Does gravity matter? Effects of semantic and syntactic inconsistencies on the allocation of attention during scene perception

Does gravity matter? Effects of semantic and syntactic inconsistencies on the allocation of attention during scene perception

Melissa L.-H. Võ

John M. Henderson

Psychology Department, University of Edinburgh, UK, & Psychology Department, Ludwig-Maximilians-Universität Munich, Germany

Psychology Department, University of Edinburgh, UK

It has been shown that attention and eye movements during scene perception are preferentially allocated to semantically inconsistent objects compared to their consistent controls. However, there has been a dispute over how early during scene viewing such inconsistencies are detected. In the study presented here, we introduced syntactic object–scene inconsistencies (i.e., floating objects) in addition to semantic inconsistencies to investigate the degree to which they attract attention during scene viewing. In Experiment 1 participants viewed scenes in preparation for a subsequent memory task, while in Experiment 2 participants were instructed to search for target objects. In neither experiment were we able to find evidence for extrafoveal detection of either type of inconsistency. However, upon fixation both semantically and syntactically inconsistent objects led to increased object processing as seen in elevated gaze durations and number of fixations. Interestingly, the semantic inconsistency effect was diminished for floating objects, which suggests an interaction of semantic and syntactic scene processing. This study is the first to provide evidence for the influence of syntactic in addition to semantic object–scene inconsistencies on eye movement behavior during real-world scene viewing.

Keywords: eye movements, semantics versus syntax, object–scene inconsistency, foveal versus extrafoveal processing

Citation: 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, 1–15, http://journalofvision.org/9/3/24/, doi:10.1167/9.3.24.

Introduction

There is ample evidence that a short glimpse of a scene is enough to extract the global meaning—the so-called gist—of a scene (e.g., Castelhano & Henderson, 2008; Oliva & Schyns, 2000; Oliva & Torralba, 2006; Potter, 1975; Thorpe, Fize, & Marlot, 1996). Extraction of gist leads to a set of expectations regarding the scene's composition, e.g., which objects a certain scene should contain or where within the scene such objects should be located. In the study presented here, we compared the effects of violating such expectations on the control of eye movements during scene viewing.

There has been ongoing debate concerning how quickly we can detect and process objects that do not fit the global gist of a scene, and whether initial eye movements can be modulated by the computation of such object–scene inconsistencies. Ever since the influential "octopus in farmyard" study by Loftus and Mackworth (1978) showed that scene inconsistencies can be detected early enough to affect initial eye movements, various research groups have either been able to replicate this finding (e.g., Becker, Pashler, & Lubin, 2007; Bonitz & Gordon, 2008; Underwood & Foulsham, 2006; Underwood, Humphreys, & Cross, 2007; Underwood, Templeman, Lamming, & Foulsham, 2008) or have found evidence that argues against an early impact of scene inconsistencies on eye movement control (e.g., De Graef, Christiaens, & d'Ydewalle, 1990; Gareze & Findlay, 2007; Henderson, Weeks, & Hollingworth, 1999; Rayner, Castelhano, & Yang, 2009). The debate is based on the paradox that the gist of a scene can be perceived within a very short glance, while in the same amount of time only a few objects can be identified (e.g., Castelhano & Henderson, 2005; Henderson & Hollingworth, 1999a, 2003; Tatler, Gilchrist, & Rusted, 2003). The question therefore is whether object–scene inconsistencies can influence eye movement control prior to foveal processing of the inconsistent object, which due to the limitations of the visual acuity in the visual periphery would imply a semantic pop-out of such inconsistencies independent of foveal processing. Thus, finding an effect of object–scene inconsistency on early eye movements during scene viewing would argue for an attraction of attention and gaze without the need of full object identification, while only finding effects upon fixation of an inconsistent object would strengthen the claim that in order to compute the inconsistency between the object and the scene a higher degree

doi: 10.1167/9.3.24 Received July 12, 2008; published March 27, 2009 ISSN 1534-7362

of object identification by means of foveal processing is necessary.

Henderson and colleagues (1999) showed that contrary to the results reported by Loftus and Mackworth (1978), there was no evidence for an effect of semantic inconsistency prior to the fixation of an inconsistent object. Participants viewed a set of line drawings of natural scenes modified from the ones used by De Graef et al. (1990) in preparation for a memory task. The scenes included either a semantically consistent object, e.g., a cocktail glass in a kitchen, or a semantically inconsistent object, e.g., a microscope in the kitchen. Scenes were paired so that the inconsistent object in one scene would serve as the consistent object in another, i.e., the microscope would be the consistent object in a laboratory. The results showed that initial saccades were not controlled by semantic inconsistencies in the visual periphery, but upon fixation the semantic inconsistency of an object affected fixation densities and durations. Inconsistent objects were fixated longer and more often than their consistent counterparts, and viewers tended to return their gaze to inconsistent objects more often than to consistent objects. Even when participants were instructed to actively search for target objects that were either consistent or inconsistent with the scene context, there was no evidence for extrafoveal processing of semantic inconsistencies.

The discrepancy between these findings and the ones reported by Loftus and Mackworth (1978) has been attributed to the differences in stimulus material used. While the scenes in the "octopus in farmyard" study were rather sparse, containing only a few objects displayed in large amounts of empty space, the scenes used by Henderson and colleagues (1999) were derived from photographs and were therefore more cluttered. In sparser scenes, inconsistent objects might more readily "pop out" of the scene than when objects first have to be segregated from their background to allow for a greater degree of processing in the periphery of the visual field.

Similar to the study by Henderson et al. (1999), De Graef and colleagues (1990), Gareze and Findlay (Experiment 5, 2007), and Rayner et al. (2009) were unable to find any evidence of a consistency effect prior to target fixation in line drawings derived from photographs. An object's inconsistency with the scene only showed an effect upon its fixation. Together, these findings argue against an extrafoveal detection of scene inconsistencies that attract early eye movements during scene viewing.

To date, most of the evidence bearing on the effect of object–scene inconsistencies has come from one type of manipulation: the semantic violation of a scene's gist. However, a different way to produce object–scene inconsistencies relates not to an object's semantic fit to the general scene gist, but to its position within the specific structure of scene elements, i.e., the scene syntax. In the early 1980s, Biederman and colleagues (e.g., Biederman, Mezzanote, & Rabinowitz, 1982) investigated the effects of different object–scene inconsistencies including "probability" (objects tend to be found in some categories of scene but not in others) and "support" (objects tend to rest on surfaces). These studies measured object detection using 150-ms scene presentations. One outcome was that both types of inconsistencies equally led to decreased object identification performance. Further, when an object was inconsistent in both support and probability, identification was even further decreased, arguing for the rapid detection of such object–scene inconsistencies. However, since object processing was measured by asking participants post-perceptually whether a certain object was absent or present, response bias and decision uncertainty rather than differences in perceptual sensitivity might have produced the results (Hollingworth & Henderson, 1998; see also Henderson & Hollingworth, 1999b).

Eye movements provide a more unobtrusive, on-line measure of attention allocation and object processing. De Graef et al. (1990) compared the effect of a variety of scene inconsistencies including syntactic violations by measuring first fixation and gaze durations on objects embedded in line drawings of scenes in which participants were instructed to search for non-objects. Each scene contained two objects that were manipulated to create sets of inconsistencies. While first fixation durations are believed to provide a measure that more closely reflects object–identification time (e.g., Friedman, 1979; Henderson et al., 1999), gaze durations mirror later processing stages. When the eye movement data were analyzed for objects that were fixated early versus late during viewing, there was no evidence that contextual information modulated object perception in the early stages of scene viewing. Only later did semantic as well as support violations lead to prolonged first fixation durations on these objects.

Taken together, these prior studies indicate that syntactic as well as semantic violations affect attention allocation during scene viewing. However, Biederman et al.'s (1982) findings of early effects were not based on eye movement data and might therefore have resulted from response biases, while De Graef et al. (1990) never tested the effect of inconsistencies individually but used pairs of different inconsistencies within each scene making it more difficult to interpret the effect of a single manipulation. Moreover, the De Graef et al. effects appeared only later during scene viewing. Finally, both studies used line drawings which might have diminished the effect of syntactic violations due to a lack of depth perception in such reduced scenes (see Becker et al., 2007; Underwood et al., 2007).

