Disentangling the effects of spatial inconsistency of targets and distractors when searching in realistic scenes

Sara Spotorno
School of Psychology, University of Dundee, Park Place, Dundee, Scotland, UK
Institut de Neurosciences de la Timone (INT), CNRS & Aix-Marseille University, Marseille, France

George L. Malcolm
Department of Psychology, The George Washington University, Washington DC, USA

Benjamin W. Tatler
School of Psychology, University of Dundee, Park Place, Dundee, Scotland, UK

Previous research has suggested that correctly placed objects facilitate eye guidance, but also that objects violating spatial associations within scenes may be prioritized for selection and subsequent inspection. We analyzed the respective eye guidance of spatial expectations and target template (precise picture or verbal label) in visual search, while taking into account any impact of object spatial inconsistency on extrafoveal or foveal processing. Moreover, we isolated search disruption due to misleading spatial expectations about the target from the influence of spatial inconsistency within the scene upon search behavior. Reliable spatial expectations and precise target template improved oculomotor efficiency across all search phases. Spatial inconsistency resulted in preferential saccadic selection when guidance by template was insufficient to ensure effective search from the outset and the misplaced object was bigger than the objects consistently placed in the same scene region. This prioritization emerged principally during early inspection of the region, but the inconsistent object also tended to be preferentially fixated overall across region viewing. These results suggest that objects are first selected covertly on the basis of their relative size and that subsequent overt selection is made considering object-context associations processed in extrafoveal vision. Once the object was fixated, inconsistency resulted in longer first fixation duration and longer total dwell time. As a whole, our findings indicate that observed impairment of oculomotor behavior when searching for an implausibly placed target is the combined product of disruption due to unreliable spatial expectations and prioritization of inconsistent objects before and during object fixation.

Introduction

A key question in visual cognition research is how we process different high-level sources of information in a scene, and how the visual system dynamically utilizes them in order to guide the eyes during scene inspection. A powerful way to investigate these issues is to introduce some inconsistencies into the scene, violating scene-schemata learnt rules about the probability of an object's occurrence or position within a scene (Biederman, Mezzanotte, & Rabinowitz, 1982; Friedman, 1979). Detecting and processing inconsistency depend on the discrepancy between global information and local information according to predictions based on viewers' experience with types of scenes and objects. Therefore, examining how this discrepancy affects object selection and further inspection within a scene provides insights into the mechanisms of scene understanding in terms of both object-context relationships and object identities.

Perspectives on inconsistency and information processing during scene viewing

There is ongoing debate about how object-scene inconsistencies influence inspection behavior: in par-

Citation: Spotorno, S., Malcolm, G. L., & Tatler, B. W. (2015). Disentangling the effects of spatial inconsistency of targets and distractors when searching realistic scenes. *Journal of Vision*, *15*(2):12, 1–21, http://www.journalofvision.org/content/15/2/12, doi:10.1167/15.2.12.

doi: 10.1167/15.2.12 Received February 15, 2014; published February 10, 2015 ISSN 1534-7362

ticular whether effects are found for measures of attentional engagement (when the object is selected for fixation) or disengagement (how long the object is fixated once selected).

Evidence in favor of attentional engagement effects of inconsistency comes from findings of earlier or more probable ocular selection of objects that are inconsistent with scene context (Becker, Pashler, & Lubin, 2007; Bonitz & Gordon, 2008; Gordon, 2004, 2006; Loftus & Mackworth, 1978; Underwood & Foulsham, 2006; Underwood, Humphreys, & Cross, 2007; Underwood, Templeman, Lamming, & Foulsham, 2008). This would mean that, before fixating the object and, thus, bringing it into high-acuity foveal vision, we are able to access aspects of its identity and its relationships with scene context. In other words, despite a progressive drop in visual acuity at increasing eccentricities, we can process the object in the extrafoveal space sufficiently to detect to some extent its implausibility with respect to scene global semantics (e.g., a cow in a kitchen) and/or spatial organization (e.g., a chandelier on the floor of a living room), both available from the first glimpse of the scene (e.g., Biederman, 1977; Potter, 1976). This initial understanding would be sufficient to trigger a saccade toward the object, in order to bring it into the fovea for subsequent detailed analysis.

