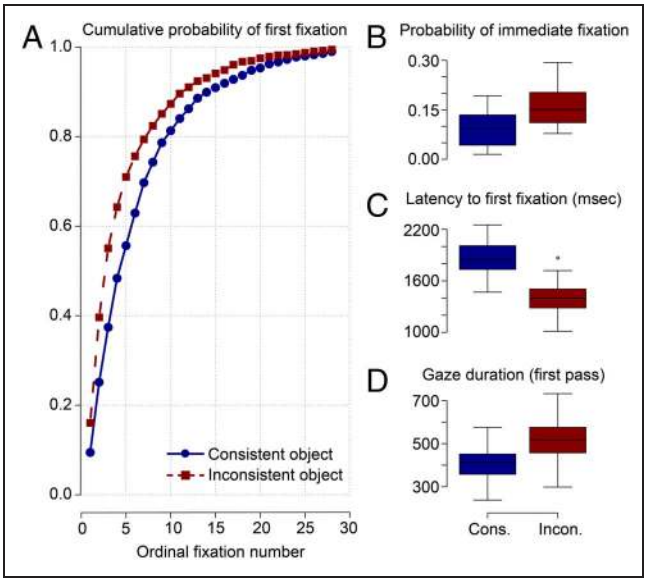

Figure 3. Eye-movement correlates of early overt attention toward consistent and inconsistent critical objects. (A) Cumulative probability of fixating the critical object as a function of the ordinal fixation number on the scene. Blue solid line = consistent object; red dashed line = inconsistent object. (B) Probability of fixating the critical object immediately, that is, with the first fixation after scene onset. (C) Latency until fixating the critical object for the first time. (D) First-pass gaze duration for the critical object, that is, the sum of all fixation durations from first entry to first exit. The size of the boxplots (B–D) represent the 25th and 75th percentiles of the measure (lower and upper quartiles). Dots indicate observations lying beyond the extremes of the whiskers. Cons. = consistent; Incon. = inconsistent.

Table 2. Cumulative Probability of Having Fixated the Critical Object as a Function of the Ordinal Number of Fixations on the Scene (Binomial Probit)

	Cumulative Probability of First Fixation			
Predictor	*β*	*SE*	*z Value*	*Pr (>|z|)*
Intercept	−1.04	0.02	−50.2	.00001
Nr. Fixation	−2.02	0.06	−35.5	.00001
Consistency	0.17	0.03	5.9	.00001
Consistency × Nr. Fixation	−0.72	0.09	−8.1	.00001

The centered predictors are Consistency (consistent: −0.5, inconsistent: 0.5) and Number of Fixation (Nr. Fixation).

Please note, however, that unlike the global test result, these first-stage values are not stringently controlled for false positives and do not establish precise effect onsets or offsets (Sassenhagen & Draschkow, 2019). We report them here as a descriptive statistic.

Finally, for purely descriptive purposes and to provide a priori information for future studies, we also plot the 95% between-participant confidence interval for the consistency effects at the central ROI (corresponding to sample-by-sample paired *t* testing without correction for multiple comparisons; see also Mudrik et al., 2014) in Figures 4 and 5.

RESULTS

Task Performance (Change Detection Task)

After the recognition phase, participants pressed a button to indicate whether or not a change had taken place within the scene. Response accuracy in this task was high (M = 85.0 ± 5.16%) and did not differ as a function of whether the study scene contained a consistent (84.6 ± 5.28%) or an inconsistent (85.3 ± 5.12%) target object.

Eye-movement Behavior

Figure 3A shows the cumulative probability of having fixated the target object as a function of the ordinal number of fixation and semantic consistency, and Table 2 reports the corresponding GLM coefficients. We found a significant main effect of Consistency; overall, inconsistent objects were looked at with a higher probability than consistent objects. As expected, the cumulative probability of looking at the critical object increased as a function of the Ordinal Number of Fixation. There was also a significant interaction between the two variables.

Complementing this global analysis, we analyzed the very first eye movement during scene exploration to assess whether observers had immediate extrafoveal access to object–scene semantics (Loftus & Mackworth, 1978). The mean probability of immediate object fixation was 12.93%; we observed a numeric advantage of inconsistent objects over consistent objects (Figure 3B), but this difference was not significant (Table 3). The latency to first fixation on the target object is another measure to capture the potency of an object in attracting early attention in extrafoveal vision (e.g., Võ & Henderson, 2009; Underwood & Foulsham, 2006). This measure is defined as the time elapsed between the onset of the scene image and the first fixation on the critical object. Importantly, this latency was significantly shorter for inconsistent as compared with consistent objects (Figure 3C, Table 3).

Moreover, we analyzed gaze duration as a measure of foveal object processing time (e.g., Henderson et al., 1999). First-pass gaze duration for a critical object is defined as the sum of all fixation durations from first entry to first exit. On average, participants looked longer at inconsistent (519 msec) than consistent (409 msec) objects

Table 3. Probability of Immediate Fixation, Latency to First Fixation, and Gaze Duration

	Probability of Immediate Fixation			*Latency to First Fixation*			*Gaze Duration*		
Predictor	*β*	*SE*	*z*	*β*	*SE*	*t*	*β*	*SE*	*t*
Intercept	−2.82	0.18	−15.36***	1,774.4	77.2	23.0***	455.5	36.55	23.33***
Consistency	0.22	0.16	1.38	−246.4	64.0	−3.85***	105.0	14.83	7.08***

The simple coded predictor is Consistency (consistent = −0.5, inconsistent = 0.5). We report the *β*, standard error, *z* value (for binomial link), and *t* value. Asterisks indicate significant predictors.