The study presented here extends previous work and at the same time aims at setting the stage for resolving the debate by using highly controlled 3D-rendered images of real-world scenes instead of line drawings or photographs, which are either less realistic or more difficult to control for bottom-up saliency. In addition, we directly compared the effects of both semantic and syntactic object–scene inconsistencies on eye movement control during scene viewing. We therefore created both semantic and syntactic

inconsistencies of objects embedded in otherwise consistent scene contexts. Semantic violations of the scene context were created by replacing a semantically plausible object within a scene, e.g., a pot in a kitchen, with an implausible object, e.g., a printer in the kitchen. We operationalized syntactic inconsistencies by violating the local scene structure, i.e., having objects that normally rest on surfaces float above the surface. This resulted in four versions of each scene generated from all possible combinations of semantic and syntactic manipulations (see Figure 1).

A key question was whether object–scene inconsistencies would attract early eye movements as, for example, Underwood and colleagues have reported (e.g., Underwood & Foulsham, 2006; Underwood et al., 2007, 2008), arguing for extrafoveal processing of scene inconsistencies, or whether scene inconsistencies would only exhibit additional processing once an inconsistent object has been fixated (e.g., Gareze & Findlay, 2007; Henderson et al., 1999; Rayner et al., 2009). In particular, we were interested in directly comparing the impact of violations regarding both the semantic and syntactic scene construction on eye movement control during scene viewing. If object identification is a prerequisite for the detection of both semantic and syntactic inconsistencies, no early effects on eye movements are expected and the eyes should not be drawn to the inconsistencies. However, once fixated, the violation of expectations regarding object–scene relations should lead to prolonged allocation of attention in order to resolve the detected anomaly. According to Itti and Baldi (2005), the difference between prior and posterior expectations about the world constitutes "surprise" in a Bayesian framework, which subsequently leads to increased allocation of human attention and gaze to surprising events. If the degree of attention allocation to an inconsistent object represents a function of expectations or the probability of encountering such inconsistencies, floating objects should lead to more and longer fixations than semantically inconsistent objects due to the fact that we are more often exposed to semantically inconsistent than floating objects. You might, for example, have come across a cocktail glass in the lab but have probably not encountered a floating microscope.

In order to investigate these questions, we recorded eye movements while participants either viewed a scene for later recognition (Experiment 1) or while searching for pre-specified target objects (Experiment 2). The second experiment was conducted to rule out the possibility that scene inconsistencies could not exhibit an early effect on

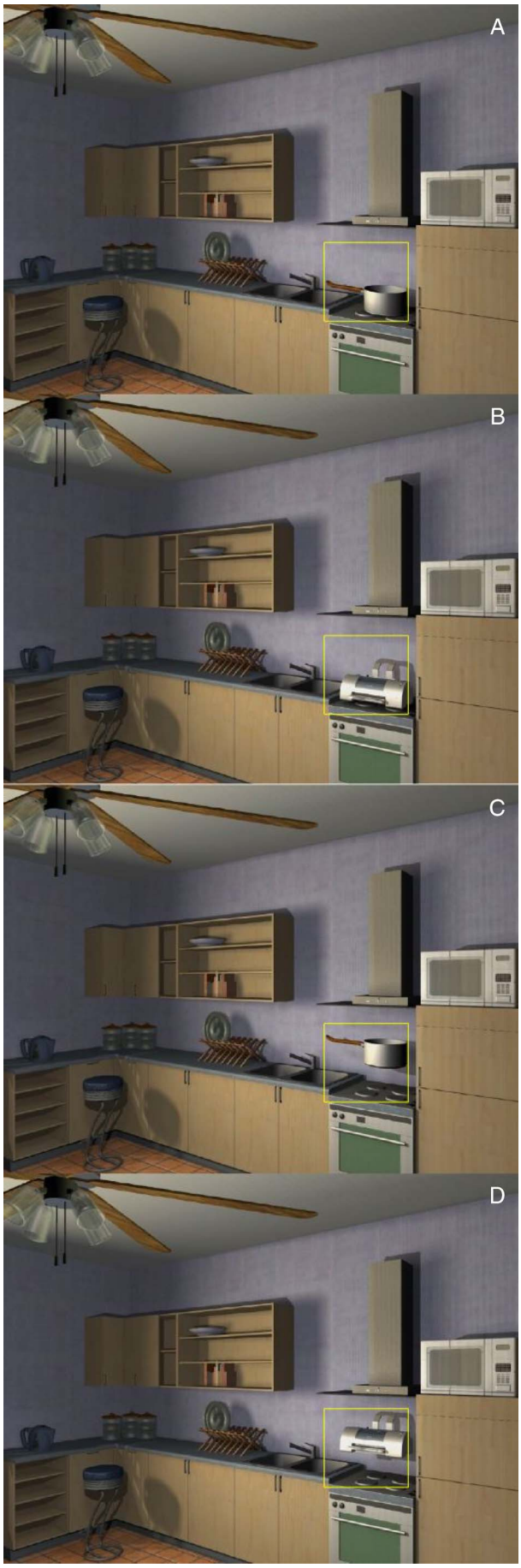

Figure 1. Sample of four versions of a kitchen scene containing (A) a semantically consistent, non-floating object; (B) a semantically inconsistent, non-floating object; (C) a semantically consistent, floating object; or (D) a semantically inconsistent, floating object. Yellow rectangles indicate scoring regions and were not shown to participants.

initial eye movements because participants were not motivated to fixate inconsistent objects quickly due to the unspeeded nature of the memorization task (Henderson et al., 1999).

The task to search for a target as fast as possible should increase the effect of object–scene inconsistency on the attraction of eye movements, since participants would be expected to show increased extrafoveal processing when moving the eyes more quickly in search of the target. The lack of effects of object–scene inconsistencies in either experiment prior to the fixation of the inconsistent object would imply that regardless of the task, object–scene inconsistencies in the periphery of the visual field cannot be processed to a degree sufficient to affect eye movement control.

Experiment 1

Method

Participants

Twenty-four students (21 female) from the University of Edinburgh ranging in age between 18 and 24 years (M = 19.8, SD = 1.83) participated in Experiment 1 for course credit or for 6£/hour. All participants reported normal or corrected-to-normal vision and were unfamiliar with the stimulus material. Two participants had to be replaced due to unstable recording of the eye.

Stimulus material

The stimulus material consisted of 20 3D-rendered images of real-world scenes. The scenes were displayed on a 21-inch computer screen (resolution 1024 × 768 pixel, 140 Hz) subtending visual angles of 25.66 (horizontal) and 19.23 (vertical) at a viewing distance of 90 cm. Each scene was manipulated so that it conformed to one of the four experimental conditions: In the consistent-surface condition, the object of interest was semantically consistent with the scene context and rested on a surface (e.g., a pot on a kitchen stove), whereas in the consistent-float condition the same object was displayed as hovering above the surface in mid-air. In the inconsistent-surface and inconsistent-float conditions, the semantically consistent object was replaced by an inconsistent object (e.g., a printer on a kitchen stove) resting on a surface or hovering in mid-air, respectively. Figure 1 displays a sample scene in its four versions.

Scenes were paired so that the semantically inconsistent object of one scene was consistent in its paired scene (e.g., a printer on an office desk). Semantically consistent and inconsistent objects were matched for size and were placed in the same position within each scene away from the center where the initial fixation was to be made. Furthermore, scenes were processed using the Itti and Koch (2000) MatLab Saliency Toolbox to determine the most salient regions according to low-level saliency calculations of brightness, color, contrast, and edge orientation. The rank order of saliency peaks—with rank 1 assigned to the most salient region of the scene—was used to ensure that consistent and inconsistent objects did not differ in their mean low-level saliency (M = 8.45, SD = 3.09 vs. M = 8.9, SD = 2.38, $p > .05$, respectively).

