Atten Percept Psychophys (2017) 79:154–168
DOI 10.3758/s13414-016-1203-7

Stuck on semantics: Processing of irrelevant object-scene inconsistencies modulates ongoing gaze behavior

Tim H. W. Cornelissen [1] · **Melissa L.-H. Võ** [1]

Published online: 19 September 2016

Abstract People have an amazing ability to identify objects and scenes with only a glimpse. How automatic is this scene and object identification? Are scene and object semantics—let alone their semantic congruity—processed to a degree that modulates ongoing gaze behavior even if they are irrelevant to the task at hand? Objects that do not fit the semantics of the scene (e.g., a toothbrush in an office) are typically fixated longer and more often than objects that are congruent with the scene context. In this study, we overlaid a letter *T* onto photographs of indoor scenes and instructed participants to search for it. Some of these background images contained scene-incongruent objects. Despite their lack of relevance to the search, we found that participants spent more time in total looking at semantically incongruent compared to congruent objects in the same position of the scene. Subsequent tests of explicit and implicit memory showed that participants did not remember many of the inconsistent objects and no more of the consistent objects. We argue that when we view natural environments, scene and object relationships are processed obligatorily, such that irrelevant semantic mismatches between scene and object identity can modulate ongoing eye-movement behavior.

Keywords Scene perception · Visual search · Scene semantics · Semantic integration · Object-scene inconsistencies

A brief glimpse of a scene can be sufficient for people to extract its global meaning or gist (Castelhano & Henderson, 2008; Oliva & Schyns, 2000; Oliva & Torralba, 2006; Potter & Faulconer, 1975; Thorpe, Fize, & Marlot, 1996). Knowing the category of a scene one is looking at leads to expectations about likely objects and their positioning within, allowing—among other things—for efficient search through naturalistic scenes (Eckstein, Drescher, & Shimozaki, 2006; Mack & Eckstein, 2011; Torralba, Oliva, Castelhano, & Henderson, 2006; Võ & Henderson, 2009, 2010; Wolfe, Võ, Evans, & Greene, 2011). When searching an image of an unfamiliar bathroom for a mirror, we tend to direct our eyes to the wall above the sink and hardly look anywhere else. Although such efficient saccadic search requires knowledge of *where* objects can be in a scene, it logically also requires knowledge of *what* objects are likely to be present in a scene.

In favor of the view that rapid gist extraction leads to expectations regarding what objects can be found within a scene, there is evidence that objects that are unlikely to appear in a scene (e.g., a football in the bathroom) seem to elicit different processing than objects that are likely to be found in it.

In eye-movement research, longer fixation durations are often assumed to reflect longer or deeper processing. A broad range of experiments has shown that objects that do not fit the global identity of a scene (so-called semantically inconsistent objects) are fixated longer and more often, compared to semantically consistent objects in the same position of the same scene (e.g., Bonitz & Gordon, 2008; De Graef, Christiaens, & d'Ydewalle, 1990; Henderson, Weeks, & Hollingworth, 1999; Loftus & Mackworth, 1978; Underwood, Templeman, Lamming, & Foulsham, 2008; Võ & Henderson, 2009).

Electronic supplementary material The online version of this article (doi:10.3758/s13414-016-1203-7) contains supplementary material, which is available to authorized users.

✉ Tim H. W. Cornelissen
cornelissen@psych.uni-frankfurt.de

[1] Scene Grammar Lab, Department of Cognitive Psychology, Goethe University Frankfurt, Theodor-W.-Adorno-Platz 6, 60323 Frankfurt am Main, Germany

But what exactly is different in the processing of these semantically inconsistent objects? Is it the perceptual identification of the object or the integration of the object with existing knowledge of its context? Studies involving electrophysiology show differential event-related potentials (ERPs) elicited by objects appearing in a congruent versus an incongruent context. The difference is similar to the difference in ERP patterns elicited by semantically inconsistent versus consistent words in a sentence during reading (Ganis & Kutas, 2003; Mudrik, Lamy, & Deouell, 2010; Mudrik, Shalgi, Lamy, & Deouell, 2014; Võ & Wolfe, 2013). The time course of these ERPs evoked by scene-inconsistent objects suggests that a difference in processing occurs at a stage of semantic integration between the scene context and the representation of the object (Ganis & Kutas, 2003), and some studies indicate that context can also influence object recognition at earlier, more perceptual stages (Demiral, Malcolm, & Henderson, 2012; Mudrik et al., 2014; Võ & Wolfe, 2013).

The fact that we perform many searches through naturalistic scenes on a daily basis and seem to do so very efficiently suggests that object and scene identification (as well as their integration) are processes that require little attentional resources and might be obligatory in that they are hard to suppress. Here we seek to test to what extent the semantic integration of scene and object identities is indeed obligatory. We will refer to scene and object identification as scene semantic processing. If scene semantic processing is obligatory, we expect such processing to occur even under circumstances that do not call for it. Would, for instance, participants still fixate semantically inconsistent objects longer, even when both scene and object meaning are irrelevant to their task and when doing so is counterproductive to completing their task as fast as possible?

In a study investigating salience and semantic consistency, Underwood and Foulsham (2006) let participants search through grayscale photographs of naturalistic scenes. The target object's position was not predicted by the identity of the scene (the target was always a small ball placed somewhere in the scene). Although the authors reported longer gaze durations on semantically inconsistent objects in the scene, it cannot be said that object identity was irrelevant to the search because the target itself was one of the objects in the scene. Moreover, identifying other objects might have helped participants to decide whether the target could have some position relative to the currently fixated object (e.g., could the target ball rest on one of the objects' surfaces?). Similarly, De Graef et al. (1990) had participants search a scene for embedded nonobjects in the presence of inconsistent objects. The search for nonobjects, the authors argued, would require little or no semantic processing of the inconsistent objects. Instead of photographs the authors used line drawings, but similar to Underwood & Foulsham (2006) the nonobjects were placed *in* the scene like the other objects therein (e.g., obeying the laws of gravity and having roughly the same size as other objects). Moreover, the shape of the target changed from scene to scene, requiring participants to identify the nonobject, at least as "not an object" on every trial. Thus, it cannot be concluded that the task rendered scene–object relations in the scenes irrelevant.

In a study not involving eye movements, Greene and Fei-Fei (2014) devised a Stroop-like task in which a word was presented on top of an image of a scene or an object. Participants were instructed to classify the words as describing objects or scenes while ignoring the images. The authors varied whether the word matched the background image or not (e.g., the word *guitar* on top of an image of a guitar or on top of an image of a different object). Results showed that participants were slower to categorize the words on top of an incongruent image for both scenes and objects, which the authors took as evidence for automatic processing of the scene or object presented in the background image. Yet it remains unclear whether this kind of processing can influence ongoing gaze behavior.

To test whether participants would get "stuck" on semantically inconsistent objects, even when scene and object meaning are irrelevant to their task, we devised a visual search experiment. Participants were instructed to search for a letter target, *T*, that had been artificially overlaid on a scene and was maximally visually dissimilar from any objects in the background scene. The location of the target was therefore in no way predicted by the meaning of the scene or by the identity of the objects within. In some of the background scenes, a semantically inconsistent object was present. Part of the instruction that participants received was to search for the *T* as fast and as accurately as possible. If participants fixated the irrelevant semantic inconsistencies longer than consistent controls, then this would indicate the processing of irrelevant scene semantics despite being counterproductive to the goal of completing the search as fast as possible.

Experiment 1 followed the outline described above. Experiment 2 served as a replication of Experiment 1, with the addition of two memory tests. The memory tests served as an indication of whether or not the inconsistent objects were more noticeable to the participants than the consistent objects. We expected that if the inconsistent objects were more noticeable, then they would be encoded into memory more deeply and recalled or recognized better in a subsequent memory test. After Experiment 2, eye-movement data from the first two experiments were collapsed to further test when during search differences between conditions arise. In Experiment 3, we aimed to disrupt the processing of semantics by overlaying an extra layer of multiple *L* distractors, among which observers had to find the *T*.

General method

Participants

We gathered data from 14 participants in each experiment (Experiment 1: Mean age = 21.9 years, *SE* = 4.1, 9 female. Experiment 2: Mean age = 25.1 years, *SE* = 6.3, 11 female. Experiment 3: Mean age = 25.1 years, *SE* = 8.3, 11 female). All observers were students participating in the experiment for course credit. All were tested for normal or corrected-to-normal visual acuity and had normal color vision as assessed by the Ishihara test. All participants gave their informed consent before taking part in the experiment.