***$p < .001$.

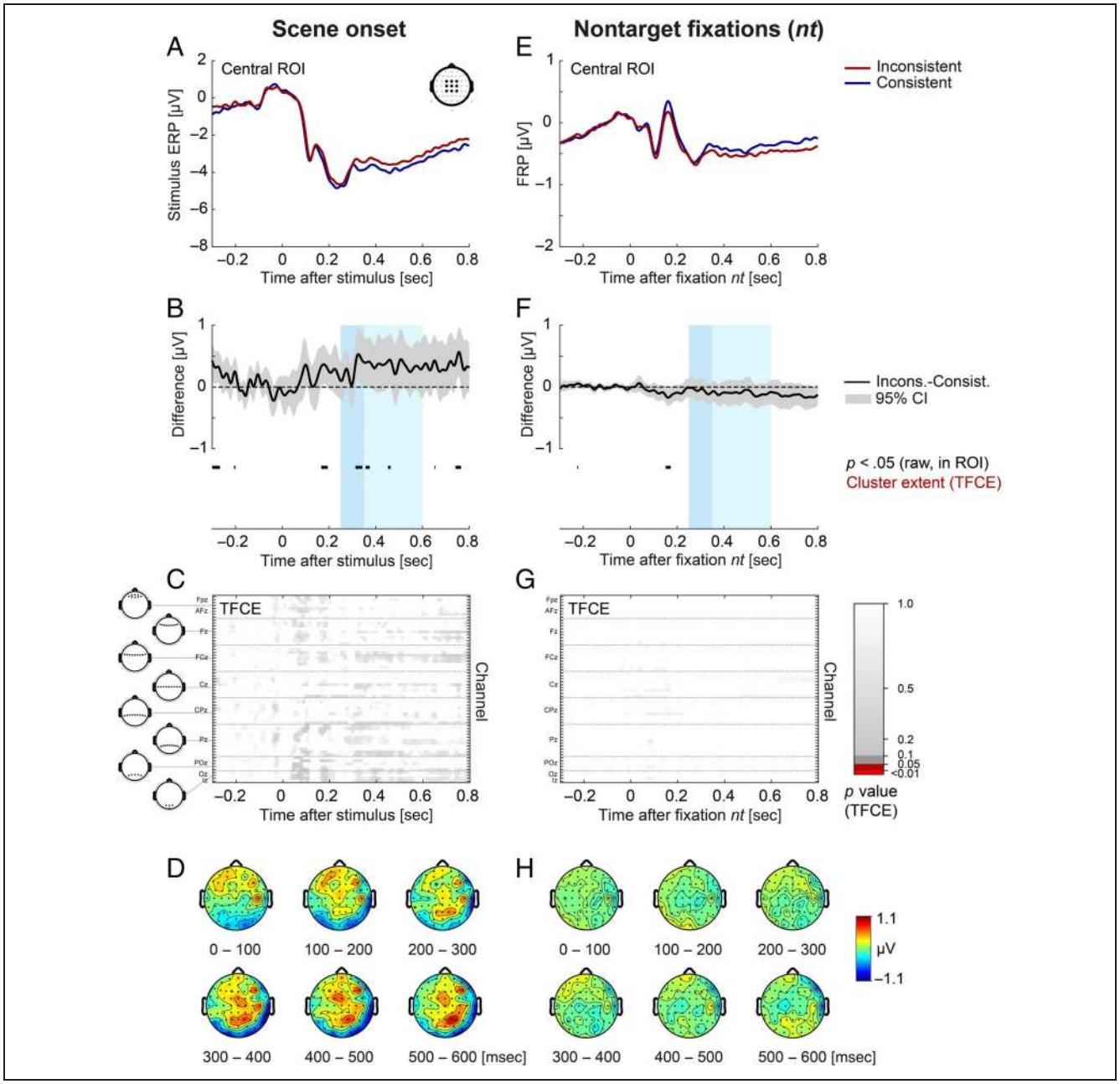

Figure 4. Stimulus ERP aligned to scene onset (left) and FRP aligned to nontarget fixations (right) as a function of object–scene consistency. (A, E) Grand-averaged ERP/FRP at the central ROI (composed of electrodes FC1, FCz, FC2, C1, Cz, C2, CP1, CPz, and CP2). Red lines represent the inconsistent condition, and blue lines represent the consistent condition. (B, F) Corresponding difference waves (inconsistent minus consistent) at the central ROI. Gray shading illustrates the 95% confidence interval (without correction for multiple comparisons) of the difference wave, with values outside the confidence interval also marked in black below the curve. The two windows used for LMM statistics (250–350 and 350–600 msec) are indicated in light blue. (C, G) Extent of the spatio-temporal clusters underlying the cluster-based permutation statistic (TFCE) computed across all electrodes/time points. There were no significant ($p < .05$) effects. (D, H) Scalp topographies of the consistency effect (inconsistent minus consistent) averaged across successive 100-msec time windows. Object–scene consistency had no significant effects on the stimulus ERP or on the FRP elicited by nontarget fixations, neither in the LMM statistic nor in the cluster permutation test. Consist. = consistent; Incons. = inconsistent.

before leaving the target object for the first time, and this difference was significant (Table 3). Table 1 summarizes additional oculomotor characteristics in the two conditions of object consistency.

Supplementary Figures S4 and S5 visualize the locations of the pretarget, target, and posttarget fixations for two example scene stimuli.

Electrophysiological Results

Figures 4 and 5 depict the ERP evoked by the presentation of the scene as well as the FRPs for the three types of fixation that were analyzed. Results focus on the midcentral ROI for which effects of object–scene consistency have been reported. Waveforms for other scalp sites are depicted in Supplementary Figures S6–S9.