Apparatus

Eye movements were recorded with an EyeLink1000 tower system (SR Research, Canada), which tracks with a resolution of .01° visual angle at a sampling rate of 1000 Hz. The position of the right eye was tracked while viewing was binocular. Experimental sessions were carried out on a computer running OS Windows XP. Stimulus presentation and response recording were controlled by Experimental Builder (SR, Research, Canada).

Procedure

Each participant received written instructions before being seated in front of the presentation screen. Participants were informed that they would be shown a series of scenes that they had to memorize for a later memory test.

At the beginning of the experiment, the eye tracker was calibrated for each participant using 9-point calibration and validation. The participant's viewing position was fixed with a chin and forehead rest. Each trial sequence was preceded by a fixation check, i.e., in order to initiate the next trial, the participants had to fixate a cross centered on the screen for 200 ms. When the fixation check was deemed successful, the fixation cross was replaced by the presentation of a scene for 15 s during which the participant inspected the scene freely in preparation for a memory task. After an inter-trial interval of 1 s, the next trial followed. Two practice trials at the beginning of the experiment allowed participants to become accustomed to the experimental set-up. The experiment lasted about 15 minutes. Subsequently, an off-line memory test was administered without recording eye movements. Since we were only interested in the eye movement data during scene memorizing, the data from the memory test will not be reported here.

Eye movement data analysis

The interest area for each target object was defined as the rectangular box that was large enough to encompass the consistent and inconsistent target objects when located on a surface as well as when floating (see Figure 1). Thus, the scoring regions were the same for all conditions to allow for better comparison. Fixation durations of less than 90 ms and more than 1000 ms were excluded as outliers. Raw data were subsequently filtered using SR

Research Data Viewer and then submitted to an analysis of variance (ANOVA) with semantic consistency (consistent vs. inconsistent) and syntactic consistency (surface vs. float) as within-subject factors.

Results

A set of measures was calculated to analyze viewers' eye movement patterns as a function of both the semantic and syntactic consistency manipulations. We have divided these measures into those that mirror extrafoveal processing of inconsistencies on the one hand and foveal processing of inconsistencies on the other.

Extrafoveal processing of scene inconsistencies

The main aim of the current study was to investigate whether initial eye movements during scene viewing would be modulated by the processing of peripheral scene inconsistencies. To investigate whether semantic as well as syntactic inconsistencies affect eye movements prior to their fixation, seven measures were examined (see Table 1): initial saccade latency, probability of correct initial saccade, probability of immediate target fixation, latency to first target fixation, number of fixations to first target fixation, and incoming saccade amplitude.

Initial saccade latency

Initial saccade latency was measured from scene onset until the initiation of the first saccade and averaged 683 ms across all conditions. There was neither an effect of semantic, $F(1,23) = 1.89$, $p > .05$, nor of syntactic consistency, $F < 1$, and no interaction, $F = 1.48$, $p > .05$.

Probability of correct initial saccade direction

The probability of correct initial saccade direction was defined as the percentage of initial saccades that were directed toward that half of the scene (left vs. right) that contained the target object and averaged 56.88%. There was neither an effect of semantic, $F(1,23) = 1.25$, $p > .05$, nor an effect of syntactic consistency, $F < 1$, and no interaction, $F < 1$.

Latency to first target fixation

Latency was measured from scene onset until the first fixation of the target object and averaged 3966 ms across all conditions. There was neither an effect of semantic, $F(1,23) = 1.15$, $p > .05$, nor an effect of syntactic consistency, $F < 1$, and no interaction, $F < 1$.

Number of fixations to first target fixation

This measure was defined as the discrete number of fixations until the target object was first fixated. The values include both the initial fixation on the scene and the first fixation on the target object. On average, participants performed 10.83 fixations to the first fixation of the target object. There was neither an effect of semantic nor an effect of syntactic consistency, and no interaction, all Fs < 1.

Incoming saccade amplitude

The amplitude of the saccade that first entered the target region was designated incoming saccade amplitude and averaged 5.53 degrees visual angle. There was no effect of semantic consistency, $F(1,23) = 1.24$, $p > .05$, no effect of syntactic consistency, $F < 1$, and no interaction, $F < 1$. Taken together, none of these measures provided evidence that extrafoveal processing of either semantic or syntactic inconsistencies could draw the eyes to peripheral scene regions.

Additionally, we analyzed the probability of immediate target fixation defined as the percentage of trials in which the initial saccade landed on the target object. There was neither an effect of semantic, $F(1,23) = 3.59$, $p > .05$, nor an effect of syntactic manipulation, $F(1,23) = 2.66$, $p > .05$, and no interaction, $F < 1$. We also analyzed the cumulative probability of target fixation after the second saccade and also found no effects, all Fs < 1. The probabilities of target fixation as a function of ordinal fixation number can be seen in Figures 2 and 3. There is no indication of an early effect of either semantic or syntactic manipulations on initial eye movements.

	Semantic			Syntax		
Measures	Consistent	Inconsistent	*F*	Surface	Float	*F*
Initial saccade latency (ms)	673 [40]	693 [32]	1.80	675 [38]	689 [34]	<1
Probability of correct initial saccade direction (%)	59.59 [3.65]	54.17 [4.82]	1.25	57.50 [4.32]	57.09 [4.15]	<1
Probability of immediate target fixation (%)	11.70 [3.78]	7.04 [2.62]	3.59	10.63 [3.89]	7.88 [2.50]	2.66
Latency to first target fixation (ms)	3808 [345]	4124 [334]	<1	4034 [365]	3898 [313]	<1
Number of fixations till target fixation	10.40 [0.98]	11.27 [0.94]	<1	11.05 [1.05]	10.61 [0.86]	<1
Number of fixations to target fixation	10.40 [0.98]	11.27 [0.94]	<1	11.05 [1.05]	10.61 [0.86]	<1
Incoming saccade amplitude in degree visual angle	5.39 [0.27]	5.67 [0.38]	1.24	5.45 [0.28]	5.61 [0.37]	<1

Table 1. Summary of mean values [standard errors] for dependent variables in Experiment 1 reflecting extrafoveal processing as a function of semantic (consistent vs. inconsistent) and syntax (surface vs. float) manipulations. Dependent variables were initial saccade latency, probability of correct initial saccade direction, probability of immediate target fixation, latency to first target fixation, number of fixations to target fixation, and incoming saccade amplitude. *Note*: *$p < .05$; **$p < .01$.

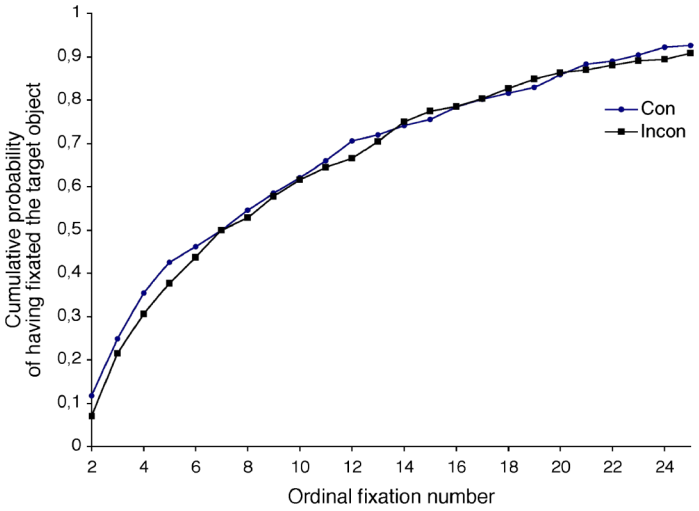

Figure 2. Cumulative probability of having fixated the target object as a function of the ordinal fixation number and semantic consistency (semantically consistent = Con, semantically inconsistent = Incon) in Experiment 1. Note that first fixations on the display were excluded because these were located on the initial fixation cross.

Foveal processing of scene inconsistencies

To investigate whether the semantic or syntactic manipulations affected object processing once the object was fixated, we calculated five additional measures that index the degree of attention allocated to the target objects. The measures were total fixation duration and fixation count, first-pass gaze duration and fixation count, and first fixation duration (see Table 2).