Stimulus material and design

As experimental images, we used 100 colored images of indoor real-world scenes taken from Võ and Wolfe (2013). These were created by photographing each of 50 different scenes in two versions: (1) with semantically "consistent" objects (e.g., a pot on the kitchen stove) and (2) with the scene-consistent object replaced by a semantically "inconsistent" object (e.g., a soccer ball on the kitchen stove). All images had a resolution of 1024 × 768 pixels. The bottom-up saliency of the critical objects was assessed using the MATLAB Saliency Toolbox (Walther & Koch, 2006). The rank order of saliency peaks assigned to the critical objects by the algorithm was used to ensure that consistent and inconsistent objects did not systematically differ in mean low-level saliency in relationship to the rest of the scene.

We took special care not to expose participants to the same scenes twice, particularly not with different critical objects. This was done so that effects of semantic consistency would not be confounded by participants detecting changes between two versions of the scene. Every participant therefore saw half the experimental scenes (25) in the consistent condition and the other half (25) in the inconsistent condition (see Fig. 1). The version of a scene a participant saw was randomized for every two participants and counterbalanced. The 50 experimental scenes never contained the target *T* to avoid instantaneous target detection resulting in too little eye-movement data on the scene. Our analyses thus only include target-absent trials.

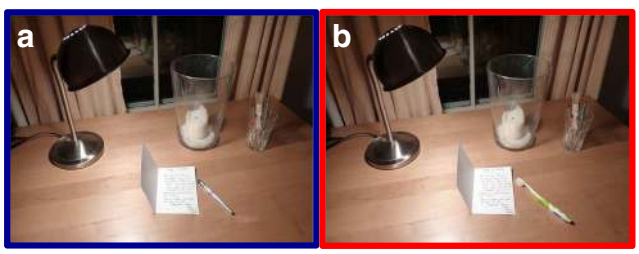

Fig. 1 Example images of the two versions of a desk scene; all objects are consistent including the pen, which is also the critical object (**a**). In the inconsistent version of the scene, the pen was replaced by a semantically inconsistent object—here, a toothbrush (**b**)

In addition to the experimental images, we used 65 *filler images* that consisted of another set of photographs of indoor scenes. These 65 filler scenes were the same for each participant and were not included for analysis. Twenty-five of the filler scenes contained a semantically inconsistent object. These 25 filler scenes with semantically inconsistent objects were created in the same way as the experimental scenes. The other 40 filler scenes were indoor scenes downloaded from the SUN database (Xiao, Hays, Ehinger, Oliva, & Torralba, 2010). In Experiments 1 and 2, a target was placed in each filler scene by inserting the 0.5° × 0.5° outline of a capital letter *T* at a random intersection of a hexagonal grid with 4 × 3 positions placed 6.0° apart. Before placing the *T*, a random direction and distance displacement was added to each position of the grid in each scene. Displacement varied from zero to half the distance between two neighboring grid points. The color of the target was always gray (128 on an 8-bit gray scale), and the *T* shape could be rotated 0°, 90°, 180°, or 270°. The target never appeared within a 2.4° radius from the center of the image. In filler scenes with a scene-inconsistent object the target never appeared on top of the inconsistent object.

The relatively small size of the target (0.5° × 0.5°), its color, and the fact that it was not part of the physical scene that was photographed, were deliberately chosen to maximize visual dissimilarity between the (critical) objects in the scenes and the target. The rationale behind this is that it could allow participants to discard objects as "not the target" based on their visual features rather than their semantic identities. Put differently, it seems highly unlikely that participants would mistake (critical) objects for the target, therefore rendering them part of the irrelevant background scene rather than an actual search distractor. The only difference in stimuli between Experiments 1 and 2 was that the counterbalancing procedure for determining presentation order and the assignment of scenes to conditions was re-run for different assignment of scenes to conditions. Changes to the stimuli in Experiment 3 are discussed in its section.

Apparatus

Eye movements were recorded with an EyeLink 1000 desktop mounted eye tracker (SR Research, Canada) at a sampling rate of 1000 Hz. Viewing was binocular, but data were recorded from the left eye only. The experiment was run on a computer running Windows 7. Stimulus presentation was controlled by MATLAB (Version 8.1.0.604), making use of the Psychophysics Toolbox (Brainard, 1997; Pelli, 1997). Subjects were seated in a dimly lit room with their heads fixated in a chinrest, in front of a 24-in. computer screen with

a refresh rate of 144 Hz and a 1920 × 1080 resolution. Viewing distance was approximately 65 cm, making the scenes subtend 23.5 × 18.1 degrees of visual angle.

Data analysis

Interest areas for the critical objects were defined as a rectangular box that was large enough to encompass the critical object in both the consistent and inconsistent versions of the scene. Thus, the interest areas were the same size for both conditions. Only the 50 experimental (target absent) trials were included for analysis, and false alarms were removed (<4 % of trials in all experiments).

Saccades and fixations were extracted from raw gaze data during recording by the EyeLink parser. Velocity and acceleration thresholds were set to the EyeLink default values of 30 degrees/s and 8,000 degrees/s^2, respectively. Fixations with durations shorter than 50 ms and longer than 2,000 ms were excluded from analysis. Duration criteria led to the exclusion of 0.7 %, 1.6 %, and 1.1 % of all fixations in Experiments 1, 2, and 3, respectively.

Procedure

Observers were first verbally informed about the task and sequence of events during a trial. Each experimental session was preceded by a 9-point eye-tracker calibration and validation procedure. Calibration was deemed successful when validation accuracy was under 0.5° for all validation points together and none of the points had an accuracy larger than 1.0°. A written reminder of the task instruction followed validation. Instructions were to search for the *T* as fast and as accurately as possible and report its absence or presence by pressing buttons on a keyboard. Each trial started with a fixation dot, automatically followed by a scene after 700 ms. In every scene, participants would search for the target letter *T*. Upon each response, the stimulus would disappear and a red or green square would indicate to the participant whether the response was correct. In case participants missed the target (i.e., reporting it absent when it was in fact present), the scene would reappear for 700 ms, with the target position indicated by a red rectangle. If no response was given within 25 seconds, the experiment automatically continued with the next trial (see also Fig. 2). None of the stimuli was ever repeated. Every participant completed 120 trials in total (five practice scenes, 50 experimental scenes, 65 filler scenes).

In Experiments 2 and 3, two subsequent surprise memory tests were added to the experimental procedure. After completing the search task, the chinrest was removed, and participants were presented with a surprise explicit recall and, subsequently, a memory recognition task.

In the *explicit memory* task, observers were presented with each of the experimental scenes once more, except this time photographed without the critical object present. Observers were informed that all scenes were taken from the search task they had just completed, except that in each scene something had been taken away or had been replaced by something else. The instruction was to first click on the position where the change had happened and subsequently type the name of the object that had been there during the search task. Unbeknown to the observers, a replacement never occurred. The instruction was only given so that observers would not simply look for empty spaces in the scene where an object might have been. We also informed observers that whether objects were replaced or taken away was completely random and that it was possible that objects were missing in all scenes or that all scenes might contain replacements.

After completing the explicit recall task, we tested subjects' *recognition memory* for critical objects by presenting both the semantically consistent and inconsistent versions of the experimental scenes side by side. In this case, participants performed a two-alternative forced-choice (2AFC) task. Instructions were to click on the version of the scene that participants thought they had seen before. There was no performance feedback during the task.

Experiment 1

To test whether participants will still fixate semantically inconsistent objects longer when scene and object meaning are irrelevant to their task, we devised a visual search experiment. Participants were instructed to search for a target that had been artificially overlaid on a scene. The location of the target was therefore in no way predicted by the meaning of the scene or by the identity of other objects in the scene. In some of the background scenes, a semantically inconsistent object was present. Part of the instruction that participants received was to search for the target as fast and as accurately as possible. If participants fixated the irrelevant semantic inconsistencies longer, then it would indicate not only the processing of irrelevant scene semantics but also would be counterproductive to the goal of completing the search as fast as possible.