Other studies have found no support for prioritizing inconsistent objects for selection and have claimed that recognition of inconsistency—and thus of an object's semantic and spatial relationship with the scene—only occurs once the object is within foveal vision. According to this perspective, known in literature as the attention disengagement perspective (see Gordon, 2004; Henderson, Weeks, & Hollingworth, 1999), object-scene inconsistency only affects the time needed to move (i.e., disengage) gaze away from the inconsistent object. Any selection of inconsistent objects would instead be determined exclusively by factors unrelated to violation of object-scene associations, such as low-level features, high-level global attributes, or just by chance (e.g., Castelhano & Heaven, 2011; De Graef, Christiaens, & d'Ydewalle, 1990; Gareze & Findlay, 2007; Henderson et al., 1999; Malcolm & Henderson, 2010; Võ & Henderson, 2009, 2011).

Note that investigations supporting a preferential engagement (i.e., earlier fixation) on inconsistent objects have also reported longer foveal inspection of these objects. Overall, studies have suggested that longer inspection arises from the need for deeper processing in order to recognize inconsistent objects (e.g., Biederman et al., 1982; Friedman, 1979; Gordon, 2004) or, after their recognition, from the attempt to solve the context/local conflict they engender (e.g., De Graef et al., 1990; Henderson et al., 1999; Hollingworth & Henderson, 1998, 1999). The effect may emerge in longer dwell time and/or more fixations (Becker et al., 2007; Bonitz & Gordon, 2008; De Graef et al., 1990; Friedman, 1979; Henderson et al., 1999; Loftus & Mackworth, 1978; Mudrik, Deouell, & Lamy, 2011; Rayner, Castelhano, & Yang, 2009; Võ & Henderson, 2009) on inconsistent objects than on consistent objects. Mixed evidence, however, exists on whether inconsistency may also affect initial foveal processing of an object, resulting in longer first fixation durations (Bonitz & Gordon, 2008; Castelhano & Heaven, 2011; De Graef et al., 1990; Võ & Henderson, 2009; Underwood et al., 2008), or only emerges in later processing and in aggregate measures that take into account multiple fixations (Becker et al., 2007; Henderson et al., 1999; Rayner et al., 2009).

Inconsistency in visual search

Evidence regarding the effects of consistency on attentional allocation in scenes comes from a range of different paradigms: free viewing (Becker et al., 2007; Gareze & Findlay, 2007; Võ & Henderson, 2009, 2011), scene memorization (Brockmole & Henderson, 2008; Gareze & Findlay, 2007; Loftus & Mackworth, 1978; Underwood & Foulsham, 2006; Underwood et al., 2007), object recognition (Gordon, 2004), change detection (Brockmole & Henderson, 2008; Friedman, 1979; Hollingworth, Williams, & Henderson, 2001; Spotorno, Tatler, & Faure, 2013; Stirk & Underwood, 2007), priming (Gordon, 2006), image rating (Bonitz & Gordon, 2008; Rayner et al., 2009), binocular rivalry (Mudrik et al., 2011), comparative visual search (i.e., finding a difference between two scenes presented at the same time: Underwood et al., 2008), object naming (Coco, Malcolm, & Keller, 2013), and visual search. Unlike most of the other paradigms, in which the findings regarding preferential selection of inconsistency are very mixed, in visual search there is more agreement between studies. To our knowledge, no previous investigation has found selection prioritization for inconsistent objects during search. Studies that examined either search for targets without contextual associations (pseudo-objects: De Graef et al., 1990; a gray ball, Underwood & Foulsham, 2006) or search for consistent and inconsistent objects (Castelhano & Heaven, 2011; Eckstein, Drescher, & Shimozaki, 2006; Henderson et al., 1999; Malcolm & Henderson, 2010; Võ & Henderson, 2009, 2011) supported detection of inconsistency exclusively in foveal vision. If there was an effect on where observers looked in scenes, this appeared to be a later selection of inconsistent objects than consistent objects, requiring more time and fixations before reaching them. This was due to ineffectual, or even a misleading, contextual guidance (see also Mack & Eckstein, 2011; Neider & Zelinsky, 2006; Võ & Wolfe, 2013).