Total fixation duration

The total fixation duration was defined as the sum of all fixation durations on the target region from scene onset until scene offset. Across all conditions, the mean total fixation duration was 1760 ms. There was a main effect of semantic consistency, $F(1,23) = 6.36$, $p < .05$, in that semantically inconsistent objects were fixated for a longer amount of time than objects that were consistent with the semantics of the scene. In addition, we observed a strong effect of the syntactic manipulation, $F(1,23) = 42.43$, $p < .01$, with floating objects looked at longer than objects resting on surfaces. The interaction failed to reach significance, $F(1,23) = 1.94$, $p > .05$.

Total fixation count

Total fixation count was defined as the sum of all fixations located in the target region from scene onset until scene offset and averaged 5.76 fixations. Similar to the total fixation duration, we observed main effects for both the semantic, $F(1,23) = 7.07$, $p = .01$, and the syntactic manipulation, $F(1,23) = 29.91$, $p < .01$, while the interaction was not significant, $F < 1$. Semantically inconsistent as well as floating objects led to a greater number of fixations than semantically consistent objects or objects resting on a surface.

First-pass gaze duration

To investigate the effect of inconsistency on the initial encoding of objects, we calculated the first-pass gaze duration, which was defined as the sum of all fixation durations from the first entry of the eyes into the target region until their first exit. It has been shown that first-pass gaze duration increases when processing semantic inconsistencies (e.g., De Graef et al., 1990; Henderson et al., 1999; Loftus & Mackworth, 1978). On average, participants spent 692 ms on the target before leaving the target region for the first time. As with the total fixation duration, we found effects for both the semantic, $F(1,23) = 8.24$, $p < .01$, and the syntactic inconsistency, $F(1,23) = 12.26$, $p < .01$. In addition, there was a significant interaction of factors, $F(1,23) = 8.64$, $p < .01$. As can be seen in Figure 4, the interaction was characterized by a strong effect of semantic inconsistency for objects resting on surfaces, $t(23) = 4.06$, $p < .01$, while this effect was eliminated for floating objects, $t(23) < 1$.

First-pass gaze count

The first-pass gaze fixation count was defined as the number of fixations from the first entry of the eyes to the target region until their first exit. Similar to the first-pass gaze duration, we observed significant main effects for semantic inconsistency, $F(1,23) = 6.10$, $p < .05$, and syntactic inconsistency, $F(1,23) = 10.87$, $p < .01$, as well as a significant interaction, $F(1,23) = 5.69$, $p < .05$. Participants fixated semantically as well as syntactically inconsistent objects more often than consistent objects and objects resting on surfaces. While there was a significant effect of semantic inconsistency for objects on surfaces, $t(23) = 3.75$, $p < .01$, this effect disappeared for floating objects, $t(23) < 1$.

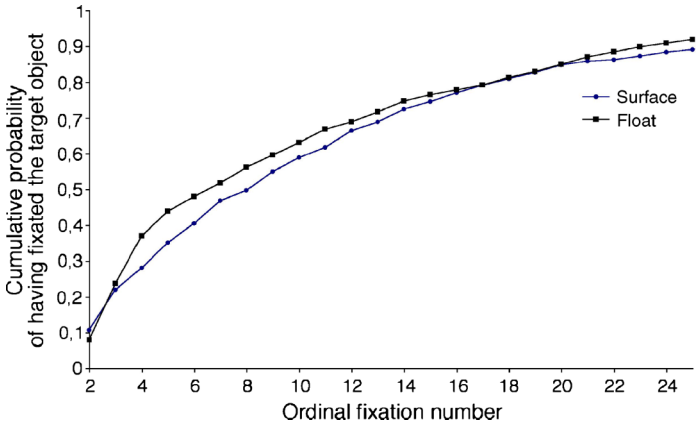

Figure 3. Cumulative probability of having fixated the target object as a function of the ordinal fixation number and syntactic manipulation (Surface, Float) in Experiment 1.

	Semantic			Syntax		
Measures	Consistent	Inconsistent	*F*	Surface	Float	*F*
Total fixation duration (ms)	1633 [87]	1887 [103]	6.36*	1489 [83]	2030 [108]	42.43**
Total fixation count	5.36 [0.32]	6.16 [0.37]	7.07**	5.00 [0.29]	6.52 [0.40]	29.91**
First-pass gaze duration (ms)	586 [56]	798 [77]	8.24*	577 [60]	806 [73]	12.26*
First-pass gaze count	1.97 [0.18]	2.53 [0.21]	6.10*	1.91 [0.16]	2.59 [0.23]	10.87**
First fixation duration (ms)	280 [14]	293 [14]	<1	268 [12]	305 [17]	10.99**

Table 2. Summary of mean values [standard errors] of Experiment 1 regarding dependent variables on foveal processing as a function of semantic (consistent vs. inconsistent) and syntax (surface vs. float) including total gaze duration and gaze count, first-pass gaze duration and gaze count, and first fixation duration. *Note*: *$p < .05$; **$p < .01$.

First fixation duration

As another indicator of initial object encoding, we analyzed the first fixation duration defined as the duration of just the initial fixation made on the target object. First fixation durations are believed to provide a measure that more directly reflects object-identification time (e.g., Friedman, 1979; Henderson et al., 1999). The average first fixation duration amounted to 286 ms. While there was no significant main effect of semantic consistency, $F < 1$, and no significant interaction, $F(1,23) = 1.19$, $p > .05$, we found a significant main effect of the syntactic manipulation, $F(1,23) = 10.99$, $p < .01$, in that the first fixation duration was increased for floating objects compared to objects resting on surfaces. This result suggests that syntactically inconsistent objects might require a greater degree of object processing during initial encoding.

In sum, the data show strong effects of both the semantic and syntactic consistency manipulation once the target object has been fixated. Further, it seems that the effect of semantic inconsistency is weakened when the target object is already syntactically inconsistent.

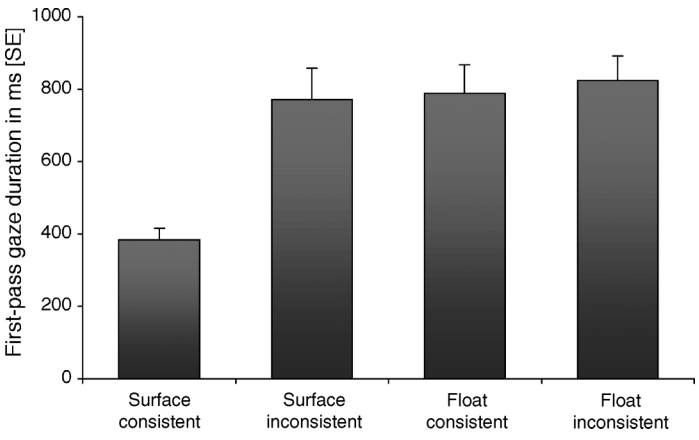

Figure 4. Mean first-pass gaze durations [standard errors] for Experiment 1 as a function of semantic (consistent versus inconsistent) and syntax (surface versus float) manipulations.

Discussion

The aim of Experiment 1 was to investigate whether scene inconsistencies would attract early eye movements prior to the fixation of inconsistent objects when participants were asked to view a scene for a later memory test. There was no evidence that initial eye movements were drawn to objects that either violated expectations of scene semantics or syntax. Object–scene inconsistencies were neither fixated earlier during scene viewing nor from a greater distance than their consistent counterparts, arguing against an extrafoveal processing of scene inconsistencies. This largely replicates the findings by Henderson and colleagues (1999), who also found no early effects of semantic inconsistencies prior to their fixation using line drawings of naturalistic scenes. However, upon fixation, objects that did not fit the semantic context of the scene attracted a higher degree of attention than consistent objects, as seen in a greater amount of fixations and therefore longer gaze durations. Again, this is in line with findings of increased object processing for inconsistent objects once fixated (e.g., De Graef et al., 1990; Gareze & Findlay, 2007; Henderson et al., 1999; Hollingworth, Williams, & Henderson, 2001; Loftus & Mackworth, 1978; Rayner et al., 2009; Underwood & Foulsham, 2006; Underwood et al., 2007, 2008).