Results

As measures of scene semantic processing, for each condition we calculated gaze duration measures for the critical objects including total dwell time (total time spent fixating the object during a trial), number of fixations (number of times the object was fixated during a trial), number of refixations (the number of times observers fixated the object then fixated elsewhere in the scene and then fixated the object again), mean fixation duration on the critical objects (mean duration of individual fixations on the object), and the duration of the first fixation on

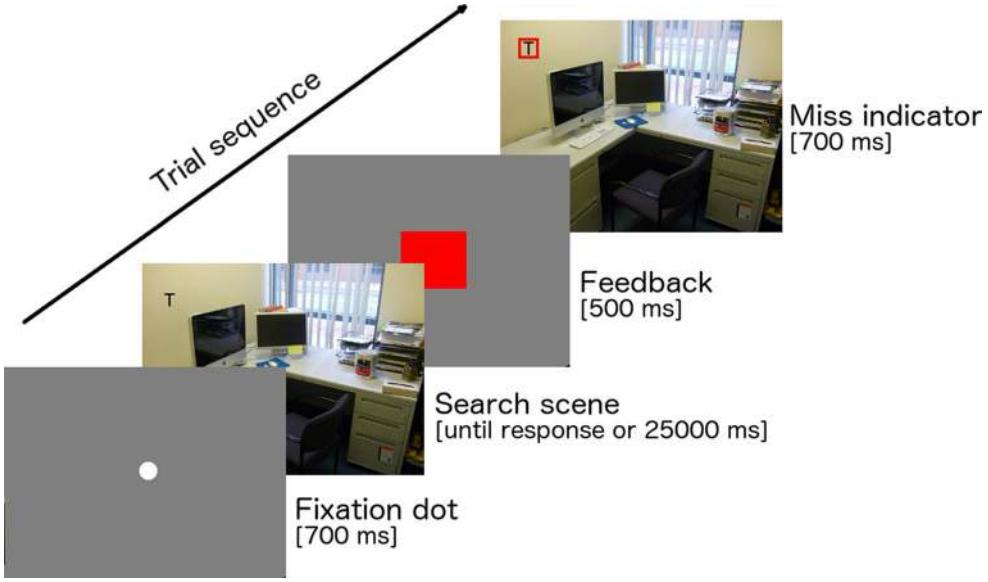

Fig. 2 Trial sequence of the search task. Note that the "miss indicator" was only shown in case of a miss response

the critical object. As descriptors of search performance we also calculated reaction times (RTs) and accuracy. Additionally, we calculated time to first fixation (the time elapsed between scene onset and participants first fixating the critical object) and the probability of fixating the critical object at least once during a trial as estimators for gaze attraction by the semantically inconsistent objects. For each measure, a mean per participant was calculated for both the inconsistent and the consistent condition and then subjected to a paired-samples *t* test.

Search performance

RTs did not statistically differ between the inconsistent (M = 9,500 ms, SE = 844 ms) and consistent condition (M = 9,240 ms, SE = 784 ms), $t(13) = -1.09$, $p = .30$. Overall accuracy of correct rejections was 96 % (SE = 1), and hit rates were 78 % (SE = 2), on average.

Gaze duration measures

As can be seen in Fig. 3, mean total dwell time was significantly higher for semantically inconsistent objects than for consistent objects, $t(13) = -3.15$, $p < .001$, as was the mean number of fixations, $t(13) = -2.45$, $p < .05$. Mean fixation duration was also significantly longer for inconsistent objects, $t(13) = -2.34$, $p < .05$, indicating that both more and longer fixations underlie the longer dwell times on semantically inconsistent objects. We did not find a significant difference in the duration of the first fixation on the critical object, $t(13) = -1.70$, $p = .11$, and no difference in the number of refixations between inconsistent objects and consistent objects, $t(13) = -1.46$, $p = .17$ (see Table 1 for all gaze-duration measures).

Gaze attraction measures

We found no difference between the probability of participants fixating the semantically inconsistent objects (M = 81 %, SE = 4) and the consistent objects (M = 77 %, SE = 3), $t(13) = -1.15$, $p = .27$. However, analysis of the time to first fixation revealed that inconsistent objects were fixated significantly earlier during a trial (M = 2,312 ms, SE = 262) than consistent objects were (M = 2.737 ms, SE = 303), $t(13) = 2.29$, $p < .05$.

Discussion

The results of Experiment 1 indicate that observers process scene semantics, even when doing so is irrelevant to their task. As mentioned earlier, scene semantics are irrelevant for the completion of this search task because, first, the identity of the scene does not provide information about the position of the

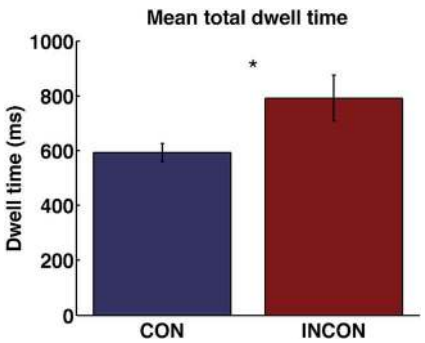

Fig. 3 Mean total dwell time on semantically consistent (CON) versus semantically inconsistent (INCON) objects. Error bars indicate standard errors. Asterisk indicates a statistically significant difference

Table 1 Summary of mean values (with standard errors) in Experiment 1 regarding dependent variables as a function of critical object consistency, including total dwell time, total number of fixations, mean fixation duration, number of refixations, and first fixation duration

Measures	Object				*df*	*t*	*p*
	Consistent		Inconsistent				
Total dwell time, in ms	592	(33)	791	(83)	13	-3.15	.008
Total number of fixations	2.5	(0.1)	3.0	(0.3)	13	-2.45	.02
Mean fixation duration, in ms	234	(7)	249	(10)	13	-2.34	.03
Number of refixations	0.80	(0.1)	0.95	(0.1)	13	-1.46	.17
First fixation duration, in ms	242	(7)	256	(10)	13	-1.70	.11

target. Second, the visual dissimilarity between critical objects and the target minimizes the need to fully identify the objects. On top of object and scene identification, object and scene identities need to be integrated to affect behavior; another processing step that would seem unnecessary or even counterproductive because object–scene relations are irrelevant to the letter search and the instruction is to search fast. Following the logic that both identification and integration are necessary to notice a scene-object mismatch, we assume that longer gaze durations on semantically inconsistent objects (i.e., mismatches) are a good indicator that scene as well as object identity processing are taking place. Participants spent more time fixating the semantically inconsistent objects than the consistent objects. This is reflected in longer total dwell times, more fixations, and longer mean fixation durations. The difference between conditions was not reflected in RTs, even though numerically RTs were on average prolonged by about 250 ms. Varying complexity and clutter in the scene, plus varying conspicuity of the target from scene to scene, might cause variance in target-absent decision times (Wolfe, 2012) larger than differences because of object consistency.

In this experiment we were mostly interested in measures upon object fixation. An ongoing debate in scene-perception literature concerns whether semantically inconsistent objects also attract gaze, indicating semantic processing in visual periphery. Even though we find earlier first fixations of semantically inconsistent objects, we refrain from drawing strong conclusions from this about attention attraction because the stimuli used were not controlled for distance to the initial fixation dot. In addition, critical objects were relatively large and therefore more easily recognizable in the visual periphery, even when further away from initial fixation.

After experimental sessions, as a check of how noticeable the semantic inconsistencies had been to participants during their search task, we asked each participant, "Did you notice anything?" Most responded that there were "odd objects" in some of the images. When asked how many they had noticed, most participants indicated "a few" or named one or two examples. In Experiment 2, we aimed to get a better grip on how noticeable the semantic inconsistencies were to participants by testing their memory for the critical objects.

Experiment 2

The goal of Experiment 2 was twofold. First, we aimed to replicate our findings from Experiment 1. Second, we aimed to measure if the semantic inconsistencies present in half of the experimental trials had been noticed by the participants. To do so, we presented participants with two surprise memory tests after the search task. One memory task to measure explicit recall, and one to measure implicit recall. If inconsistent objects in the scene were more noticeable to participants during search, we would expect stronger memory encoding for inconsistent objects than for consistent ones.

Procedure

The experimental procedure for the search task was the same as in Experiment 1. After completing the search task, the chinrest was removed and participants were presented with a surprise explicit recall and, subsequently, a memory recognition task (detailed in the General method section).

Results

Percentage of correctly remembered object identities and positions were added as measures of explicit memory. Position responses were deemed correct when observers' clicks fell within the same areas of interest as used for the analysis of eye movements. Correctness of object naming was judged manually by the experimenters. Percentage of correctly remembered scenes in the 2AFC task was added as a measure of implicit memory. All explicit and implicit recall measures were subjected to individual paired-samples *t* tests.

Search performance

Correct rejection rates in Experiment 2 were 97 % on average (*SE* = 1). Hit rates were 75 % (*SE* = 1). As in Experiment 1, we found no difference in RTs between the inconsistent (*M* = 9,240 ms, *SE* = 901) and consistent condition (*M* = 9,246 ms, *SE* = 888), $t(13) < 1$.