Typically, when studying search for consistent and inconsistent objects, only consistent and inconsistent targets have been compared, reporting better performance and more efficient eye guidance for consistent targets. We argue here that using only this kind of comparison may fail to reveal aspects of how inconsistency influences search. When we search for an object that is placed in an inconsistent location, two potential factors may influence oculomotor behavior: (a) ineffective contextual guidance disrupting search and (b) an attentional trigger effect due to extrafoveal (covert) detection of inconsistency. As such, whether we observe an earlier selection of inconsistent objects or not may reflect the relative balance of increased search time arising from inappropriate contextual guidance and decreased search time arising from extrafoveal detection of inconsistency. In order to conclude whether there is evidence for extrafoveal detection of object-context inconsistency, it is important to tease apart these two potential influences on search behavior when searching for inconsistently placed objects.

How can we disentangle these two competing (and opposite) effects? First, we need to have a "pure measure" of eye guidance due to spatial expectations regarding the target, activated before scene presentation. We will call this effect a "spatial expectation guidance effect." Second, we need to examine whether there is an attention engagement effect arising from online processing of object-scene inconsistency. We will call this effect a "spatial inconsistency guidance effect."

It is also essential to consider that in visual search within scenes, the visual system utilizes not only contextual guidance, but at least one other source of high-level information: target template, that is to say its working memory representation (see Zelinsky, 2008). When looking for consistent targets, contextual guidance and template guidance are always combined (Castelhano & Heaven, 2010; Ehinger, Hidalgo-Sotelo, Torralba, & Oliva, 2009; Kanan, Tong, Zhang, & Cottrell, 2009; Malcolm & Henderson, 2010), having a considerably higher impact on eye movements and search performance than any low-level salience (for review: Tatler, Hayhoe, Land, & Ballard, 2011). Therefore, we need to consider whether the precision of template information alters the relative strength of any expectation guidance effect and inconsistency guidance effect.

The present study

The objective of this study was to isolate any attentional impact of inconsistency per se, while analyzing the respective eye guidance of target template and spatial expectations in visual search within realistic scenes. When placed in a location that does not match expectations—that may be activated before scene presentation—concerning where to find it, the target object is necessarily inconsistent with scene context. Therefore, understanding search behavior in this situation always requires a distinction between the influence of the mismatch with the (preactivated) spatial expectations concerning the target and the influence of violations detected during simultaneous object-context processing in scene viewing (see also Demiral, Malcolm, & Henderson, 2012). In order to distinguish between these two potential components, we used scenes with a clearly defined boundary that separated two regions: one plausible and the other implausible for the target. Scene spatial violations were created by switching the locations of two objects, each from one of the regions. One of these objects was designated as the inconsistent target, implausibly placed with respect to both viewers' expectations about where to find it and object-context associations detected during scene processing. The other object was designated as the inconsistent distractor and placed in the region where participants would instead have expected to find the target. The inconsistent target and the inconsistent distractor occurred among several objects, all consistent with scene context, inserted in either the expected target region or the unexpected target region (see Figure 1 and Method).