Additionally, we found that objects that violated expectations of their syntactic properties, i.e., were floating, also resulted in increased processing compared to objects resting on surfaces. Especially when first encountering an object, the effects of the semantic and syntactic manipulations interacted in such a way that while non-floating objects showed clear modulation according to their semantic fit to the scene context, this semantic inconsistency effect was eliminated when the object was floating. A semantically consistent but floating object held gaze to the same degree as an object that was resting on a surface but incongruent with the scene semantics. A double inconsistency, i.e., a semantically inconsistent and floating object, did not yield more processing time than each individual inconsistency. Thus, it seems that once an object violated syntactic regularities, it no longer mattered whether it fit the overall gist of the

scene or not. The syntactic manipulation also affected the first fixation duration on an object, which was prolonged for floating objects, whereas the semantic manipulation did not affect initial encoding time. This implies that initial encoding of syntactically inconsistent objects required more time than the encoding of syntactically consistent objects.

A possible explanation for the strong impact of the syntactic manipulation is the probability of encountering such an inconsistency in everyday life. While we do come across misplaced objects from time to time, we rarely encounter floating objects. The stronger and more restricted the scene priors are, the greater the effect of their violations seems to be. Contrary to this view, Biederman and colleagues (1982) did not find stronger effects for the support compared to the semantic manipulation in their tachistoscopic object detection paradigm. Besides the lack of control for response biases, another reason for the lack of stronger effects of additional processing for floating objects could have been the task itself. While participants in the Biederman et al. study had to decide whether a certain object was present or absent, participants in our study were asked to memorize objects for later recognition. Floating objects can enable easier figure-ground segmentation due to their position in the scene, e.g., in the Biederman et al. study the couch floating in the sky was in an uncluttered scene region. This is particularly true in line drawings as used in the Biederman et al. study and could have increased performance in the syntactic violation condition, counteracting the detrimental effect of having to resolve the syntactic violation. As a result, semantic and syntactic inconsistencies yielded similar detection performance.

In sum, the data of Experiment 1 did not lend support to the claim that extrafoveal processing of object inconsistencies in scenes can guide eye movements to these inconsistent objects. Rather, our data clearly speak for a limited region around the fovea in which semantic as well as syntactic inconsistencies can be processed to a degree that attention allocation and eye movements are modulated.

Experiment 2

The data of Experiment 1 seem to imply that scene inconsistencies cannot be processed when they are outside of foveal viewing, but they exhibit strong effects on attention allocation and eye movement control once fixated. According to Henderson and colleagues (1999), an alternative interpretation of the data from Experiment 1 could be that the scene inconsistencies could not exhibit an early effect on initial eye movements because participants were not motivated to fixate inconsistent objects quickly due to the unspeeded nature of the memorization task. In contrast to the 15-s viewing time in our study, participants in the Loftus and Mackworth (1978) study only had 4 s to inspect a scene, which could have increased the need for extrafoveal processing.

To address this possibility, we conducted a second experiment using the same experimental design. Instead of allowing participants to view each scene for 15 s, we asked them to search for pre-specified target objects as quickly as possible. If inconsistency is processes extrafoveally, the additional motivation to quickly find the target object should increase the effect of object–scene inconsistency on the attraction of eye movements, such that semantically and syntactically inconsistent objects should be fixated earlier and with a greater incoming saccade amplitude than their consistent counterparts.

Method

Participants

Twenty-four students (16 female) from the University of Edinburgh ranging in age between 19 and 26 years (M = 21.8, SD = 2.5) participated in Experiment 2 for course credit or for 6£/hour. All participants reported normal or corrected-to-normal vision, and none had taken part in Experiment 1. One participant had to be replaced due to misunderstandings of target words.

Stimulus material

The search scenes were identical to the scenes used in Experiment 1. All target objects of Experiment 1 served as search targets in Experiment 2, which were pre-specified by target words preceding each search scene. The 20 target words were displayed in uppercase black Arial typeset centered on a gray background (RGB: 51, 51, 51). Target words were chosen to be comprehensible and unambiguous in indicating the target object.

Apparatus

The apparatus was identical to the one used in Experiment 1. A joy-pad was added to collect reaction time data.

Procedure

As in Experiment 1, each participant received written instructions before being seated in front of the presentation screen. Participants were informed that they would be presented with a series of scenes, each of which contained a pre-specified target object that they had to find as quickly as possible. Once found, they were to press a button on a joy-pad.

At the beginning of the experiment, the eye tracker was calibrated for each participant. Each trial sequence was preceded by a fixation check. When the fixation check

was deemed successful, the fixation cross was replaced by the presentation of a word (2000 ms) indicating the identity of the target object. An additional fixation cross followed (500 ms) to make sure that after reading the target word the eyes were repositioned at the center of the screen when the search scene appeared. Participants were instructed to search the scene for the target object as fast as possible and to indicate detection of the target by holding fixation on the object and pressing a joy-pad button. The search scene was displayed for 15 s or until button press. After an inter-trial interval of 1 s, the next trial followed. Two practice trials were administered at the beginning of the experiment. The experiment lasted a total of about 10 minutes. Again, an off-line memory test without recording eye movements followed. The data of the memory test will not be reported here.

Results

In the following analyses, trials were excluded that did not result in successful target search or were subject to unstable tracking of the eye (3.94%). As in Experiment 1, raw data were preprocessed by the SR Research Data Viewer and then submitted to an analysis of variance (ANOVA) with semantic consistency (consistent vs. inconsistent) and syntactic consistency (surface vs. float) as within-subject factors.

Eye movement data recorded after fixation of the target object were sparse and truncated because the scene disappeared once participants had pushed the button to indicate that they had found the target object. Due to this artificial termination of fixations by the button press, we only report extrafoveal processing measures for Experiment 2.

Extrafoveal processing of scene inconsistencies

In addition to the dependent variables reported in Experiment 1, reaction times are reported since participants had to press a button as soon as the target object had been found (see Table 3).

Initial saccade latency

Initial saccade latency for search averaged 464 ms across all conditions. There was neither an effect of semantic nor syntactic consistency, and no interaction, all $Fs < 1$.

Probability of correct initial saccade direction

The probability of correct initial saccade direction was defined as the percentage of initial saccades that were directed toward that half of the scene (left vs. right) that contained the target object and averaged 60.47%. There was neither an effect of semantic nor syntactic consistency, and no interaction, all $Fs < 1$.

Reaction time

RT was defined as the time elapsed from scene onset until button press and averaged 2015 ms across all conditions. There was neither an effect of semantic, $F < 1$, nor an effect of syntactic consistency, $F(1,23) = 2.71$, $p > .05$, and no interaction, $F < 1$.

Latency to first target fixation

The time to the first fixation of a target object was much shorter than in Experiment 1, with an average latency of 1282 ms. However, as in Experiment 1, there were no main effects for either the semantic or syntactic manipulation, $F(1,23) = 1.11$, $p > .05$ and $F < 1$, respectively, and no interaction, $F(1,23) = 1.37$, $p > .05$.

Number of fixations to target fixation

The average number of fixations needed to find the target object was 3.90 fixations. Neither the main effects nor the interaction reached significance, all Fs < 1.

Incoming saccade amplitude

The amplitude of the saccade entering the target region for the first time averaged 6.25° visual angle across all conditions. Again, there was no effect of semantic inconsistency, $F(1,23) = 1.34$, $p > .05$, no effect of syntactic inconsistency, $F(1,23) = 1.38$, $p > .05$, and no interaction, $F(1,23) = 1.56$, $p > .05$.