Gaze duration measures

Mean total dwell time was significantly higher for semantically inconsistent objects than for consistent objects, $t(13) = -3.94$, $p < .001$ (see Fig. 4 and Table 2). As was the mean number of fixations, $t(13) = -3.05$, $p < .01$. Different from Experiment 1, mean fixation duration was not significantly different for inconsistent objects, $t(13) = -1.81$, $p = .09$. There was, however, an effect on number of refixations. Contrary to Experiment 1, there was a higher number of refixations on the inconsistent than on the consistent objects, $t(13) = -3.08$, $p < .01$. Like in Experiment 1, there was no effect of object consistency on the duration of the first fixation, $t(13) < 1$.

Gaze attraction measures

The probability of fixating the inconsistent object (*M* = 0.79, *SE* = 0.03) was not significantly different from the probability of fixating the consistent object (*M* = 0.80, *SE* = 0.03), $t(13) < 1$. In line with results from Experiment 1, inconsistent objects were fixated significantly earlier during a trial (*M* = 1,758 ms, *SE* = 182) than consistent objects were (*M* = 2,560 ms, *SE* = 300), $t(13) = 2.60$, $p < .05$.

Memory recall

Mean percentage of correctly recalled positions of critical objects was 28 % (*SE* = 2) for consistent and 29 % (*SE* = 3) for inconsistent objects. Analysis yielded no significant difference between consistencies, $t(13) < 1$. During free recall, participants correctly named missing objects that were semantically consistent in 11 % (*SE* = 2) of trials and semantically inconsistent objects in 9 % of trials (*SE* = 2), with no significant difference between consistencies $t(13) < 1$.

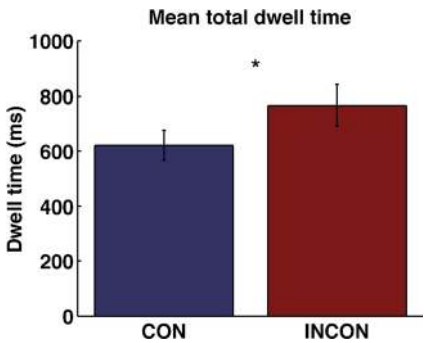

Fig. 4 Mean total dwell time on semantically consistent (CON) versus semantically inconsistent (INCON) objects. Error bars indicate standard errors. Asterisk indicates a statistically significant difference

Recognition memory

In the 2AFC task, overall performance reached 62 % correct (*SE* = 3). Using a Bonferroni-corrected alpha level of 0.0167 (0.05/3) per test, overall performance was better than chance, $t(13) = 3.17$, $p = .007$.

Analyzing performance as a function of consistency of the critical object revealed that participants on average reached 72 % correct (*SE* = 5) for consistent objects, making performance better than chance, $t(13) = 4.51$, $p < .001$. Participants performed no better than chance (*M* = 52 %, *SE* = 7) for semantically inconsistent objects, $t(13) < 1$.

Discussion

Replicating findings from Experiment 1, our results indicate that observers process scene semantics, even when doing so is irrelevant to their task. Interestingly, neither recall nor recognition memory measures show evidence of stronger encoding of semantically inconsistent objects. Gaze duration measures reflect the same pattern as in Experiment 1 except for two. First, mean fixation durations were again increased for inconsistent objects, but this difference failed to reach statistical significance ($p = .09$). Second, the number of refixations to the critical object was significantly larger for inconsistent objects in Experiment 2, as opposed to in Experiment 1. This indicated that perhaps longer dwell times are not only due to more and on average longer fixations but also due to participants looking back toward the inconsistent objects more often over the course of a trial than to the consistent objects.

In Experiment 2, we tested participants' memory as a check on whether or not inconsistent objects had been more noticeable to the participants compared to their consistent counterparts. Our results indicate that despite longer gaze durations on inconsistent objects, these objects did not somehow stand out to a degree that enhanced memory performance for them. During free recall, participants first had to click where they thought the presented scene had changed since the search task and subsequently name the object that was there during the search task. Participants managed to name few critical objects and no more of the inconsistent ones than the consistent ones. Perhaps free recall was difficult because participants were focused on the search task and not on memorizing the objects in the scene. Recognition memory, however, showed better recall of the consistent objects than the inconsistent objects. This was unexpected because, on average, more time was

Table 2 Summary of mean values (with standard errors) in Experiment 2 regarding dependent variables as a function of critical object consistency, including total dwell time, total number of fixations, mean fixation duration, number of refixations, and first fixation duration

Measures	Object				*df*	*t*	*p*
	Consistent		Inconsistent				
Total dwell time, in ms	621	(55)	766	(76)	13	-3.94	.002
Total number of fixations	2.5	(0.2)	3.0	(0.3)	13	-3.05	.009
Mean fixation duration, in ms	242	(6)	251	(8)	13	-1.81	.09
Number of refixations	0.8	(0.1)	1.1	(0.2)	13	-3.08	.009
First fixation duration, in ms	242	(7)	244	(10)	13	<1	.80

spent looking at the inconsistent objects. It could be the case that memory for consistent objects is generally better than for inconsistent objects (e.g., Gronau & Shachar, 2015). Alternatively, perhaps participants had a bias to choose the consistent version of a scene over the inconsistent version. Doing so would make for more correct "consistent" decisions, but not necessarily for better memory. Without a proper control condition in the 2AFC task—for instance, where participants choose between two semantically consistent versions of the same scene—we cannot distinguish response bias from "true" memory. Nevertheless, it is an interesting observation that inconsistent objects were not remembered better *despite* participants having looked at these objects longer. We will return to this observation in the General Discussion.

Most important to our hypotheses, though, our data do not indicate better memory for inconsistent objects than for consistent ones. The fact that the inconsistent objects are not recalled or recognized better than consistent objects serves as an indication that the observed effect of consistency on gaze duration does not stem from semantic inconsistencies somehow being more noticeable to participants during the search task.

A question that remains after Experiment 1 and its replication in Experiment 2 is at what point during a trial do participants start getting "stuck" on the irrelevant, semantically inconsistent objects. From an extreme standpoint, one could, for instance, argue—because of the less-clear criteria for terminating search in target-absent trials (Wolfe, 2012), and given the relatively long average response times in our experiments—that instead of processing irrelevant scene semantics *during* visual search, participants first decide that the target is absent and then attend to objects and their consistency with the overall scene meaning before responding (even though this would go directly against the instruction to complete the task as fast as possible). If participants indeed searched first and only attended to semantic inconsistencies once they had given up, we would expect differences between conditions in gaze duration measures to arise relatively late during trials. To test this, we calculated the point where differences between conditions became statistically significant and compared it to the mean RT in target-present trials. We assume that RTs in target-present trials reflect the minimal amount of time participants spend searching in target-absent trials, a conservative assumption. RTs in target-absent trials are generally longer than in target-present trials (Wolfe, 1998, 2012). As a measure of gaze duration, we chose total dwell time because it captures both duration of fixations and number of fixations in one measure.

As a first step, eye-movement data from Experiment 1 and 2 were collapsed, bringing the number of participants to 28. Collapsing was done to maximize the amount of eye-movement data underlying the averages for early time points. Second, we calculated mean total dwell time per participant for 245 subsequent time points which, in steps of 100 ms, ranged from 100 ms to 25,000 ms (the maximum trial duration). Total dwell time for time points beyond the RT of a trial was set to the total dwell time upon response. Calculating total dwell time up to each time point was done in the same way as prior analyses. Then, at each time point, means were submitted to a paired-samples *t* test. We corrected for multiple comparisons by temporal clustering of *t* values as described by Maris and Oostenveld (2007) and implemented by Pernet, Chauveau, Gaspar, and Rousselet (2011). The cluster correction effectively eliminates small clusters of significant time points that likely result from false positives. Correction was performed with 1,000 resampling trials and alpha = .05. Analysis revealed that the difference in total dwell time between conditions is already statistically significant from the 1,000-ms time point onwards (see also Fig. 5). This is considerably earlier than the average RT in target-present trials, which was 2,703 ms (*SD* = 635). We therefore conclude that even though response times were longer than 9 seconds on average in both experiments, participants were already processing irrelevant scene semantics (i.e., getting "stuck" on semantically inconsistent objects) early during a trial while still performing visual search, and not merely after terminating the search for the target. For further qualitative assessment of eye movements and viewing behavior over time, we