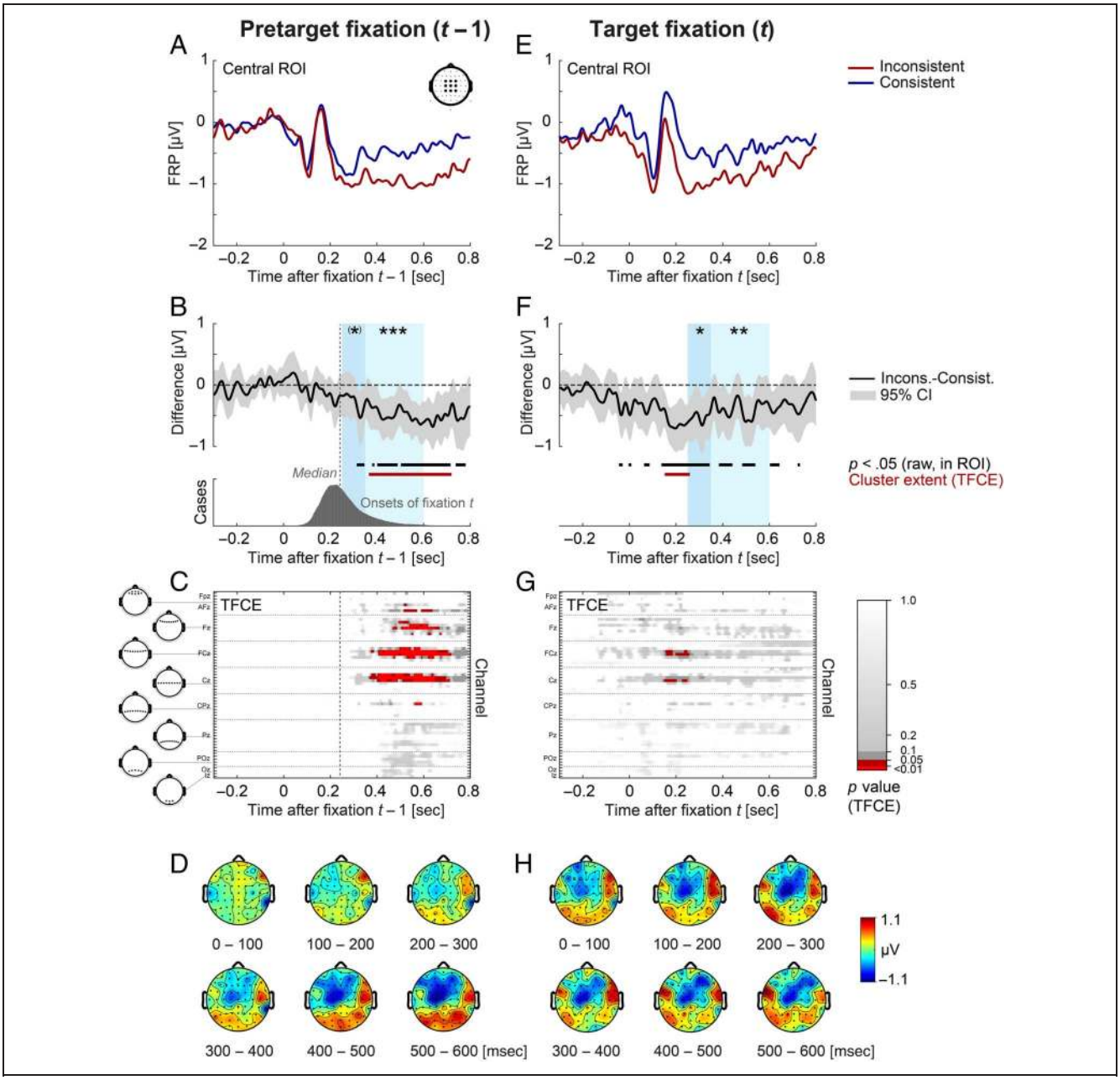

Figure 5. Grand-averaged FRP elicited by pretarget fixation (left) and target fixation (right) as a function of object–scene consistency. (A, E) Grand-averaged FRPs at the central ROI. (B, F) Difference waves at the central ROI. In B, the gray distribution shows the onset of fixation *t* relative to the onset of the pretarget fixation *t* − 1, with the vertical dotted line indicating the median latency (260 msec). (C, G) Results of cluster-based permutation testing (TFCE). The extent of the clusters from the first stage of the permutation test (marked in red) provides some indication which spatio-temporal features of the waveforms likely contributed to the overall significant effect of consistency. The temporal extent of the clusters is also illustrated by the red bars in B and F. (D, H) Scalp topographies of the consistency effect (inconsistent minus consistent) across successive 100-msec time windows. A frontocentral N400 effect emerged in the FRP time-locked to fixation *t* − 1 and reached significance shortly after the eyes had moved on to fixation *t*. This effect then continued during fixation *t* reaching a maximum of 200 msec after the start of the target fixation. Consist. = consistent; Incons. = inconsistent.

Scene-onset ERP

The left panels of Figure 4 show the grand-averaged ERP aligned to scene onset. Although inspection of the scalp maps indicated slightly more positive amplitudes over central right-hemispheric electrodes in the inconsistent condition, these differences were not statistically significant. Specifically, no effect of Consistency was found with the LMM analysis in the early or late time window (see Table 4 for detailed LMM results). Similarly, the TFCE test across all channels and time points yielded no significant Consistency effect (all *p*s > .2; see Figure 4C). Thus, we found no evidence that the semantic consistency of the target object influences the neural response to the initial presentation of the scene.

Table 4. Mixed-Effects Models for the ERPs/FRPs at the Mid-central ROI for Two Temporal Windows of Interest as Predicted by *Consistency*

Type of Event	*Analysis Window*	β	*SE*	*t Value*
Scene onset	Early (250–350 msec)	0.28	0.39	0.71
	Late (350–600 msec)	0.34	0.39	0.37
nt	Early (250–350 msec)	−0.06	0.07	−0.79
	Late (350–600 msec)	−0.09	0.08	−1.10
$t - 1$	Early (250–350 msec)	−0.28	0.15	−1.77(*)
	Late (350–600 msec)	−0.46	0.12	−3.76***
t	Early (250–350 msec)	−0.52	0.17	−3.03**
	Late (350–600 msec)	−0.38	0.15	−2.43*
t (control analysis with baseline before fixation *t*)	Early (250–350 msec)	−0.34	0.16	−2.20*
	Late (350–600 msec)	−0.20	0.17	−1.14

Temporal windows of interest: Early = 250–350 msec; Late = 350–600 msec. Consistency is defined as: Consistent = −0.5, Inconsistent = 0.5.

(*)$p < .1$.

*$p < .05$.

**$p < .01$.

***$p < .001$.

Nontarget Fixations, nt

Next, we tested whether fixations on scenes with an inconsistent object evoke a globally different neural response than those on scenes containing a consistent object. As the right panels of Figure 4 show, this was not the case: Consistency had no effect on the FRP for nontarget (*nt*) fixations, neither in the LMM analysis (see Table 4) nor in the TFCE statistic (all ps $> .2$; see Figure 4G).

Pretarget Fixation, $t - 1$

Figure 5 depicts the FRPs aligned to the pretarget and target fixations. Importantly, in the FRP aligned to the pretarget fixation $t - 1$, waveforms began to clearly diverge between the two consistency conditions, developing into a long-lasting frontocentral negativity in the inconsistent as compared with the consistent condition (Figure 5A and B; see also Supplementary Figure S8). The scalp distribution of this difference, shown in Figure 6, closely resembled the frontocentral N400 (and N300) previously reported in ERP studies on object–scene consistency (e.g., Mudrik et al., 2014; Võ & Wolfe, 2013). In the LMM analyses conducted on the midcentral ROI, this effect was marginally significant ($p < .1$) for the early time window (250–350 msec) but became highly significant between 350 and 600 msec ($p < .001$; Table 4). The TFCE test across all channels and time points also revealed a significant effect of consistency on the pretarget FRP ($p < .05$). Figure 5C also shows the extents of the underlying spatio-temporal clusters, computed in the first stage of the TFCE procedure. Between 372 and 721 msec after fixation onset, we observed a cluster of 14 frontocentral electrodes that was shifted slightly to the left hemisphere. This N400 modulation on the pretarget fixation could be seen even in traditionally averaged FRP waveforms without any control of overlapping potentials (see Supplementary Figure S3). In summary, we were able to measure a significant frontocentral N400 modulation during natural scene viewing that already emerged in FRPs aligned to the pretarget fixation.

On average, the target fixation *t* occurred at a median latency of 240 msec (± 18 msec) after fixation $t - 1$, as marked by the vertical dashed line in Figure 5B. If we take the extent of the cluster from the TFCE test as a rough approximation for the likely onset of the effect in the FRP, this means that, on average, at the time when the electrophysiological consistency effect started (372 msec), the eyes had been looking at the target object for only 132 msec (372 minus 240 msec).