	Semantic			Syntax		
Measures	Consistent	Inconsistent	*F*	Surface	Float	*F*
Initial saccade latency (ms)	459 [18]	470 [21]	<1	459 [21]	468 [18]	<1
Probability of correct initial saccade direction (%)	62.36 [4.54]	58.58 [4.85]	<1	61.08 [4.79]	59.86 [4.60]	<1
Probability of immediate target fixation (%)	17.02 [3.83]	19.79 [4.19]	<1	16.77 [3.98]	20.04 [4.03]	<1
Reaction time (ms)	1964 [146]	2067 [121]	2.71	1936 [122]	2094 [144]	<1
Latency to first target fixation (ms)	1237 [104]	1327 [104]	1.13	1248 [89]	1317 [119]	<1
Number of fixations till target fixation	3.83 [0.21]	3.98 [0.23]	<1	3.97 [0.23]	3.84 [0.02]	<1
Number of fixations to target fixation	3.83 [0.21]	3.98 [0.23]	<1	3.97 [0.23]	3.84 [0.02]	<1
Incoming saccade amplitude in degree visual angle	6.02 [0.37]	6.45 [0.39]	1.34	6.45 [0.43]	6.02 [0.34]	1.38

Table 3. Summary of mean values [standard errors] of Experiment 2 for extrafoveal processing as a function of semantic (consistent vs. inconsistent) and syntactic conditions (surface vs. float). Dependent variables were initial saccade latency, probability of correct initial saccade direction, probability of immediate target fixation, reaction time, latency to first target fixation, number of fixations to target fixation, and incoming saccade amplitude. *Note*: *$p < .05$; **$p < .01$.

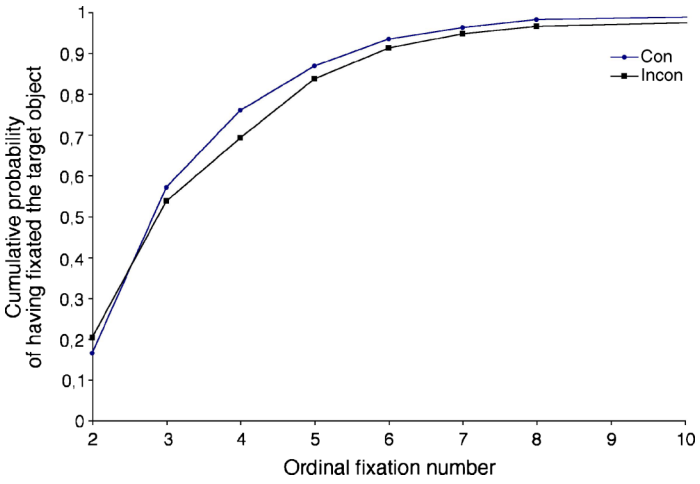

Figure 5. Cumulative probability of having fixated the target object as a function of the ordinal number and semantic consistency (semantically consistent = Con, semantically inconsistent = Incon) in Experiment 2.

Probability of immediate target fixation

As in Experiment 1, there was no effect of either the semantic or syntactic manipulation on the probability that the first saccade would be directed at the target object, nor an interaction, Fs < 1. The same was true for the cumulative probability of the second saccade landing on the target object, Fs < 1. The probabilities of target fixations as a function of ordinal fixation number can be seen in Figures 5 and 6.

These data suggest that despite the instruction to search for target objects as quickly as possible, eye movements were not modulated by extrafoveal processing of scene inconsistencies.

Discussion

The purpose of Experiment 2 was to test whether the lack of extrafoveal effects of scene inconsistencies in Experiment 1 was due to the task, which did not motivate participants to actively move their eyes quickly to objects displayed in the scene. In order to check whether task instruction really had an effect on eye movement behavior when inspecting the scenes, we compared mean gaze durations and saccade amplitudes for Experiment 1 versus Experiment 2. We found that the average time the eyes spent in each fixation on the scene—excluding the time spent fixating target objects—differed significantly between Experiments 1 and 2, $t(47) = 10.39$, $p < .01$, with longer mean fixation durations for eye movements in the memorization task (M = 322 ms) than in the search task (M = 268 ms). Mean saccade amplitude—excluding those saccades that originated from or entered the target object—was shorter during memorization ($M = 3.40°$ visual angle) than during search ($M = 4.36°$ visual angle), $t(47) = 12.88$, $p < .01$. Thus, the task significantly affected eye movement behavior when viewing the scenes.

With the task to search for target objects in Experiment 2, participants might have tried to process more information from the periphery of their visual field, which could have increased the detection of inconsistencies outside the focus of a current fixation. However, despite the task to actively search for target objects, there was again no evidence that scene inconsistencies attracted eye movements. Neither semantically nor syntactically inconsistent target objects were found faster than their consistent controls. The first saccade was neither initiated earlier nor more often directed toward the target for inconsistent objects. Also, the amplitude of the saccade entering the target region did not vary as a function of the semantic or syntactic manipulation, arguing against the view that scene inconsistencies attract attention prior to their fixation. This is in line with findings by Henderson and colleagues (1999), who also found no evidence for extrafoveal processing of semantic inconsistencies despite engaging participants in an active search task.

Rather than finding search benefits for inconsistent objects, as would be expected if inconsistency captures attention, previous studies have instead found search benefits for consistent objects (e.g., Eckstein, Drescher, & Shimozaki, 2006; Henderson et al., 1999; Neider & Zelinsky, 2006). Neider and Zelinsky (2006), for example, had participants search for scene-constrained or scene-unconstrained targets and found that targets that were constrained by the scene context were found faster and with fewer eye movements. Also Eckstein and colleagues (2006) found that endpoints of initial saccades were closer to the target region when target objects were located in

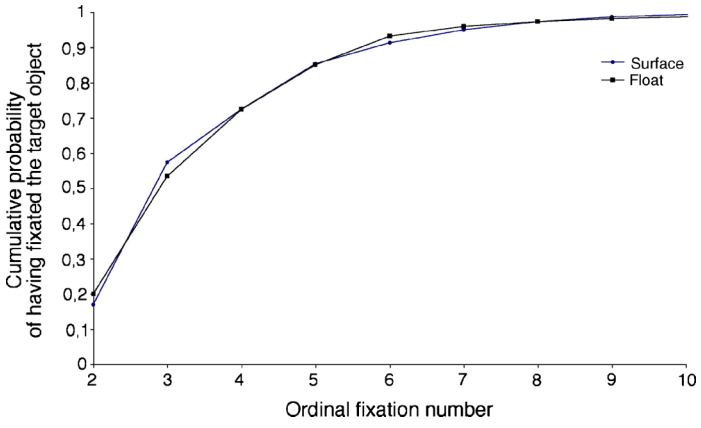

Figure 6. Cumulative probability of having fixated the target object as a function of the ordinal number and syntactic manipulation (Surface, Float) in Experiment 2.

expected rather than unexpected locations. Torralba, Oliva, Castelhano, and Henderson (2006) observed that participants limited their search to scene regions likely to contain a search target even when other highly salient regions were present elsewhere. In our study, we did not find search benefits for consistent objects as opposed to inconsistent objects either early or late in scene viewing. This could be due to the fact that contrary to these other studies, in our study the semantically consistent objects were not specifically chosen to be contextually constrained to a certain expected location within a scene, e.g., plates in a kitchen. The difference between expected and unexpected objects in our study referred to expecting a certain object within a scene in general and was therefore not a matter of the objects' spatial constraints.

Furthermore, the syntactic manipulation—even though altering the expected spatial location of an object—did not involve the displacement of an object to a distant, improbable region of the scene but simply had objects hover above surfaces. Thus, such an inconsistency is less apt to modulate search performance that the manipulations used by Eckstein and colleagues (2006) and Neider and Zelinsky (2006).

General discussion

The departure point of the present study was to shed new light on the discussion regarding the processing of object–scene inconsistencies during scene viewing. Therefore, we created 3D-rendered images of naturalistic scenes with a high degree of realism, while allowing for highly controlled manipulations of objects within a scene. A key question was whether inconsistent objects in the periphery of the visual field would be able to control initial eye movements prior to their fixation. In addition, we were interested in the direct comparison of two different scene inconsistencies: violations of scene semantics on the one hand and violations of scene syntax on the other. In the following we will discuss these issues in greater detail.