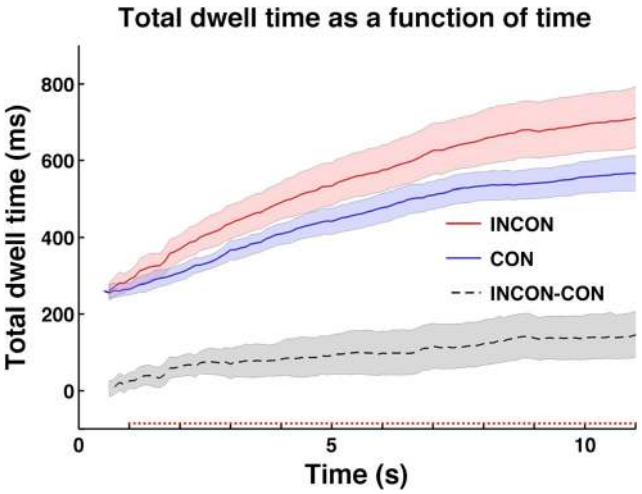

Fig. 5 Mean total dwell time on semantically consistent (CON) versus semantically inconsistent (INCON) objects at different time points during Experiments 1 and 2, combined. INCON-CON indicates the difference between consistencies. Shaded areas represent 95 % percentile bootstrap confidence intervals, 1,000 resampling trials (Wilcox, 2005). Red marks plotted along the *x*-axis indicate statistical significance of the difference between consistencies at that time point, with $p < .05$ and corrected for multiple comparisons. (Color figure online)

calculated and visually inspected the development of both fixation durations and saccade amplitudes as a function of time. To this end, all fixations and saccades within a trial were included. Saccade amplitude rapidly increased and then more gradually decreased with time, while fixation duration gradually increased (for plots, see the Supplementary Materials). This pattern is consistent with previous literature (Godwin, Reichle, & Menneer, 2014; Mills, Hollingworth, Van der Stigchel, Hoffman, & Dodd, 2011; Over, Hooge, Vlaskamp, & Erkelens, 2007; Unema, Pannasch, Joos, & Velichkovsky, 2005) and is thought to indicate a coarse-to-fine search strategy of scanning a search array in a global manner initially, before progressing to more fine and local scanning (Over et al., 2007; but see also Godwin et al., 2014). Furthermore, Over et al. (2007) found such coarse-to-fine eye-movement behavior regardless of whether conspicuity of the target varies and is unknown to the searcher (like in our experiments) or is constant across trials.

Having replicated our findings from Experiment 1 with regard to processing task-irrelevant semantics and having established that differences in gaze behavior between conditions already emerge relatively early within sometimes long-lasting trials, we wondered how obligatory such processing is and what role attention plays in the processing of scene semantics. Perhaps if there is more relevant visual information to focus attention on that is not part of the background scene, the processing of scene semantics can be disrupted. In Experiment 3, we therefore aimed to interfere with the processing of scene semantics by adding artificially overlaid distractors.

Experiment 3

In an attempt to disrupt the processing of irrelevant scene semantics, we overlaid more than one search element on the background scenes from Experiments 1 and 2. We reason that a grid of search elements might form an extra layer on top of the scene that participants could focus their attention on rather than on the background scene. If more attention is focused on the grid, perhaps less attention is available to automatically process object–scene inconsistencies, resulting in participants not getting "stuck" on irrelevant semantics anymore. Conversely, if no or very little attention is needed to process object–scene inconsistencies, we expect to find results similar to those from Experiment 1 and Experiment 2.

Stimuli

In Experiment 3, the scene photographs from Experiments 1 and 2 were used once more, assigned to conditions in the same manner. The only adaptation to the stimuli was the number of search elements and their positions. Rather than placing a single target only in the target-present scenes, distractors were placed on every scene, replacing one by a target in target-present trials. Eighteen search elements were distributed across 30 possible positions in a hexagonal grid (6 columns, 5 rows). Spacing between grid positions was 200 pixels (5°) with a random direction jitter of 30 pixels (0.7°) added to each position. Elements were placed so that no element was within a 2.4° radius from the image center. Also, no element was ever placed within any of the critical object areas of interest so that there would be no effect of visibility of search elements confounding gaze duration measures. (See Fig. 6 for a schematic example with enlarged search elements.)

Procedure

The experimental procedure for the search task was the same as in Experiment 2. After completing the search task, the chinrest was removed and participants were presented with an explicit recall and, subsequently, a memory recognition task.

Results

Search performance

Correct rejection rates in Experiment 3 were 97 % on average (*SE* = 1). Hit rates were 74 % (*SE* = 3). Analysis revealed no

Fig. 6 Schematic example of the semantically inconsistent version of a desk scene (see Fig. 1b), as used in Experiment 3. Note that, for display purposes, search elements have been enlarged by a factor 2.5

difference in reaction times between the inconsistent (M = 9,820 ms, SE = 593) and consistent condition (M = 9,859 ms, SE = 629), $t(13) < 1$.

Gaze duration measures

As shown in Table 3 and Fig. 7, the overall pattern of results remained similar to Experiments 1 and 2. However, the grid manipulation decreased the differences between the semantically consistent and inconsistent objects to become statistically nonsignificant for any of the gaze duration measures included here.

Gaze attraction measures

The probability of fixating the inconsistent object (M = 70 %, SE = 3) was not significantly different from the probability of fixating the consistent object (M = 65 %, SE = 4), $t(13) < 1$. Contrary to Experiments 1 and 2, we found no evidence of inconsistent objects being fixated significantly earlier during a trial (M = 3,424, SE = 302) than consistent objects were (M = 3,671, SE = 330), $t(13) = 1.14$, $p = .28$.

Table 3 Summary of mean values (with standard errors) in Experiment 3 regarding dependent variables as a function of critical object consistency, including total dwell time, total number of fixations, mean fixation duration, number of refixations, and first fixation duration

Measures	Object				*df*	*t*	*p*
	Consistent		Inconsistent				
Total dwell time, in ms	456	(37)	532	(31)	13	-1.74	.11
Total number of fixations	2.0	(0.1)	2.2	(0.1)	13	-1.44	.17
Mean fixation duration, in ms	226	(7)	234	(10)	13	-1.23	.24
Number of refixations	0.5	(0.1)	0.6	(0.1)	13	-0.90	.39
First fixation duration, in ms	226	(8)	232	(9)	13	-0.66	.52

Memory recall

Mean percentage of correctly recalled positions of critical objects was 31 % (SE = 2) for consistent and 28 % (SE = 3) for inconsistent objects. Analysis yielded no significant difference between consistencies, $t(13) < 1$. When naming missing objects, participants correctly named missing objects in 5 % (SE = 2) of all semantically inconsistent trials and 7 % (SE = 1) of all semantically consistent trials, with no significant difference between consistencies $t(13) = 1.27$, $p = .22$.

Recognition memory

In the 2AFC task, overall performance reached 59 % correct (SE = 2). Using a Bonferroni-corrected alpha level of 0.0167 (0.05/3) per test, overall performance was better than chance, $t(13) = 4.64$, $p < .001$.

Analyzing performance as a function of consistency of the critical object revealed that participants on average reached 71 % correct (SE = 4) for consistent objects, making performance better than chance $t(13) = 5.37$, $p < .001$. Participants performed no better than chance (M = 47 %, SE = 4) for semantically inconsistent objects, $t(13) < 1$.

Comparison between experiments

To better understand the influence of the grid of search elements, data from Experiments 2 and 3 were additionally submitted to 2 × 2 (Object Consistency × Grid) ANOVAs, with the main difference between experiments (i.e., the grid) as between-subjects factor.

Search performance

Analysis of *reaction times* in target-absent trials revealed no main effect of grid, $F(1, 26) < 1$, no main effect of object consistency, $F(1, 26) < 1$, and no interaction, $F(1, 26) < 1$.

Gaze duration measures

Total dwell time Analysis revealed a main effect of grid, $F(1, 26) = 8.40$, $p < .01$. Compared to Experiment 2, total dwell time on critical objects was reduced with the grid present (for means, see Table 4). Analysis of dwell times also yielded a significant main effect of object consistency, $F(1, 26) = 15.02$, $p < .001$, indicating dwell time was longer for inconsistent objects. No significant interaction was found, $F(1, 26) = 1.48$, $p = .24$.