Target Fixation, t

An anterior N400 effect was also clearly visible in the FRP aligned to fixation *t*. In the LMM analysis at the central ROI, the effect was significant in both the early (250–350 msec, $p < .01$) and late (350–600 msec, $p < .05$) windows (see Table 4). However, compared with the effect aligned to the pretarget fixation, this N400 was significant at only a few electrodes in the TFCE statistic (Cz, FCz, and FC1; see Figure 6). Aligned to the target fixation *t*, the N400 also peaked extremely early, with the maximum of the difference curve already observed at 200 msec after fixation onset (Figure 5F). Qualitatively, a frontocentral negativity was already visible much earlier than that, within

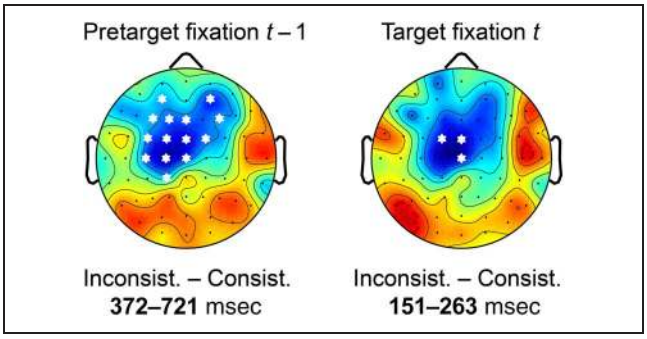

Figure 6. Scalp distribution of frontocentral N400 effects in the time windows significant in the TFCE statistic (see also Figure 5). White asterisks highlight the spatial extent of the clusters observed in the first stage of the TFCE permutation test for both intervals. In the FRP aligned to the pretarget fixation (left), clusters extended from 372 to 721 msec and across 14 frontocentral channels. In the FRP aligned to the target fixation (right), clusters extended from 151 to 263 msec at three frontocentral channels. Consist. = consistent; Inconsist. = inconsistent.

the first 100 msec after fixation onset (Figure 5H). The TFCE permutation test confirmed an overall effect of consistency ($p < .05$) on the target-locked FRP. Figure 5G also shows the extents of the underlying first-stage clusters. For the target fixation, clusters only extended across a brief interval between 151 and 263 msec after fixation onset, an interval during which the N400 effect also reached its peak.

Figure 5F shows that, numerically, voltages at the central ROI were more negative in the inconsistent condition during the baseline interval already, that is, before the critical object was fixated. To understand the role of activity already present before fixation onset, we repeated the FRP analyses for fixation t after applying a standard baseline correction, with the baseline placed immediately before the target fixation itself (−200 to 0 msec). This way, we eliminate any weak N400-like effects that may have already been ongoing before target fixation onset. Interestingly, in the resulting FRP waveforms, the target-locked N400 effects were weakened: The N400 effect now failed to reach significance in the TFCE statistic and in the LMM analysis for the second window (350–600 msec; see the last row of Table 4) and only remained significant for the early window (250–350 msec). This indicates that some N400-like negativity was already ongoing before target fixation onset.

To summarize, we found no immediate influences of object–scene consistency in ERPs time-locked to scene onset. However, N400 consistency effects were found in FRPs aligned to the target fixation (t) and in those aligned to the pretarget fixation ($t - 1$).

DISCUSSION

Substantial research in vision science has been devoted to understanding the behavioral and neural mechanisms underlying object recognition (e.g., Loftus & Mackworth, 1978; Biederman, 1972). At the core of this debate are the type of object features that are accessed (e.g., low-level vs. high-level), the time course of their processing (e.g., preattentive vs. attentive), and the region of the visual field in which these features can be acquired (e.g., foveal vs. extrafoveal). A particularly controversial topic is whether and how quickly semantic properties of objects are available outside foveal vision.

In the current study, we approached these questions from a new perspective by coregistering eye movements and EEG while participants freely inspected images of real-world scenes in which a critical object was either consistent or inconsistent with the scene context. As a novel finding, we demonstrate a fixation-related N400 effect during natural scene viewing. Moreover, behavioral and electrophysiological measures converge to suggest that the extraction of object–scene semantics can already begin in extrafoveal vision, before the critical object is fixated.

It is a rather undisputed finding that inconsistent objects, such as a flashlight in a bathroom, require increased processing when selected as targets of overt attention. Accordingly, several eye-movement studies have reported longer gaze durations on inconsistent than consistent objects, probably reflecting the greater effort required to resolve the conflict between object meaning and scene context (e.g., Cornelissen & Võ, 2017; Henderson et al., 1999; De Graef et al., 1990). In addition, a number of traditional ERP studies using steady-fixation paradigms have found that inconsistent objects elicit a larger negative brain response at frontocentral channels (an N300/N400 complex) as compared with consistent objects (e.g., Coco et al., 2017; Mudrik et al., 2010; Ganis & Kutas, 2003).

However, previous research with eye movements remained inconclusive on whether semantic processing can take place before foveal inspection of the object. Evidence in favor of extrafoveal processing of object–scene semantics comes from studies in which inconsistent objects were selected for fixation earlier than consistent ones (e.g., Borges et al., 2019; LaPointe & Milliken, 2016; Underwood et al., 2008). However, other studies have not found evidence for earlier selection of inconsistent objects (e.g., Võ & Henderson, 2009, 2011; Henderson et al., 1999; De Graef et al., 1990). Parafoveal and peripheral vision are known to be crucial for saccadic programming (e.g., Nuthmann, 2014). Therefore, any demonstration that semantic information can act as a source of guidance for fixation selection in scenes implies that some semantic processing must have occurred prior to foveal fixation, that is, in extrafoveal vision.

ERPs are highly sensitive to semantic processing (Kutas & Federmeier, 2011) and provide an excellent temporal resolution to investigate the time course of object processing. However, an obvious limitation of existing ERP studies is that observers were not allowed to explore the scene with saccadic eye movements, thereby constraining their

normal attentional dynamics. Instead, the critical object was usually large and/or placed near the point of fixation. Hence, these studies were unable to establish whether semantic processing can take place before foveal inspection of the critical object.

In the current study, we addressed this problem by simultaneously recording behavioral and brain-electric correlates of object processing. Specifically, we analyzed different eye-movement responses that tap into extrafoveal and foveal processing along with FRPs time-locked to the first fixation on the critical object (t) and the fixation preceding it ($t - 1$). We also analyzed the scene-onset ERP evoked by the trial-initial presentation of the image. Recent advances in linear deconvolution methods for EEG (e.g., Ehinger & Dimigen, 2019) allowed us to disentangle the overlapping brain potentials produced by the scene onset and the subsequent fixations and to control for the modulating influence of saccade amplitude on the FRP.