Constructing the search scenes in this way allowed us to tease apart aspects of the manner in which inconsistency may influence search behavior. First, by comparing behavior with respect to the inconsistent target and the inconsistent distractor, we were able to consider the relative strength of the target template guidance effect and the spatial expectation guidance effect. While both objects were inconsistent with scene context, they differed with respect to their correspondence with guidance sources supplied by template and expectations: The target matched template features but did not match expectations about where it should be found (as it was placed in an implausible region); conversely, the inconsistent distractor did not match search template but it did match expectations concerning target location within the scene. Thus, both objects were inconsistently placed, allowing us to control for this factor in eye guidance and isolate relative importance of target appearance and expectations about target location in guiding search, when these two sources of information are in conflict.

Second, we were able to isolate the spatial inconsistency guidance effect on object selection and subsequent object inspection by comparing the inconsistent distractor with the consistent distractors placed in the same scene region. The crucial distinction between the inconsistent distractor and the consistent distractors was their spatial relationship with scene context. Only the inconsistent distractor violated rules about plausi-

a. Target object	b. Inconsistent distractor	c. Consistent distractor
- Matching with template - Mismatching with expectations about target location - Inconsistency with context	- Matching with template - Matching with expectations about target location - Inconsistency with context	- Mismatching with template - Matching with expectations about target location - Consistency with context

a vs. b
Relative guidance by **target template** and by **expectations about target location**

b vs. c
Effect of **inconsistency**

Figure 1. Schematic representation of comparisons within the switched scene arrangement. In this example, the vacuum is the target and the painting is the inconsistent distractor and the two scene regions analyzed are the floor (target expected region) and the wall (target unexpected region). The diagram describes object's properties with respect to matching with template, matching with expectations about where the target will be probably located within the scene, and level of consistency of object position with scene context. It also indicates sources of eye guidance investigated by comparing either the target object and the inconsistent distractor or the inconsistent distractor and the other objects (consistent distractors) plausibly placed in the same scene region (target expected region).

bility of placement in the scene. Both types of objects did not match template features, but did match the viewer's expectations about where to find the target (as they were all placed in a target plausible scene region).

In order to investigate the issues of relative strength of contribution in search and interplay between guidance by template, guidance by spatial expectation and guidance by inconsistency, we manipulated the availability of prior information concerning the target object, by using the target's name (at a basic category level, like "dog") or its precise picture as cue. Cueing the target with a verbal label leads to abstract representation based on long-term knowledge about typical target features, while cueing the target with its picture enables the viewer to form a detailed representation of what to look for. Several studies have examined the role of the level of detail of target template in visual search (e.g., Bravo & Farid, 2009; Castelhano, Pollatsek, & Cave, 2008; Castelhano & Heaven, 2010; Malcolm & Henderson, 2009, 2010; Maxfield & Zelinsky, 2012; Schmidt & Zelinsky, 2009, 2011; Vickery et al., 2005; Wolfe, Horowitz, Kenner, Hyle, & Vasan, 2004; Yang & Zelinsky, 2009), showing that a specific template facilitates efficient oculomotor behavior and enhances search performance. By comparing search following picture and word cues, we were able to consider whether varying the availability of specific information regarding the features of the target object influences the relative contribution of expectations and inconsistency upon search behavior in scenes.

Method

Participants

Thirty-two native English-speaking students (12 males), aged 18–34 ($M = 21$, $SD = 4.35$) participated for course credit and gave informed consent in accordance

with the institutional review board of the University of Dundee. All participants were naive about the purpose of the study and reported normal or corrected-to-normal vision.

Apparatus

Eye movements were recorded using an EyeLink 1000 (SR Research, Canada) at a sampling rate of 1000 Hz. Viewing was binocular, but only the dominant eye was tracked. Experimental sessions were carried out on a Dell Optiplex755 computer running OS Windows XP. Stimuli were shown on a ViewSonic G90f-4 19-in. CRT monitor, with a resolution of 800 × 600 pixels, and a refresh rate of 100 Hz. A chinrest stabilized the eyes about 63 cm away from the display. Manual responses were made on a response pad. Stimulus presentation and response recording was controlled by Experiment Builder (SR Research, Canada).