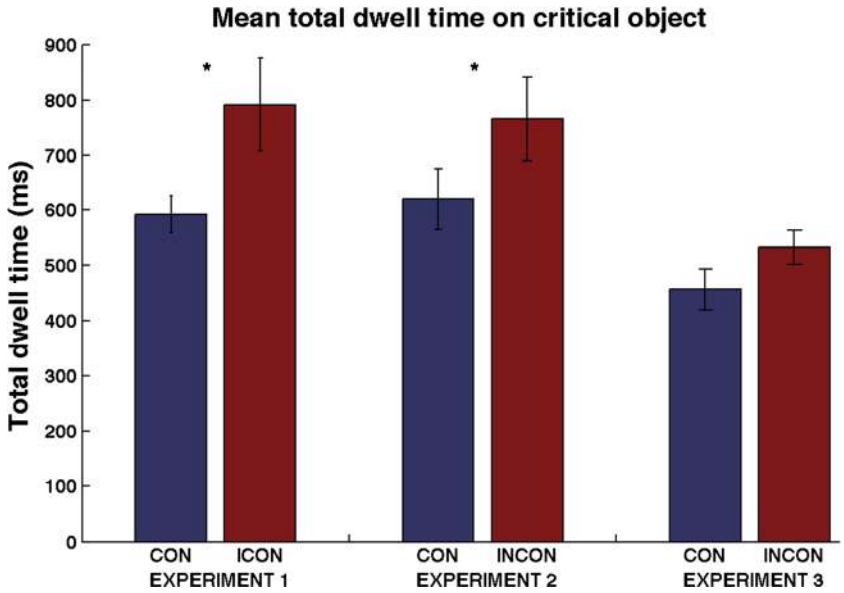

Fig. 7 Mean total dwell time on semantically consistent (CON) versus semantically inconsistent (INCON) objects for each experiment. Error bars indicate standard errors. Asterisks indicate statistically significant differences

Total number of fixations Total number of fixations was significantly reduced with a grid of elements present compared to Experiment 2, $F(1, 26) = 5.21$, $p < .05$. There was also a significant main effect of object consistency, $F(1, 26) = 9.70$, $p < .01$, but no significant interaction, $F(1, 26) < 1$.

Mean fixation durations A main effect of object consistency did not reach significance, $F(1, 26) = 3.67$, $p = .07$, just as the main effect of grid, $F(1, 26) = 9.70$, $p = .07$. No significant interaction was found, $F(1, 26) < 1$.

Number of refixations There was a main effect of grid $F(1, 26) = 7.89$, $p < .01$, reflecting fewer refixations of critical objects in the presence of multiple search elements. A significant main effect of object consistency also became apparent, $F(1, 26) = 7.07$, $p < .05$ Indicating more refixations to semantically inconsistent objects. We found no significant interaction effect, $F(1, 26) = 1.64$, $p = .21$.

Duration of the first fixation Analysis yielded no significant main effect of grid $F(1, 26) = 1.69$, $p = .21$. Similarly we found no main effect of object consistency, $F(1, 26) < 1$, and no significant interaction effect, $F(1, 26) < 1$.

Gaze attraction measures

Time to first fixation showed a main effect of object consistency, $F(1, 26) = 7.75$, $p < .01$, indicating inconsistent objects were fixated earlier than consistent controls. As for differences between experiments, analysis revealed that there was a

Table 4 Summary of mean values (with standard errors) in Experiment 2 and 3 regarding dependent variables as a function of critical object consistency, including reaction time, total dwell time, total number of fixations, mean fixation duration, number of refixations, first fixation duration, time to first fixation, and probability of fixating the critical object. *F* values and *p* values are given for the main effect of grid

Measures	Experiment 2				Experiment 3				*F*	*p*
	Consistent		Inconsistent		Consistent		Inconsistent			
Reaction time, in ms	9,246	(888)	9,240	(901)	9,859	(629)	9,820	(593)	<1	.82
Total dwell time, in ms	621	(55)	766	(76)	456	(37)	532	(31)	8.40	.0006
Total number of fixations	2.5	(0.2)	3.0	(0.3)	2.0	(0.1)	2.2	(0.1)	5.21	.004
Mean fixation duration, in ms	242	(6)	251	(8)	226	(7)	234	(10)	9.70	.07
Number of refixations	0.8	(0.1)	1.1	(0.2)	0.5	(0.1)	0.6	(0.1)	7.89	.01
First fixation duration, in ms	242	(7)	244	(10)	226	(8)	232	(9)	1.69	.50
Time to first fixation, in ms	2560	(300)	1758	(182)	3671	(330)	3424	(302)	15.29	.01
Probability of fixation	0.80	(0.03)	0.79	(0.03)	0.65	(0.04)	0.70	(0.03)	12.51	.002

significant main effect of grid, $F(1, 26) = 15.29$, $p < .001$, indicating that time to first fixation was longer in Experiment 3 than it was in Experiment 2.

Comparing the *probability of fixating the critical object* showed a main effect of grid, $F(1, 26) = 12.51$, $p < .01$, that indicates a smaller probability of fixating the critical objects with multiple search elements present. Furthermore, we found no evidence for a main effect of object consistency, $F(1, 26) < 1$, and no interaction effect, $F(1, 26) < 1$.

Memory recall

Comparison of *memory recall* in Experiments 2 and 3 yielded no significant main effects of grid, $F(1, 26) = 1.81$, $p = .19$, of condition, $F(1, 26) = 1.63$, $p = .21$, and no interaction effect, $F(1, 26) = 1.48$, $p = .24$. Analysis of *proportion of correctly remembered object positions* revealed no main effect of grid, no main effect of condition, and no interaction effect, all Fs < 1.

Discussion

First, comparisons between Experiments 2 and 3 show that placing multiple search elements on top of the scenes in the form of an overlaid grid reduced total dwell time, number of fixations, and number of refixations on critical objects compared to Experiment 2. Additional calculations and subsequent visual inspection of the area covered by participants' gaze over time (see the Supplementary Materials) showed that participants looked at a greater proportion of the scene and did so faster in Experiment 3 than in Experiments 1 and 2. This indicates that gaze was more spread out with the grid in place than without it, even though the exact same background scenes were presented. Statistical analyses of Experiment 3 did not reveal significant differences between object consistencies for any of the gaze duration measures included here, despite the pattern of results being the same as in Experiment 1 and Experiment 2. However, the comparison of Experiments 2 and 3 revealed main effects of object consistency on dwell time, number of fixations, and number of refixations. Judging by the individual analyses of Experiments 2 and 3, these main effects are likely driven by the differences in Experiment 2 rather than representing a statistically reliable difference between semantically consistent and inconsistent objects in Experiment 3.

Furthermore, memory measures from Experiment 3 show the same pattern as in Experiment 2, with no significant difference in recall performance between experiments. Recognition memory in Experiment 3 was, again, better for consistent objects than for inconsistent objects.

General discussion

When we view the world around us, we usually know within a glimpse what we are looking at. We have knowledge of what environment we are currently in and expectations about objects likely to appear in that environment. Whether we just look around a room or search for a specific object in it, we seem to be constantly identifying things. How else would you know that you found a pen to write down that phone number, if you had not identified the pen? The ease with which we find our way around our visual world, search for, and identify objects in it suggests that such behavior relies on processes that require little attentional resources and might be obligatory, in a sense that they are hard to suppress. To explore how obligatory the processing of scene and object identity are, we designed a search task in which scene and object meaning provided no information about target location. We hypothesized that if scene and object identification are difficult to suppress, participants would process both scene and object identity, plus their semantic relationship, even when scene identity provided no information about possible target locations and objects could hardly be mistaken for the target. More specifically, we hypothesized that if scene and object identification are not easily suppressed, participants would fixate objects that are semantically inconsistent with the scene they are placed in longer than semantically consistent objects. Fixating semantically inconsistent objects longer would not only indicate processing of irrelevant scene semantics. Given that task instruction was to complete the search as fast and as accurately as possible, it would also be somewhat counterproductive to the goal of finding the target letter as fast as possible.

Experiment 1 showed that participants do fixate semantically inconsistent objects longer and more often despite the irrelevance of scene semantics to the search. In Experiment 2, we replicated these findings and added two surprise memory tests. Both memory tests showed that memory for semantically inconsistent objects was no better than for consistent controls. This indicates that the semantically inconsistent objects were not more noticeable than the semantically consistent objects during search, despite participants getting "stuck" on them. Additional analyses of gaze duration on critical objects showed that despite the relatively long target-absent responses, effects of object consistency arise as early as 1 second into a trial. These additional analyses provide further evidence against the explanation that observers in this task only got stuck late during target-absent trials. Rather, semantic processing of objects and their context takes place early on during scene viewing.