The eye-movement behavior showed no evidence for Hypothesis A, as outlined in the Introduction, according to which semantic information can exert an immediate effect on eye-movement control (Loftus & Mackworth, 1978). Specifically, the mean probability of immediate object fixation was fairly low (12.9%) and not modulated by Consistency. Instead, the data lend support to Hypothesis B, according to which extrafoveal processing of object–scene semantics is possible but takes some time to unfold. In particular, the results for the latency to first fixation of the critical object show that inconsistent objects were, on average, looked at sooner than consistent objects (cf. Bonitz & Gordon, 2008; Underwood et al., 2008). At the same time, we observed longer gaze durations on inconsistent objects, replicating previous findings (e.g., Võ & Henderson, 2009; Henderson et al., 1999; De Graef et al., 1990). Thus, we found not only behavioral evidence for the extrafoveal processing of object–scene (in)consistencies but also differences in the subsequent foveal processing.

The question then remains why existing eye-movement studies have provided very different results, ranging from rapid processing of semantic information in peripheral vision to a complete lack of evidence for extrafoveal semantic processing. Researchers have suggested that the outcome may depend on factors related to the critical object or the scene in which it is located. Variables that may (or may not) facilitate the appearance of the incongruency effect include visual saliency (e.g., Underwood & Foulsham, 2006; Henderson et al., 1999), image clutter (Henderson & Ferreira, 2004), and the critical object's size and eccentricity (Gareze & Findlay, 2007). Therefore, an important question for future research is to identify the specific conditions under which extrafoveal semantic information can be extracted or when the three outlined hypotheses and/or outcomes would prevail.

Returning to the present data, the FRP waveforms showed a negative shift over frontal and central scalp sites when participants fixated a scene-inconsistent object. This result is in agreement with traditional ERP studies that have shown a frontocentral N300/N400 complex after passive foveal stimulation (e.g., Coco et al., 2017; Mudrik et al., 2014; Võ & Wolfe, 2013; Ganis & Kutas, 2003) and extends this finding for the first time to a natural viewing situation with eye movements. Regarding the time course, the present data suggest that the effect was already initiated during the preceding fixation ($t - 1$) but then carried on through fixation (t) on the target object.

As a cautionary note, we emphasize that it is not trivial to unambiguously ascribe typical N400 (and N300) effects in the EEG to either extrafoveal or foveal processing. The reason is that these canonical congruency effects only begin 200–250 msec after stimulus onset (Draschkow et al., 2018; Mudrik et al., 2010). This means that even a purely extrafoveal effect would be almost impossible to measure during the pretarget fixation ($t - 1$) itself, because it would only emerge at a time when the eyes are already moving to the target object. That being said, three properties of the observed FRP consistency effect suggest that it was already initiated during the pretarget fixation.

First, because of the temporal jitter introduced by variable fixation durations, an effect that only arises in foveal vision should be the most robust in the FRP averages aligned to fixation t but latency-jittered and attenuated in those aligned to fixation $t - 1$. However, the opposite was the case: At least qualitatively, a frontocentral N400 effect was seen at more electrodes (Figure 6) and for longer time intervals (Figure 5) in the FRP aligned to the pretarget fixation as compared with the actual target fixation. The second argument for extrafoveal contributions to the effect is the forward shift in its time course. Relative to fixation t, the observed N400 occurred almost instantly: As the effect topographies in Figure 5H show, the frontocentral negativity for inconsistent objects was qualitatively visible within the first 100 msec after fixation onset, and the effect reached its peak after only 200 msec. Clusters underlying the TFCE test were also restricted to an early time range between 151 and 263 msec after fixation onset and therefore to a much earlier interval to what we would expect from the canonical N300 or N400 effect elicited by foveal stimulation.

Of course, it is possible that even purely foveal N400 effects may emerge earlier during active scene exploration with eye movements as compared with the latencies established in traditional ERP research. For example, it is reasonable to assume that, during natural vision, observers preprocess some low-level (nonsemantic) features of the soon-to-be fixated object in extrafoveal vision (cf. Nuthmann, 2017). This nonsemantic preview benefit might then speed up the timeline of foveal processing (including the latency of semantic access) once the object is fixated (cf. Dimigen, Kliegl, & Sommer, 2012, for reading). Moreover, if eye movements are permitted, observers have more time to build a representation of the scene before they foveate the target, and this

increased contextual constraint may also affect the N400 timing (but see Kutas & Hillyard, 1984). Importantly, however, neither of these two accounts could explain why the N400 effect is stronger—rather than much weaker—in the waveforms aligned to fixation $t - 1$ as compared with fixation t. The fact that the eye movement data also provided clear evidence in favor of extrafoveal processing further strengthens our interpretation of the N400 timing.

Finally, we found that the N400 consistency effect aligned to the target fixation (t) became weaker (and nonsignificant in two of the three statistical measures considered) if the baseline interval for the FRP analysis was placed directly before this target fixation. Again, this indicates that at least a weak frontocentral negativity in the inconsistent condition was already present during the baseline period before the target was fixated. Together, these results are difficult to reconcile with a pure foveal processing account and are more consistent with the notion that semantic processing of the object was at least initiated in extrafoveal vision (and then continued after it was foveated).

Crucially, we did not find any effect of target consistency in the traditional ERP aligned to scene onset. In line with the behavioral results, this goes against the most extreme Hypothesis A postulating that object semantics can be extracted from peripheral vision already at the first glance of a scene (Loftus & Mackworth, 1978). Similarly, there was no effect of consistency on the FRPs evoked by the nontarget fixations on the scene (Figure 4); this was also the case in a control analysis that only included nontarget fixations that occurred earlier than $t - 1$ and at an extrafoveal distance between 3° and 7° from the target object (see Supplementary Figure S10). All these analyses suggest that the semantic information of the critical object started during fixation $t - 1$. However, from any given fixation, there are many candidate locations that could potentially be chosen for the next saccade (cf. Tatler, Brockmole, & Carpenter, 2017). Thus, it is conceivable that observers may have partially acquired semantic information of the critical object outside foveal vision before fixation $t - 1$, but without selecting it as a saccade target. Such reasoning leaves open the possibility that observers may have already picked up some information about the target object's semantics during these occasions.

Taken together, our behavior and electrophysiological findings are consistent with the claim formulated in Hypothesis B that objects can be recognized outside the fovea or even in the visual periphery, at least to some degree. Indirectly, our results also speak to the debate about the unit of saccade targeting and, by inference, attentional selection during scene viewing. Finding effects of object–scene semantics on eye guidance is evidence in favor of object- and meaning-based, rather than image-based, guidance of attention in scenes (e.g., Henderson, Hayes, Peacock, & Rehrig, 2019; Hwang, Wang, & Pomplun, 2011).