Materials

Forty-eight full-color photographs (800 × 600 pixels, 31.8° × 23.8°) of realistic scenes from a variety of categories (outdoor and indoor, natural and man-made) were used as experimental scenes. Each of them included two distinct regions (e.g., floor and wall). For each of the two scene regions, seven objects, taken from Hemera Images database (Hemera Technologies, Gatineau, Canada) or Google Images, were modified and inserted using Adobe Photoshop CS (Adobe, San Jose, CA).

Two versions of each experimental scene were made, corresponding to two types of scene arrangements (see Figure 2). In the normal scene arrangement, all the inserted objects were placed in highly plausible positions. Then, two of the inserted objects (one from each scene region) were selected and their positions interchanged in order to create the switched scene arrangement, in which they were thus both placed in low-plausible locations. One of these two objects was designated as being the target (in either the normal or the switched arrangement), while the other was designated as being the inconsistent distractor in the switched arrangement. The remaining six objects included in each scene region (consistent distractors) maintained the same position in both arrangements. The designated target object in each scene was counterbalanced across participants. Overall, the two critical objects embedded in each scene did not differ in their eccentricity from the center of the scene: $M = 11.2°$ ($SD = 2.2°$, range = 7.1°–15.6°) and $M = 11.1°$ ($SD = 2.1°$, range = 5.7°–15.3°), $t < 1$, $p = 0.765$. Moreover, the two objects did not differ in the area of the scene they occupied: $M = 2.5\%$ ($SD = 2.3\%$, range = 0.4%–10.2%) and $M = 2.4\%$ ($SD = 1.7\%$, range = 0.4%–7.2%), $t < 1$, $p = 0.701$.

The plausibility of position of the two objects in both arrangements was rated on Likert scales (from 1, low, to 6, high) in a previous pilot study (detailed in Spotorno, Malcolm, & Tatler, 2014) by 10 judges, who did not take part in the main experiment. They considered simplified versions of the scenes, not containing consistent distractors in either of the two regions. As scores were not normally distributed, we report here median and interquartile range values, with results of Wilcoxon signed-ranks tests (two tailed). This study indicated that locations were well chosen in order to manipulate object spatial consistency and that, in scenes with switched arrangement, placements of the two objects were of similar inconsistency ($Mdn = 1.20$, $IQR = 0.75$ and $Mdn = 1.20$, $IQR = 0.40$, $Z = -1.50$, $p = 0.134$). In the normal scene arrangement, both objects were placed in highly probable locations ($Mdn = 5.60$, $IQR = 0.60$ and $Mdn = 5.60$, $IQR = 0.80$, $Z < 1$, $p = 0.540$). The same study evaluated matching between these objects and their name, quality of their insertion in the scene, object relevance for scene meaning and object visual salience, using the same six-point Likert scales. The results indicated that the targets were clearly defined by the name chosen to cue them (overall: $Mdn = 5.67$, $IQR = 0.90$). Insertions of the two critical objects were of good and comparable quality in normal ($Mdn = 4.20$, $IQR = 1.40$ and $Mdn = 4.40$, $IQR = 1.35$, $Z < 1$, $p = 0.501$) and switched arrangements ($Mdn = 4.00$, $IQR = 1.00$ and $Mdn = 4.40$, $IQR = .95$, $Z = -1.04$, $p = 0.298$). The two objects had similar level of subjective salience (normal arrangement: $Mdn = 4.80$, $IQR = 1.20$ and $Mdn = 4.60$, $IQR = 1.55$, $Z < 1$, $p = 0.659$; switched arrangement: $Mdn = 4.40$, $IQR = 1.30$ and $Mdn = 4.10$, $IQR = 1.60$, $Z = -1.67$, $p = 0.094$) and semantic relevance for the scene's gist (normal arrangement: $Mdn = 4.40$, $IQR = 0.95$ and $Mdn = 4.40$, $IQR = 1.35$, $Z < 1$, $p = 0.329$; switched arrangement: $Mdn = 3.80$, $IQR = 1.35$ and $Mdn = 3.70$, $IQR = 1.00$, $Z < 1$, $p = 0.986$).