The fact that recognition memory for consistent objects was better than for inconsistent objects is a finding that invites further discussion. Note that there seems to be no general consensus in the literature about whether objects in a consistent surrounding are encoded into memory better or whether

this holds for objects *in*consistent with their surroundings (see the introduction to Gronau & Shachar, 2015, for a brief comprehensive overview). Based on the current paradigm, we cannot distinguish whether our findings of better memory for consistent objects stem from a response bias or from truly better memory. Perhaps for future research and stimulus development it would be interesting to eliminate response bias for consistent objects as a possible factor (e.g., by creating two consistent and two inconsistent versions of each scene and presenting these to the participant; cf. Hollingworth & Henderson, 1998). Similarly, in an ideal stimulus set objects would also be counterbalanced across scenes and balanced for typicality. We are currently creating such a stimulus set for future use.

The more intriguing point about the current results, however, is that memory performance for consistent objects was better *despite* the fact that more time was spent looking at the inconsistent objects. It seems intuitive that longer gaze should have given participants more time to encode the inconsistent objects into memory.

We have already put forward the possibility of a bias in the decision process involved in the recognition task. Perhaps memory was poor in general, as we find for the inconsistent objects, leading participants to choose what seemed most plausible in the 2AFC task. This explanation would entail that although semantic consistency influenced gaze behavior, little or no incidental memory was formed for objects in the background scene, irrespective of the objects' consistencies.

An alternative explanation is that longer gaze durations do not imply more time for incidental memory encoding, but might reflect difficulties in processes that usually facilitate memory encoding, for instance, object identification. Gordon (2006) for example, proposes that inconsistent objects attract attention due to difficulty in identifying an inconsistent object. Although Gordon (2006) uses short presentation times and a paradigm that does not allow for eye movements, the hypothesized difficulty might be resolved once the object is close enough to the fovea for identification (but see also Võ & Henderson, 2010, 2011, who dispute gaze attraction to inconsistent objects).

Finally, in Experiment 3, we looked at the strength of irrelevant scene semantic processing by placing an extra layer consisting of multiple search elements on the scene rather than only one target. Adding distractors led to a statistically non-significant increase in target-absent RT of about 600 ms (see Table 4). Additionally, participants covered more of the scene with their gaze, and did so faster in Experiment 3 than in Experiments 1 and 2 (see the Supplementary Materials). This indicates that gaze was in general more spread out with the grid in place than without it, even though the exact same background scenes were presented. If the processing of semantic inconsistencies between objects and scenes requires attention, then providing participants with other, more relevant visual information to focus attention on should reduce the effects of object inconsistency. Note that the question of whether object and scene recognition by themselves require attention is subject to debate (e.g., Evans & Treisman, 2005; Li, VanRullen, Koch, & Perona, 2002). While both might be recognized with little or no attention, the processing of the congruency of an object and its surrounding scene might require focused attention after all. Gronau and Shachar (2014), for instance, investigated spatial consistency between isolated objects and found attention was necessary for integration (but see also Munneke, Brentari, & Peelen, 2013, who find semantic consistency effects regardless of the focus of spatial attention on an object or its surrounding scene). Similarly, Mudrik, Breska, Lamy, and Deouell (2011) demonstrated semantic integration even without visual awareness of the stimulus. Yet that finding has been challenged in a recent study by Moors, Boelens, van Overwalle, and Wagemans (2016). At first glance, our results from Experiment 3 seem to favor the view that scene semantic processing relies on attention. With multiple distractors to focus attention on, no significant differences between object consistencies in gaze duration measures were found in Experiment 3. This could be taken as evidence that scene semantic processing was disrupted by our grid manipulation and that diverting visual attention from the critical objects to the grid left less attentional resources for semantic processing. However, more nuanced conclusions might be necessary.

First, the directions of all differences between conditions were similar to Experiments 1 and 2. In addition, the number of fixations on and the probability of fixating the critical object at all—regardless of it being either consistent or inconsistent with its scene context—were significantly reduced in Experiment 3. This led to fewer fixations being included in the analysis, possibly lowering statistical power along with the duration differences in total dwell time in Experiment 3. Moreover, there is strong evidence that fixation locations and the focus of attention are closely coupled (Deubel & Schneider, 1996). That, however, does not rule out the possibility that scene semantics were still processed in the visual periphery with little need for focused attention, while gaze (i.e., overt attention) was directed to the more task-relevant grid of distractors and away from the semantic inconsistencies. The finding that, with the grid of distracters in place, time to first fixation was longer and not significantly different between consistencies further supports the notion that gaze was preferentially allocated to the added distracters.

Our results therefore can be taken to imply that, on the one hand, semantic processing cannot be totally suppressed when semantics are irrelevant and gaze is directed away from semantic inconsistencies, but, on the other hand, also provide evidence that semantic processing is not obligatory enough to

always modulate gaze duration upon fixation of an inconsistency.

Related to the notion that we cannot rule out semantic processing based on the absence of a gaze duration effects, we want to point out that the current work focuses on measures of gaze modulation once the eyes have fixated semantic inconsistencies. However, we also found earlier fixations of inconsistent as compared to consistent objects in Experiments 1 and 2, implying processing of scene semantics before foveating the critical objects. This adds to the ongoing discussion as to what degree semantic processing is taking place in the visual periphery. Measures of gaze attraction by semantically inconsistent objects have been used to shed light on this issue but delivered mixed results (Bonitz & Gordon, 2008; De Graef et al., 1990; Henderson et al., 1999; Underwood et al., 2008; Võ & Henderson, 2009, 2011). Differences in the types of stimulus material used across the different study might have caused the mixed findings so far. The degree to which objects can be identified in the visual periphery will greatly depend on the visual properties of these, particularly their size and the degree of visual clutter surrounding the objects. It is also possible that gaze attraction is not a sufficiently sensitive measure of semantics processing. What if the gist of a scene includes object semantics, but this is not always reflected in gaze behavior? Perhaps what is needed to settle the debate about gaze attraction by semantically inconsistent objects and about object semantics in scene gist, is a measure or a combination of measures that capture the processing of object identities and gaze attraction separately. Perhaps a combination of EEG measures mentioned earlier (Ganis & Kutas, 2003; Mudrik et al., 2010, 2014; Võ & Wolfe, 2013) and eye tracking can serve the purpose (cf. Dimigen, Sommer, & Hohlfeld, 2011).

To conclude, our results show that scene and object identification are still taking place even when circumstances render the processing of semantic relationships irrelevant. In addition, we have demonstrated that this processing of irrelevant scene semantics can influence ongoing gaze behavior even when this is counterproductive to current task demands. It is important to note that the "strange" objects that participants got "stuck" on were not noticeable to a degree during search that participants formed stronger memories of them. The lack of a memory effect also implies a dissociation between the time spent looking at an object and the degree to which it is being encoded into memory. In conclusion, it seems that we cannot completely switch off the semantic processing of our environment, even if occupied with a thoroughly nonsemantic task.

Acknowledgments This work was supported by Deutsche Forschungsgemeinschaft (DFG) Grant VO 1683/2-1 awarded to M. L.-H. Võ.

References

Bonitz, V. S., & Gordon, R. D. (2008). Attention to smoking-related and incongruous objects during scene viewing. *Acta Psychologica, 129*(2), 255–263. doi:10.1016/j.actpsy.2008.08.006

Brainard, D. H. (1997). The psychophysics toolbox. *Spatial Vision, 10,* 433–436.

Castelhano, M. S., & Henderson, J. M. (2008). The influence of color on the perception of scene gist. *Journal of Experimental Psychology: Human Perception and Performance, 34*(3), 660–675. doi:10.1037/0096-1523.34.3.660

De Graef, P., Christiaens, D., & d'Ydewalle, G. (1990). Perceptual effects of scene context on object identification. *Psychological Research, 52*(4), 317–329. doi:10.1007/BF00868064

Demiral, Ş. B., Malcolm, G. L., & Henderson, J. M. (2012). ERP correlates of spatially incongruent object identification during scene viewing: Contextual expectancy versus simultaneous processing. *Neuropsychologia, 50*(7), 1271–1285. doi:10.1016/j.neuropsychologia.2012.02.011

Deubel, H., & Schneider, W. X. (1996). Saccade target selection and object recognition: Evidence for a common attentional mechanism. *Vision Research, 36*(12), 1827–1837.