In summary, our findings converge to suggest that the visual system is capable of accessing semantic features of objects in extrafoveal vision to guide attention toward objects that do not fit to the scene's overall meaning. They also highlight the utility of investigating attentional and neural mechanisms in parallel to uncover the mechanisms underlying object recognition during the unconstrained exploration of naturalistic scenes.

Acknowledgments

This research was supported by the Leverhulme Trust (grant ECF-014-205) and Fundação para a Ciência e Tecnologia (grant PTDC/PSI-ESP/30958/2017) to M. I. C., while he was a Research Fellow at the University of Edinburgh. The authors thank Benedikt Ehinger for helpful discussions on EEG deconvolution techniques.

Reprint requests should be sent to Moreno I. Coco, School of Psychology, The University of East London, Water Lane, London E16 2RD, United Kingdom, or Olaf Dimigen, Institut für Psychologie, Humboldt-Universität zu Berlin, Unter den Linden 6, 10099 Berlin, or via e-mail: moreno.cocoi@gmail.com or olaf.dimigen@hu-berlin.de.

Notes

1. We did not include random slopes for two reasons: For Participant, the inclusion of a random slope led to a small variance and a perfect correlation between intercept and slope. For the random effect Scene, only the change trials were fully counterbalanced in terms of location and consistency, meaning that the slope for Consistency could not be estimated for the no-change trials.
2. Other low-level variables, such as local image features in the currently foveated image region (e.g., luminance, spatial frequency), are also known to modulate the FRP waveform. In the model presented here, we did not include these other covariates because (1) their influence on the FRP waveform is small compared with that of saccade amplitude and (2) the properties of the target object (such as its visual saliency) did not differ between the two levels of object consistency (see Materials and Rating section). For reasons of simplicity, saccade amplitude was included as a linear predictor in the current model, although its influence on the FRP becomes nonlinear for large saccades (e.g., Dandekar et al., 2012). However, virtually identical results were obtained when we included it as a nonlinear (spline) predictor instead (Dimigen & Ehinger, 2019).
3. In theory, a more elegant model would include Type as a three-level predictor, with the levels of pretarget, target, and nontarget fixation. In principle, this would allow us to dissociate which parts of the N400 consistency effects are elicited by fixation $t - 1$ versus fixation t. The practical disadvantage of this approach is that the overlapping activities from both $t - 1$ and t would then be estimated on comparatively fewer observations (compared with the extremely stable estimate for the numerous nontarget fixations). This is critical because, compared with the limited amount of jitter in natural fixation durations, N400 effects are a long-lasting response, which makes the deconvolution more challenging. Specifically, we found that, with the three-level model, model outputs became extremely noisy and did not yield significant consistency effects for any EEG time-locking point. By defining either fixation $t - 1$ or fixation t as the critical fixation in two separate runs of the model and by treating all other fixations as nontarget fixations, the estimation becomes very robust. This simpler model still removes most of

the overlapping activity from other fixations. However, the consistency-specific activity evoked by fixation $t - 1$ (i.e., the N400 effect) will not be removed from the FRP aligned to the fixation t and vice versa.

REFERENCES

Andrews, S., & Veldre, A. (2019). What is the most plausible account of the role of parafoveal processing in reading? *Language and Linguistics Compass*, *13*, e12344.

Antes, J. R. (1974). The time course of picture viewing. *Journal of Experimental Psychology*, *103*, 62–70.

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, *68*, 255–278.

Bates, D., Mächler, M., Bolker, B. M., & Walker, S. C. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*, 1–48.

Belke, E., Humphreys, G. W., Watson, D. G., Meyer, A. S., & Telling, A. L. (2008). Top–down effects of semantic knowledge in visual search are modulated by cognitive but not perceptual load. *Perception & Psychophysics*, *70*, 1444–1458.

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

Bonitz, V. S., & Gordon, R. D. (2008). Attention to smoking-related and incongruous objects during scene viewing. *Acta Psychologica*, *129*, 255–263.

Borges, M. T., Fernandes, E. G., & Coco, M. I. (2019). Age-related differences during visual search: The role of contextual expectations and cognitive control mechanisms. *Aging, Neuropsychology, and Cognition*. https://doi.org/10.1080/13825585.2019.1632256.

Brouwer, A.-M., Reuderink, B., Vincent, J., van Gerven, M. A. J., & van Erp, J. B. F. (2013). Distinguishing between target and nontarget fixations in a visual search task using fixation-related potentials. *Journal of Vision*, *13*, 17.

Cimminella, F., Della Sala, S., & Coco, M. I. (in press). Parallel and extra-foveal processing of object semantics during visual search. *Attention, Perception, & Psychophysics*. https://doi.org/10.3758/s13414-019-01906-1.

Coco, M. I., Araujo, S., & Petersson, K. M. (2017). Disentangling stimulus plausibility and contextual congruency: Electrophysiological evidence for differential cognitive dynamics. *Neuropsychologia*, *96*, 150–163.

Cornelissen, T. H. W., Sassenhagen, J., & Võ, M. L.-H. (2019). Improving free-viewing fixation-related EEG potentials with continuous-time regression. *Journal of Neuroscience Methods*, *313*, 77–94.

Cornelissen, T. H. W., & Võ, M. L.-H. (2017). Stuck on semantics: Processing of irrelevant object–scene inconsistencies modulates ongoing gaze behavior. *Attention, Perception, & Psychophysics*, *79*, 154–168.

Dandekar, S., Privitera, C., Carney, T., & Klein, S. A. (2012). Neural saccadic response estimation during natural viewing. *Journal of Neurophysiology*, *107*, 1776–1790.

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

Debener, S., Thorne, J., Schneider, T. R., & Viola, F. C. (2010). Using ICA for the analysis of multi-channel EEG data. In M. Ullsperger & S. Debener (Eds.), *Simultaneous EEG and fMRI: Recording, analysis, and application* (pp. 121–133). New York: Oxford University Press.

De Graef, P., Christiaens, D., & d'Ydewalle, G. (1990). Perceptual effects of scene context on object identification. *Psychological Research*, *52*, 317–329.

Delorme, A., & Makeig, S. (2004). EEGLAB: An open sorce toolbox for analysis of single-trial EEG dynamics including independent component analysis. *Journal of Neuroscience Methods*, *134*, 9–21.

Devillez, H., Guyader, N., & Guérin-Dugué, A. (2015). An eye fixation-related potentials analysis of the P300 potential for fixations onto a target object when exploring natural scenes. *Journal of Vision*, *15*, 20.

Dimigen, O. (2020). Optimizing the ICA-based removal of ocular EEG artifacts from free viewing experiments. *Neuroimage*, *207*, 116117.