In order to manipulate the template of the target, picture and word cues were created. To create the picture cues, each object was pasted in the middle of a white background, appearing exactly as it subsequently did in the scene regarding size, color, etc. To create the word cues, 48 verbal labels (up to three words) of the objects (font: Courier, color: black, font size: 72 point), subtending 2.14° in height, were centered on a white background.

Seventy-eight further scenes were added to the experiment, four for practice and the others as fillers. These scenes were always presented with a normal arrangement, and an inserted object was used as the target in each of them. Thirty-nine picture cues and 39 word cues were created for these scenes.

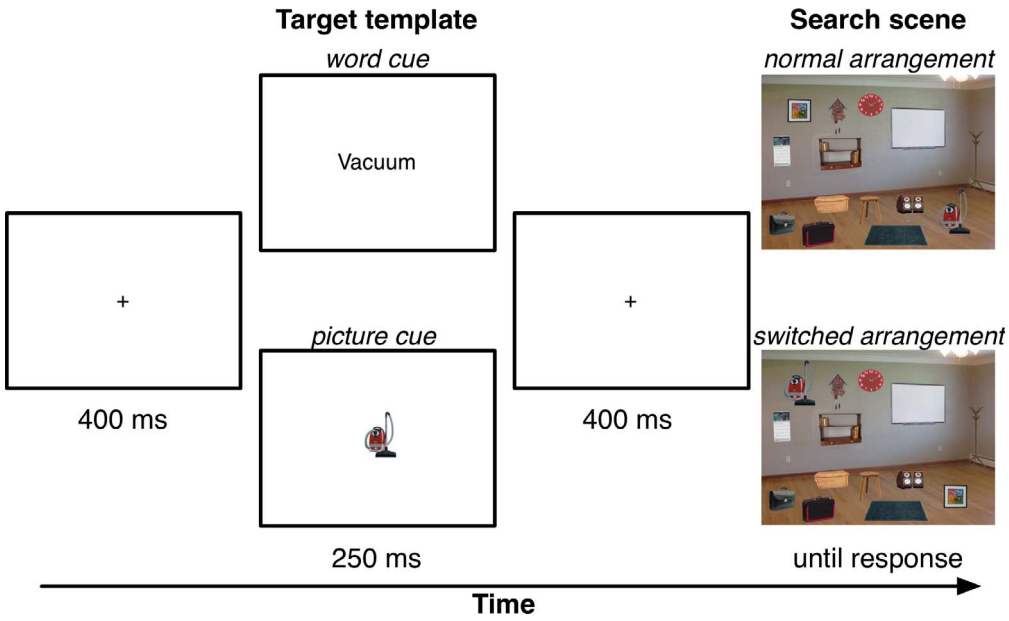

Figure 2. Example of screen shots of trials. This example shows the two types of target template and the two scene arrangements. Please note that each trial started with a drift check screen (here not depicted).

Procedure

Prior to the experiment each participant underwent a randomized nine-point calibration procedure, which was validated in order to ensure that the average error was less than 0.5° and the maximum error in one of the calibration points was less than 1°. Recalibrations were performed during the task if necessary. Before each trial sequence, a single-point calibration check was applied as the participant fixated a dot in the center of the screen. When the single-point calibration check was deemed successfully (error less than 1°), the experimenter initiated the trial. The trial sequence is depicted in Figure 2. Participants responded with a button press as soon as they located the target object within the scene.

The experiment had a 2 (Template Type) × 2 (Scene Arrangement) design. Each scene was displayed only once during the experiment. Half of the experimental scenes were cued with target's picture, the other half with target's name. Moreover, each participant saw half of them with normal arrangement and the other half with switched arrangement. All of the experimental manipulations were counterbalanced across participants for the experimental scenes. The filler scenes were presented only with normal arrangement, meaning that 75% of all the scenes was viewed with targets in high-probability locations. This percentage ensured that participants would recognize scene context as a potentially reliable source of guidance throughout the experiment. Experimental and filler scenes were intermixed and presented in a random order. The eye movements from the filler trials were not analyzed. The experiment lasted for about 30 min.