Dimigen, O., Sommer, W., & Hohlfeld, A. (2011). Coregistration of eye movements and EEG in natural reading: Analyses and review. *Journal of Experimental Psychology: General, 140*(4), 552–572. doi:10.1037/a0023885)

Eckstein, M. P., Drescher, B. A., & Shimozaki, S. S. (2006). Attentional cues in real scenes, saccadic targeting, and Bayesian priors. *Psychological Science, 17*(11), 973–980. doi:10.1111/j.1467-9280.2006.01815.x

Evans, K., & Treisman, A. (2005). Perception of objects in natural scenes: Is it really attention free? *Journal of Experimental Psychology: Human Perception and Performance, 31*(6), 1476–1492. doi:10.1037/0096-1523.31.6.1476

Ganis, G., & Kutas, M. (2003). An electrophysiological study of scene effects on object identification. *Cognitive Brain Research, 16*(2), 123–144. doi:10.1016/S0926-6410(02)00244-6

Godwin, H. J., Reichle, E. D., & Menneer, T. (2014). Coarse-to-fine eye movement behavior during visual search. *Psychonomic Bulletin & Review, 21*(5), 1244–1249.

Gordon, R. D. (2006). Selective attention during scene perception: Evidence from negative priming. *Memory & Cognition, 34*(7), 1484–1494.

Greene, M. R., & Fei-Fei, L. (2014). Visual categorization is automatic and obligatory: Evidence from Stroop-like paradigm. *Journal of Vision, 14*(1), 1–11. doi:10.1167/14.1.14

Gronau, N., & Shachar, M. (2014). Contextual integration of visual objects necessitates attention. *Attention, Perception, & Psychophysics, 76*(3), 695–714.

Gronau, N., & Shachar, M. (2015). Contextual consistency facilitates long-term memory of perceptual detail in barely seen images. *Journal of Experimental Psychology: Human Perception and Pperformance, 41*(4), 1095.

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*(1), 210–228. doi:10.1037/0096-1523.25.1.210

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

Li, F. F., VanRullen, R., Koch, C., & Perona, P. (2002). Rapid natural scene categorization in the near absence of attention. *Proceedings of the National Academy of Sciences, 99*(14), 9596–9601. doi:10.1073/pnas.092277599

Loftus, G. R., & Mackworth, N. H. (1978). Cognitive determinants of fixation location during picture viewing. *Journal of Experimental Psychology: Human Perception and Performance, 4*(4), 565–572. doi:10.1037/0096-1523.4.4.565

Mack, S. C., & Eckstein, M. P. (2011). Object co-occurrence serves as a contextual cue to guide and facilitate visual search in a natural viewing environment. *Journal of Vision, 11*(9), 9. doi:10.1167/11.9.9

Maris, E., & Oostenveld, R. (2007). Nonparametric statistical testing of EEG- and MEG-data. *Journal of Neuroscience Methods, 164*(1), 177–190. doi:10.1016/j.jneumeth.2007.03.024

Mills, M., Hollingworth, A., Van der Stigchel, S., Hoffman, L., & Dodd, M. D. (2011). Examining the influence of task set on eye movements and fixations. *Journal of Vision, 11*(8), 17.

Moors, P., Boelens, D., van Overwalle, J., & Wagemans, J. (2016). Scene integration without awareness: No conclusive evidence for processing scene congruency during continuous flash suppression. *Psychological Science*. doi:10.1177/0956797616642525

Mudrik, L., Breska, A., Lamy, D., & Deouell, L. Y. (2011). Integration without awareness: Expanding the limits of unconscious processing. *Psychological Science, 22*(6), 764–770.

Mudrik, L., Lamy, D., & Deouell, L. Y. (2010). ERP evidence for context congruity effects during simultaneous object–scene processing. *Neuropsychologia, 48*(2), 507–517. doi:10.1016/j.neuropsychologia.2009.10.011

Mudrik, L., Shalgi, S., Lamy, D., & Deouell, L. Y. (2014). Synchronous contextual irregularities affect early scene processing: Replication and extension. *Neuropsychologia, 56*(C), 447–458. doi:10.1016/j.neuropsychologia.2014.02.020

Munneke, J., Brentari, V., & Peelen, M. (2013). The influence of scene context on object recognition is independent of attentional focus. *Frontiers in Psychology, 4,* 552.

Oliva, A., & Schyns, P. G. (2000). Diagnostic colors mediate scene recognition. *Cognitive Psychology, 41*(2), 176–210. doi:10.1006/cogp.1999.0728

Oliva, A., & Torralba, A. (2006). Building the gist of a scene: The role of global image features in recognition. In S. Martinez-Conde, S. Macknik, M. M. Martinez, J.-M. Alonso, & P. U. Tse (Eds.), *Visual perception, Part 2—Fundamentals of awareness: Multisensory integration and high-order perception* (Vol. 155, pp. 23–36). Amsterdam: Elsevier. doi:10.1016/S0079-6123(06)55002-2

Over, E. A. B., Hooge, I. T. C., Vlaskamp, B. N. S., & Erkelens, C. J. (2007). Coarse-to-fine eye movement strategy in visual search. *Vision Research, 47*(17), 2272–2280.

Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: Transforming numbers into movies. *Spatial Vision, 10,* 437–442.

Pernet, C. R., Chauveau, N., Gaspar, C., & Rousselet, G. A. (2011). LIMO EEG: A toolbox for hierarchical LInear MOdeling of ElectroEncephaloGraphic data. *Computational Intelligence and Neuroscience, 2011*(3), 1–11. doi:10.1155/2011/831409

Potter, M. C., & Faulconer, B. A. (1975). Time to understand pictures and words. *Nature, 253*(5491), 437–438. doi:10.1038/253437a0

Thorpe, S., Fize, D., & Marlot, C. (1996). Speed of processing in the human visual system. *Nature, 381*(6582), 520–522.

Torralba, A., Oliva, A., Castelhano, M. S., & Henderson, J. M. (2006). Contextual guidance of eye movements and attention in real-world scenes: The role of global features in object search. *Psychological Review, 113*(4), 766–786. doi:10.1037/0033-295X.113.4.766

Underwood, G., & Foulsham, T. (2006). Visual saliency and semantic incongruency influence eye movements when inspecting pictures. *The Quarterly Journal of Experimental Psychology, 59*(11), 1931–1949. doi:10.1080/17470210500416342

Underwood, G., Templeman, E., Lamming, L., & Foulsham, T. (2008). Is attention necessary for object identification? Evidence from eye movements during the inspection of real-world scenes. *Consciousness and Cognition, 17*(1), 159–170. doi:10.1016/j.concog.2006.11.008

Unema, P. J., Pannasch, S., Joos, M., & Velichkovsky, B. M. (2005). Time course of information processing during scene perception: The relationship between saccade amplitude and fixation duration. *Visual Cognition, 12*(3), 473–494.

Võ, M. L.-H., & Henderson, J. M. (2009). Does gravity matter? Effects of semantic and syntactic inconsistencies on the allocation of attention during scene perception. *Journal of Vision, 9*(3), 24. doi:10.1167/9.3.24

Võ, M. L.-H., & Henderson, J. M. (2010). The time course of initial scene processing for eye movement guidance in natural scene search. *Journal of Vision, 10*(3), 1–13. doi:10.1167/10.3.14

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*(9), 1816–1823. doi:10.1177/0956797613476955

Walther, D., & Koch, C. (2006). Modeling attention to salient proto-objects. *Neural Networks, 19*(9), 1395–1407. doi:10.1016/j.neunet.2006.10.001

Wilcox, R. R. (2005). *Introduction to robust estimation and testing*. Waltham: Academic Press.

Wolfe, J. M. (1998). What can 1 million trials tell us about visual search? *Psychological Science, 9*(1), 33–39. doi:10.1111/1467-9280.00006

Wolfe, J. M. (2012). When do I quit? The search termination problem in visual search. In M. D. Dodd & J. H. Flowers (Eds.), *The influence of attention, learning, and motivation on visual search* (pp. 183–208). New York: Springer. doi:10.1007/978-1-4614-4794-8_8

Wolfe, J. M., Võ, M. L.-H., Evans, K., & Greene, M. R. (2011). Visual search in scenes involves selective and nonselective pathways. *Trends in Cognitive Sciences, 15*(2), 77–84. doi:10.1016/j.tics.2010.12.001

Xiao, J., Hays, J., Ehinger, K. A., Oliva, A., & Torralba, A. (2010). *SUN database: Large-scale scene recognition from abbey to zoo* (pp. 3485–3492). Paper presented at the Computer Vision and Pattern Recognition (CVPR), 2010 I.E. Conference on, IEEE. doi:10.1109/CVPR.2010.5539970