Foveal versus extrafoveal processing of scene inconsistencies

Is there "semantic pop-out" effect in scene perception? That is, can an object–scene inconsistency be detected before the object has been foveally processed and identified? According to Loftus and Mackworth (1978), an object that does not fit the semantics of a scene exhibits control over eye movements very early during scene viewing, affecting initial eye movements prior to fixation of the inconsistent object. They not only found that inconsistent objects were fixated longer, but also earlier (in fact, immediately) when inspecting a scene for later recognition. Also, saccades entering the region of an inconsistent object were longer in amplitude than when entering the region of a consistent object, thus arguing in favor of an extrafoveal processing of inconsistencies in the visual periphery.

However, this interpretation has not gone unchallenged, with a number of studies showing that initial eye movements are not influenced by the processing of scene inconsistencies prior to their fixation (e.g., De Graef et al., 1990; Gareze & Findlay, 2007; Henderson et al., 1999; Rayner et al., 2009). Recently, Underwood and colleagues reinstated the claim that full identification of objects is not necessary for object–scene inconsistencies in the visual periphery to attract eye movements prior to foveal processing (e.g., Underwood & Foulsham, 2006; Underwood et al., 2007, 2008). They used photographs instead of line drawings, manipulating both visual bottom-up saliency and the consistency of objects embedded in scenes, and found that semantically inconsistent objects were fixated earlier than their consistent counterparts.

In contrast to the early findings of Loftus and Mackworth (1978) as well as recent findings by Underwood and colleagues (e.g., Underwood & Foulsham, 2006; Underwood et al., 2007, 2008), we have not found early effects of scene inconsistencies, either when participants viewed a scene for later recognition (Experiment 1), or when they were instructed to actively search for target objects (Experiment 2). Thus, we argue that in complex naturalistic scenes, foveal processing of object–scene inconsistencies is necessary in order to influence the allocation of attention and exhibit control over initial eye movements.

How can these conflicting results be explained? The source of differences might lie in a combination of differences in presentation times, tasks, and stimulus material used across studies.

For example, Neider and Zelinsky (2006) as well as Eckstein and colleagues (2006) observed effects of object–scene inconsistencies on the initial saccade, arguing that scene-based guidance is active early during scene viewing. However, neither study manipulated semantic object–scene inconsistencies, but rather had semantically consistent objects placed in highly improbable locations within the scene, e.g., a chimney next to a tree (Eckstein et al., 2006) or a tank in the sky (Neider & Zelinsky, 2006). Furthermore, the misplacement of objects in these studies cannot be simply compared to the syntactic inconsistencies used in our study, since the violation of object locations in our study involved objects placed in probable locations but with the oddity of having them float above surfaces on which they would normally be found. Thus, their findings of effects on initial saccades are not due to the processing of object–scene inconsistencies in the visual periphery but due to misleading contextual guidance (Torralba et al., 2006).

Further, Loftus and Mackworth (1978) as well as Underwood and colleagues (2007) used shorter presentation times (4 s and 5 s, respectively) in the memorization

task than Henderson and colleagues (1999) or we did (both 15 s). Shorter presentation times might have motivated participants to move their eyes more quickly, thus widening the scope of attention allocation in order to process more information. However, the lack of evidence for extrafoveal processing of scene inconsistencies when participants were instructed to search for target objects in Experiment 2 of the present study (see also Henderson et al., 1999) argues against the idea that the different results across studies are solely due to the use of different presentation times or tasks.

Bonitz and Gordon (2008) found that semantically inconsistent objects were fixated earlier than their consistent counterparts. However, they did not report effects on the incoming saccade length or initial eye movements, which would have implied semantic pop-out of such inconsistencies. Also, while inconsistent objects were fixated earlier than consistent objects, this effect only occurred after several eye movements on the scene which leaves the question open if the earlier fixation of inconsistent objects might have been due to not direct, but very proximate fixations of inconsistent objects. In line with this explanation, Gareze and Findlay (2007) have found effects of eccentricity on the detection of semantic inconsistencies. In order to claim that semantically inconsistent objects not only hold, but readily attract attention without prior fixation, evidence for the modulation of especially initial eye-movements is necessary.

The differences in stimulus material might have been a greater source of variance across studies. As has been discussed earlier, the scenes Loftus and Mackworth (1978) used were rather sparse and might have increased the impact of extrafoveal processing since there were only a few easily identifiable objects present. Also, inconsistent objects might have been more visually conspicuous attracting eye movements by means of low-level visual salience, since this was not controlled. In contrast, De Graef et al. (1990), Gareze and Findlay (2007), and Henderson et al. (1999) used more complex line drawings, which could have decreased the effect of extrafoveal processing. Underwood and colleagues (2007), on the other hand, used color photographs that were edited post hoc, which might have introduced artificial low-level conspicuousness without the Itti and Koch (2000) algorithm detecting it, but with an effect on human observers. For example, some of the inconsistent objects in Experiment 2 of the Underwood et al. (2007) study seem visually odd due to inappropriate shadows and other artifacts caused by the cut-out and pasting process. To be more exact, the shadow of the congruent object in Figure 2a is still visible in its complementary scenes where the congruent object was replaced by incongruent or bizarre objects (Figures 2b and 2c).

In our study, we used 3D-rendered scenes that allowed for object–scene manipulations without the need to edit the stimulus material post hoc. Additionally, the scenes displayed a high degree of photorealism regarding colors, texture, and illumination of both the scenes and the embedded objects. Thus, consistent as well as inconsistent objects alike blended into the scenes.

Comparing the mean saliency rank values of objects in our study (mean rank about 8.5)—calculated using the Itti and Koch (2000) algorithms as the rank of the scene region that contained the target object in relation to the rest of the scene—with the mean rank values of objects in the scenes Underwood et al. (2007) used (mean rank about 3), reveals that the objects of interest in our scenes ranged relatively low in visual salience compared to other scene regions, whereas objects in the study by Underwood and colleagues ranged higher in visual salience within the scene context. According to Underwood and Foulsham (2006), objects of low salience values should especially exhibit effects of semantic inconsistency prior to and upon fixation of the object. Thus, on this view, our stimuli should have been more apt to produce effects of extrafoveal processing than those used by Underwood and Foulsham. However, this was not the case.

The reason might lie in the definition of high and low salience. Whereas Underwood and Foulsham defined high and low salience objects by comparing saliency rank values between two objects of a scene, it might be more important to relate the visual salience of an object to the entire scene in which it is embedded. In our case, a mean rank value of about eight implies that seven other regions in the scene were visually more conspicuous, while in the study by Underwood et al. (2007) on average only two other regions were more conspicuous. Thus, at least during free scene viewing, the effect of scene inconsistencies might depend on the relative visual salience of the inconsistent object. Specifically, there might be a greater impact of extrafoveal scene inconsistencies when there are not as many higher salient regions in the scene attracting gaze on the basis of low-level features. Follow-up studies explicitly manipulating the saliency ranks of inconsistent objects in relation to other parts of the scene might be able to shed more light on this possible source of variance across studies.

Gravity matters: Differential processing of semantic and syntactic object–scene violations

Most of the evidence on the allocation of gaze to inconsistent objects in naturalistic scenes has come from semantic violations of the scene context. However, another kind of inconsistency involves the local structural setting, i.e., the syntactic context, in which an object is placed within a scene. Certain constituents of a scene require certain syntactic structures. In this study, we directly compared the effects of semantic and syntactic violations on eye movement control.

Whereas semantic inconsistencies referred to objects that did not fit the semantic context of the scene, syntactic

inconsistencies were created by making objects float. Similar to our findings regarding semantic inconsistencies, we have produced the first evidence that syntactic inconsistencies do not attract gaze prior to the fixation of the inconsistent object. Thus, neither semantic nor syntactic anomalies seem to be sufficiently processed outside of foveal viewing to control eye movements. Rather the detection of local syntactic structures seems to require fixation of the critical object. This result is in line with findings by Tatler, Gilchrist, and Land (2005) suggesting that direct fixation of an object is required to extract meaningful position information, since only then can the position information of a stored representation be compared to position information of the one currently processed. This is also consistent with the results from transsaccadic change detection experiments, where it has been shown that changes to the structural relationship between an object and its scene (e.g., rotation of an object in a scene) is typically only detected when the object has been fixated before and after the change (Hollingworth & Henderson, 2002; Hollingworth, Schrock, & Henderson, 2001).