Dimigen, O., & Ehinger, B. V. (2019). Analyzing combined eye-tracking/EEG experiments with (non)linear deconvolution models. *BioRxiv*. https://doi.org/10.1101/735530.

Dimigen, O., Kliegl, R., & Sommer, W. (2012). Trans-saccadic parafoveal preview benefits in fluent reading: A study with fixation-related brain potentials. *Neuroimage*, *62*, 381–393.

Dimigen, O., Sommer, W., Hohlfeld, A., Jacobs, A. M., & Kliegl, R. (2011). Coregistration of eye movements and EEG in natural reading: Analyses and review. *Journal of Experimental Psychology: General*, *140*, 552–572.

Draschkow, D., Heikel, E., Võ, M. L.-H., Fiebach, C. J., & Sassenhagen, J. (2018). No evidence from MVPA for different processes underlying the N300 and N400 incongruity effects in object–scene processing. *Neuropsychologia*, *120*, 9–17.

Dyck, M., & Brodeur, M. B. (2015). ERP evidence for the influence of scene context on the recognition of ambiguous and unambiguous objects. *Neuropsychologia*, *72*, 43–51.

Ehinger, B. V., & Dimigen, O. (2019). Unfold: An integrated toolbox for overlap correction, non-linear modeling, and regression-based EEG analysis. *PeerJ*, *7*, e7838.

Feldman, J. (2003). What is a visual object? *Trends in Cognitive Sciences*, *7*, 252–256.

Fenske, M. J., Aminoff, E., Gronau, N., & Bar, M. (2006). Top–down facilitation of visual object recognition: Object-based and context-based contributions. *Progress in Brain Research*, *155*, 3–21.

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

Garcia-Diaz, A., Fdez-Vidal, X. R., Pardo, X. M., & Dosil, R. (2012). Saliency from hierarchical adaptation through decorrelation and variance normalization. *Image and Vision Computing*, *30*, 51–64.

Gareze, L., & Findlay, J. M. (2007). Absence of scene context effects in object detection and eye gaze capture. In R. P. G. van Gompel, M. H. Fischer, W. S. Murray, & R. W. Hill (Eds.), *Eye movements: A window on mind and brain* (pp. 617–637). Oxford, UK: Elsevier.

Hauk, O., Davis, M. H., Ford, M., Pulvermüller, F., & Marslen-Wilson, W. D. (2006). The time course of visual word recognition as revealed by linear regression analysis of ERP data. *Neuroimage*, *30*, 1383–1400.

Henderson, J. M., & Ferreira, F. (2004). Scene perception for psycholinguists. In J. M. Henderson & F. Ferreira (Eds.), *The interface of language, vision, and action: Eye movements and the visual world* (pp. 1–58). New York: Psychology Press.

Henderson, J. M., Hayes, T. R., Peacock, C. E., & Rehrig, G. (2019). Meaning and attentional guidance in scenes: A review of the meaning map approach. *Vision*, *3*, 19.

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

Hohenstein, S., & Kliegl, R. (2014). Semantic preview benefit during reading. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *40*, 166–190.
Hwang, A. D., Wang, H.-C., & Pomplun, M. (2011). Semantic guidance of eye movements in real-world scenes. *Vision Research*, *51*, 1192–1205.
Itti, L., Koch, C., & Niebur, E. (1998). A model of saliency-based visual attention for rapid scene analysis. *IEEE Transaction on Pattern Analysis and Machine Intelligence*, *20*, 1254–1259.
Jung, T.-P., Humphries, C., Lee, T.-W., Makeig, S., McKeown, M. J., Iragui, V., et al. (1998). Extended ICA removes artifacts from electroencephalographic recordings. *Advances in Neural Information Processing Systems*, *10*, 894–900.
Kamienkowski, J. E., Ison, M. J., Quiroga, R. Q., & Sigman, M. (2012). Fixation-related potentials in visual search: A combined EEG and eye tracking study. *Journal of Vision*, *12*, 4.
Kaunitz, L. N., Kamienkowski, J. E., Varatharajah, A., Sigman, M., Quiroga, R. Q., & Ison, M. J. (2014). Looking for a face in the crowd: Fixation-related potentials in an eye-movement visual search task. *Neuroimage*, *89*, 297–305.
Kliegl, R., Dambacher, M., Dimigen, O., Jacobs, A. M., & Sommer, W. (2012). Eye movements and brain electric potentials during reading. *Psychological Research*, *76*, 145–158.
Kretzschmar, F., Bornkessel-Schlesewsky, I., & Schlesewsky, M. (2009). Parafoveal versus foveal N400s dissociate spreading activation from contextual fit. *NeuroReport*, *20*, 1613–1618.
Kristensen, E., Rivet, B., & Guérin-Dugué, A. (2017). Estimation of overlapped eye fixation related potentials: The general linear model, a more flexible framework than the ADJAR algorithm. *Journal of Eye Movement Research*, *10*, 1–27.
Kutas, M., & Federmeier, K. D. (2011). Thirty years and counting: Finding meaning in the N400 component of the event-related brain potential (ERP). *Annual Review of Psychology*, *62*, 621–647.
Kutas, M., & Hillyard, S. A. (1984). Brain potentials during reading reflect word expectancy and semantic association. *Nature*, *307*, 161–163.
Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest package: Tests in linear mixed effects models. *Journal of Statistical Software*, *82*, 1–26.
LaPointe, M. R. P., & Milliken, B. (2016). Semantically incongruent objects attract eye gaze when viewing scenes for change. *Visual Cognition*, *24*, 63–77.
Loftus, G. R., & Mackworth, N. H. (1978). Cognitive determinants of fixation location during picture viewing. *Journal of Experimental Psychology: Human Perception and Performance*, *4*, 565–572.
Mackworth, N. H., & Morandi, A. J. (1967). The gaze selects information details within pictures. *Perception & Psychophysics*, *2*, 547–552.
Maris, E., & Oostenveld, R. (2007). Nonparametric statistical testing of EEG- and MEG-data. *Journal of Neuroscience Methods*, *164*, 177–190.
Matuschek, H., Kliegl, R., Vasishth, S., Baayen, H., & Bates, D. (2017). Balancing Type I error and power in linear mixed models. *Journal of Memory and Language*, *94*, 305–315.
Mensen, A., & Khatami, R. (2013). Advanced EEG analysis using threshold-free cluster-enhancement and non-parametric statistics. *Neuroimage*, *67*, 111–118.
Moores, E., Laiti, L., & Chelazzi, L. (2003). Associative knowledge controls deployment of visual selective attention. *Nature Neuroscience*, *6*, 182–189.
Mudrik, L., Lamy, D., & Deouell, L. Y. (2010). ERP evidence for context congruity effects during simultaneous object–scene processing. *Neuropsychologia*, *48*, 507–517.
Mudrik, L., Shalgi, S., Lamy, D., & Deouell, L. Y. (2014). Synchronous contextual irregularities affect early scene processing: Replication and extension. *Neuropsychologia*, *56*, 447–458.
Niefind, F., & Dimigen, O. (2016). Dissociating parafoveal preview benefit and parafovea-on-fovea effects during reading: A combined eye tracking and EEG study. *Psychophysiology*, *53*, 1784–1798.
Nikolaev, A. R., Meghanathan, R. N., & van Leeuwen, C. (2016). Combining EEG and eye movement recording in free viewing: Pitfalls and possibilities. *Brain and Cognition*, *107*, 55–83.
Nuthmann, A. (2013). On the visual span during object search in real-world scenes. *Visual Cognition*, *21*, 803–837.
Nuthmann, A. (2014). How do the regions of the visual field contribute to object search in real-world scenes? Evidence from eye movements. *Journal of Experimental Psychology: Human Perception and Performance*, *40*, 342–360.
Nuthmann, A. (2017). Fixation durations in scene viewing: Modeling the effects of local image features, oculomotor parameters, and task. *Psychonomic Bulletin & Review*, *24*, 370–392.
Nuthmann, A., de Groot, F., Huettig, F., & Olivers, C. N. L. (2019). Extrafoveal attentional capture by object semantics. *PLoS One*, *14*, e0217051.
Nuthmann, A., & Einhäuser, W. (2015). A new approach to modeling the influence of image features on fixation selection in scenes. *Annals of the New York Academy of Sciences*, *1339*, 82–96.
Nuthmann, A., & Henderson, J. M. (2010). Object-based attentional selection in scene viewing. *Journal of Vision*, *10*, 20.
Plöchl, M., Ossandón, J. P., & König, P. (2012). Combining EEG and eye tracking: Identification, characterization, and correction of eye movement artifacts in electroencephalographic data. *Frontiers in Human Neuroscience*, *6*, 278.
Rämä, P., & Baccino, T. (2010). Eye fixation-related potentials (EFRPs) during object identification. *Visual Neuroscience*, *27*, 187–192.
Rayner, K. (2014). The gaze-contingent moving window in reading: Development and review. *Visual Cognition*, *22*, 242–258.
Rayner, K., Balota, D. A., & Pollatsek, A. (1986). Against parafoveal semantic preprocessing during eye fixations in reading. *Canadian Journal of Psychology*, *40*, 473–483.
Sassenhagen, J., & Draschkow, D. (2019). Cluster-based permutation tests of MEG/EEG data do not establish significance of effect latency or location. *Psychophysiology*, *56*, e13335.
Satterthwaite, F. E. (1946). An approximate distribution of estimates of variance components. *Biometrics Bulletin*, *2*, 110–114.
Serences, J. T. (2004). A comparison of methods for characterizing the event-related BOLD timeseries in rapid fMRI. *Neuroimage*, *21*, 1690–1700.
Simola, J., Le Fevre, K., Torniainen, J., & Baccino, T. (2015). Affective processing in natural scene viewing: Valence and arousal interactions in eye-fixation-related potentials. *Neuroimage*, *106*, 21–33.
Smith, N. J., & Kutas, M. (2015a). Regression-based estimation of ERP waveforms: I. The rERP framework. *Psychophysiology*, *52*, 157–168.
Smith, N. J., & Kutas, M. (2015b). Regression-based estimation of ERP waveforms: II. Nonlinear effects, overlap correction, and practical considerations. *Psychophysiology*, *52*, 169–181.
Smith, S. M., & Nichols, T. E. (2009). Threshold-free cluster enhancement: Addressing problems of smoothing, threshold dependence and localisation in cluster inference. *Neuroimage*, *44*, 83–98.

Stoll, J., Thrun, M., Nuthmann, A., & Einhäuser, W. (2015). Overt attention in natural scenes: Objects dominate features. *Vision Research*, *107*, 36–48.
Tatler, B. W., Brockmole, J. R., & Carpenter, R. H. S. (2017). LATEST: A model of saccadic decisions in space and time. *Psychological Review*, *124*, 267–300.
Thickbroom, G. W., Knezevič, W., Carroll, W. M., & Mastaglia, F. L. (1991). Saccade onset and offset lambda waves: Relation to pattern movement visually evoked potentials. *Brain Research*, *551*, 150–156.
Underwood, G., & Foulsham, T. (2006). Visual saliency and semantic incongruency influence eye movements when inspecting pictures. *Quarterly Journal of Experimental Psychology*, *59*, 1931–1949.
Underwood, G., Templeman, E., Lamming, L., & Foulsham, T. (2008). Is attention necessary for object identification? Evidence from eye movements during the inspection of real-world scenes. *Consciousness and Cognition*, *17*, 159–170.
Ušćumlić, M., & Blankertz, B. (2016). Active visual search in non-stationary scenes: Coping with temporal variability and uncertainty. *Journal of Neural Engineering*, *13*, 016015.
Võ, M. L.-H., & Henderson, J. M. (2009). Does gravity matter? Effects of semantic and syntactic inconsistencies on the allocation of attention during scene perception. *Journal of Vision*, *9*, 24.
Võ, M. L.-H., & Henderson, J. M. (2011). Object–scene inconsistencies do not capture gaze: Evidence from the flash-preview moving-window paradigm. *Attention, Perception, & Psychophysics*, *73*, 1742–1753.
Võ, M. L.-H., & Wolfe, J. M. (2013). Differential electrophysiological signatures of semantic and syntactic scene processing. *Psychological Science*, *24*, 1816–1823.
Weiss, B., Knakker, B., & Vidnyánszky, Z. (2016). Visual processing during natural reading. *Scientific Reports*, *6*, 26902.
Winkler, I., Debener, S., Müller, K.-R., & Tangermann, M. (2015). On the influence of high-pass filtering on ICA-based artifact reduction in EEG-ERP. Paper presented at the *2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)* (pp. 4101–4105). Milan, Italy: IEEE.
Wolfe, J. M., Alvarez, G. A., Rosenholtz, R., Kuzmova, Y. I., & Sherman, A. M. (2011). Visual search for arbitrary objects in real scenes. *Attention, Perception, & Psychophysics*, *73*, 1650–1671.
Wu, C.-C., Wick, F. A., & Pomplun, M. (2014). Guidance of visual attention by semantic information in real-world scenes. *Frontiers in Psychology*, *5*, 54.
Yan, M., Richter, E. M., Shu, H., & Kliegl, R. (2009). Readers of Chinese extract semantic information from parafoveal words. *Psychonomic Bulletin & Review*, *16*, 561–566.