While we were not able to find early effects of scene inconsistencies, both semantic and syntactic scene violations led to increased gaze on inconsistent objects once fixated. Previous studies have reported increased fixation densities and durations for semantically inconsistent objects, implying prolonged allocation of attention necessary to resolve the object–scene inconsistency (e.g., De Graef et al., 1990; Gareze & Findlay, 2007; Henderson et al., 1999; Hollingworth et al., 2001; Loftus & Mackworth, 1978; Rayner et al., 2009; Underwood & Foulsham, 2006; Underwood et al., 2007, 2008).

Extending previous work, we provide here the first evidence for an increased degree of attention allocation to objects that were syntactically inconsistent with the scene context: floating objects were fixated longer and more often than objects that rested on surfaces. Interestingly, we found an interaction of both types of inconsistencies during the first inspection of an object: Whereas non-floating objects showed longer gaze durations when they were semantically inconsistent, this inconsistency effect was eliminated when objects were floating. Thus, when objects violated the scene's syntactic structure, their semantic fit to the rest of the scene was rendered secondary.

We propose that the stronger effect of the syntactic violation is due to the lower probability of encountering such an object–scene inconsistency in everyday life. Coming across a floating cocktail glass in a kitchen will be more disturbing than finding a microscope on the kitchen counter. This disturbance can also be regarded as an extreme degree of surprise and therefore interpreted within the framework of surprise theories. Itti and Baldi (2005), for example, formulated surprise in a Bayesian framework as the difference between prior expectations of an observer about the world and new incoming data. Surprise then quantifies as the difference between the prior and posterior beliefs. The stronger the mismatch, the stronger the computed surprise, which will—when the mismatch is strong enough—lead to increased deployment of attention and human gaze to the surprising event. According to this framework, the mismatch between prior expectations regarding an object and its current representation is more extreme for syntactic compared to semantic violations. However, the surprise model in its current form computes surprise from low-level stimulus properties. Our data argue for increased attention allocation to surprising events based on prior experiences and higher-level cognitive processes of an observer. Cognitive factors should be included in models that aspire to account for human gaze in naturalistic scenes.

Conclusions

The study presented here allows us to draw two important conclusions regarding the debate on whether object–scene inconsistencies in the visual periphery can be processed to a degree sufficient to modulate eye movement control.

First, we found that neither semantic nor support violations led to earlier fixations of inconsistent objects compared to their consistent counterparts. Only upon fixation did both inconsistencies affect the deployment of attention and eye movements with inconsistent objects being fixated more often and longer than consistent objects. Thus, the findings of our study clearly speak against an extrafoveal influence of object–scene inconsistencies on initial eye movements during scene viewing.

Second, a direct comparison of semantic and syntactic violations showed that once fixated, both inconsistencies interactively modulated eye movement behavior. We therefore propose that the effect of object–scene inconsistency on eye movements varies as a function of prior beliefs and expectations. This further promotes the idea that we automatically assign certain expectations to objects within a scene regarding their semantic and syntactic integration, which can subsequently influence how long we hold our gaze when viewing scenes in the real world.

Acknowledgments

This project was supported within the DFG excellence initiative research cluster "cognition for technical systems-CoTeSys" by the DFG (De 336/2) and by Grant RES-062-23-1092 from the Economic and Social Science Research Council of the UK to JMH.

Commercial relationships: none.
Corresponding author: Melissa Le-Hoa Võ.
Email: melissa.vo@ed.ac.uk.

Address: Visual Cognition Unit, Psychology Department, 7 George Square, S32, University of Edinburgh, Edinburgh, EH8 9JZ, United Kingdom.

References

Becker, M. W., Pashler, H., & Lubin, J. (2007). Object-intrinsic oddities draw early saccades. *Journal of Experimental Psychology: Human Perception and Performance, 33,* 20–30. [PubMed]

Biederman, I., Mezzanote, R. J., & Rabinowitz, J. C. (1982). Scene perception: Detecting and judging objects undergoing relational violations. *Cognitive Psychology, 14,* 143–177. [PubMed]

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

Castelhano, M. S., & Henderson, J. M. (2005). Incidental visual memory for objects in scenes. *Visual Cognition, 12,* 1017–1040.

Castelhano, M. S., & Henderson, J. M. (2008). The influence of color on the activation of scene gist. *Journal of Experimental Psychology: Human Perception and Performance, 34,* 660–675. [PubMed]

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

Eckstein, M. P., Drescher, B. A., & Shimozaki, S. S. (2006). Attentional cues in real scenes, saccadic targeting, and Bayesian priors. *Psychological Science, 17,* 973–980. [PubMed]

Friedman, A. (1979). Framing pictures: The role of knowledge in automatized encoding and memory for gist. *Journal of Experimental Psychology: General, 108,* 316–355. [PubMed]

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

Henderson, J. M., & Hollingworth, A. (1999a). High-level scene perception. *Annual Review of Psychology, 50,* 243–271. [PubMed]

Henderson, J. M., & Hollingworth, A. (1999b). The role of fixation position in detecting scene changes across saccades. *Psychological Science, 5,* 438–443.

Henderson, J. M., & Hollingworth, A. (2003). Eye movements and visual memory: Detecting changes to saccade targets in scenes. *Perception & Psychophysics, 65,* 58–71. [PubMed]

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

Hollingworth, A., & Henderson, J. M. (1998). Does consistent scene context facilitate object perception? *Journal of Experimental Psychology: General, 127,* 398–415. [PubMed]

Hollingworth, A., & Henderson, J. M. (2002). Accurate visual memory for previously attended objects in natural scenes. *Journal of Experimental Psychology: Human Perception and Performance, 28,* 113–136.

Hollingworth, A., Schrock, G., & Henderson, J. M. (2001). Change detection in the flicker paradigm: The role of fixation position within the scene. *Memory & Cognition, 29,* 296–304. [PubMed]

Hollingworth, A., Williams, C. C., & Henderson, J. M. (2001). To see and remember: Visually specific information is retained in memory from previously attended objects in natural scenes. *Psychonomic Bulletin & Review, 8,* 761–768. [PubMed]

Itti, L., & Baldi, P. (2005). Bayesian surprise attracts human attention. *Proceedings in neural information processing systems* (vol. 19, pp. 1–8). Cambridge, MA: MIT Press.

Itti, L., & Koch, C. (2000). A saliency-based search mechanism for overt and covert shifts of visual attention. *Vision Research, 40,* 1489–1506. [PubMed]

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. [PubMed]

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

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

Oliva, A., & Torralba, A. (2006). Building the gist of a scene: The role of global image features in recognition. *Progress in Brain Research, 155,* 23–36. [PubMed]

Oliva, A., & Torralba, A. (2007). The role of context in object recognition. *Trends in Cognitive Science, 11,* 520–527. [PubMed]

Potter, M. C. (1975). Meaning in visual search. *Science, 187,* 965–966. [PubMed]

Rayner, K., Castelhano, M. S., & Yang, J. (2009). Viewing task influences eye movements during active scene perception. *Journal of Experimental*

Psychology: Learning, Memory, & Cognition, 35, 254–259.

Tatler, B. W., Gilchrist, I. D., & Land, M. F. (2005). Visual memory for objects in naturalistic scenes: From fixations to object files. *Quarterly Journal of Experimental Psychology A: Human Experimental Psychology, 58,* 931–960. [PubMed]

Tatler, B. W., Gilchrist, I. D., & Rusted, J. (2003). The time course of abstract visual representation. *Perception, 32,* 579–592. [PubMed]

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

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,* 766–786. [PubMed]

Underwood, G., & Foulsham, T. (2006). Visual saliency and semantic incongruency influence eye movements when inspecting pictures. *Quarterly Journal of Experimental Psychology, 59,* 1931–1949. [PubMed]

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

Underwood, G., Templeman, E., Lamming, L., & Foulsham, T. (2008). Is attention necessary for object identification? Evidence from eye movements during the inspection of real-world scenes. *Consciousness and Cognition, 17,* 159–170. [PubMed]