ROIs definition and data analyses

The regions of interest (ROIs) for scoring eye movements were defined as the smallest fitting rectangle which encompassed the object. They were determined for each of the seven objects placed in each of the two scene regions. A saccade was considered as being directed toward a specific ROI if its angular direction was within 22.5° of the angular direction to the center of the ROI. A fixation was considered as being on a specific ROI if the center of gaze indicated by the eye tracker fell within 0.5° of the boundary of the ROI.

A set of larger ROIs, which encompassed the entire target expected region, was also defined (two ROIs for each scene, one for each possible target object). This enabled us to select fixations made in this region in order to carry out analyses limited to trials with a switched scene arrangement in which the region was entered (see the section Extrafoveal processing of object-scene inconsistencies).

Raw data were parsed into saccades and fixations using the SR Research algorithm. Subsequent analyses of saccades and fixations were conducted using routines written in Matlab 2012b. We discarded from analyses trials in which participants were not maintaining the central fixation (i.e., fixating within 2° radius from scene's center) when the scene appeared (7.21%) and trials with errors (6.81%). Responses were considered correct if participants looked at the target when

Figure 3. Fixation density distributions for each experimental condition for an example scene. Distributions comprise data across all search epochs from all participants and were created by iteratively adding Gaussians centered at each fixation location, each with full width at half maximum of 2° of visual angle. Hotter colors denote greater fixation density. The first fixation in each trial (which began on the central pre-trial marker) is not included in these distributions.

pressing the buttons or during the immediately preceding fixation. Trials with first saccade latency shorter than 50 ms (4.48%) were also excluded, as these were likely to reflect anticipatory responses rather than responses based on processing of the information in the 2were run using the lmer() function of the lme4 package (Bates, Maechler, Bolker, & Walker, 2014) in the R programming environment (The R Foundation for Statistical Computing, Version 3.0.0, 2013). We ran linear mixed models (LMMs), or generalized LMMs (GLMMs) for binomial data, with participants, scenes, and trials specified as random factors.

LMMs and GLMMs have many advantages over traditional analysis of variance (ANOVA) models. Crucially, they allow a simultaneous estimation of between-subject and between-item variance. In addition, they are known to be more robust than ANOVAs when a design is not fully balanced as a result of data excluded using the criteria explained above (see Kliegl, Masson, & Richter, 2010).

For each model, we report the t values, or the z values for binomial data, of the predictors, and the associate p values. When necessary, p values were estimated using Markov Chain Monte Carlo sampling derived from the pval.fcn() function in the languageR library. When an interaction was significant, follow-up LMMs or GLMMs were carried out in order to analyze simple effects.

We also ran independent or paired-sample t tests (two tailed) in order to compare the size and the eccentricity of the inconsistent distractor with the size and the eccentricity of the biggest consistent distractor included in the same region (expected for the target) in scenes with a switched arrangement; one-sample t tests (one-tailed) were run to compare selection probabilities for the inconsistent distractor and the biggest consistent distractor to chance (see the section Extrafoveal processing of object-scene inconsistencies).

Graphics were created using the ggplot2 package (Wickham, 2009).

Results

There were clear differences in viewing behavior between the experimental conditions, in fixation density distributions (an example is supplied by Figure 3) and in the temporal dynamics of eye movement behavior. These differences are explored in the sections that follow.

Eye guidance across search

Template type (picture vs. word), scene arrangement (normal vs. switched), and their interaction were entered as fixed factors in GLMMs and LMMs that analyzed oculomotor behavior across the overall search process. We divided search into three phases (see Malcolm & Henderson, 2009, 2010): (a) search initiation, considering direction and latency of the first

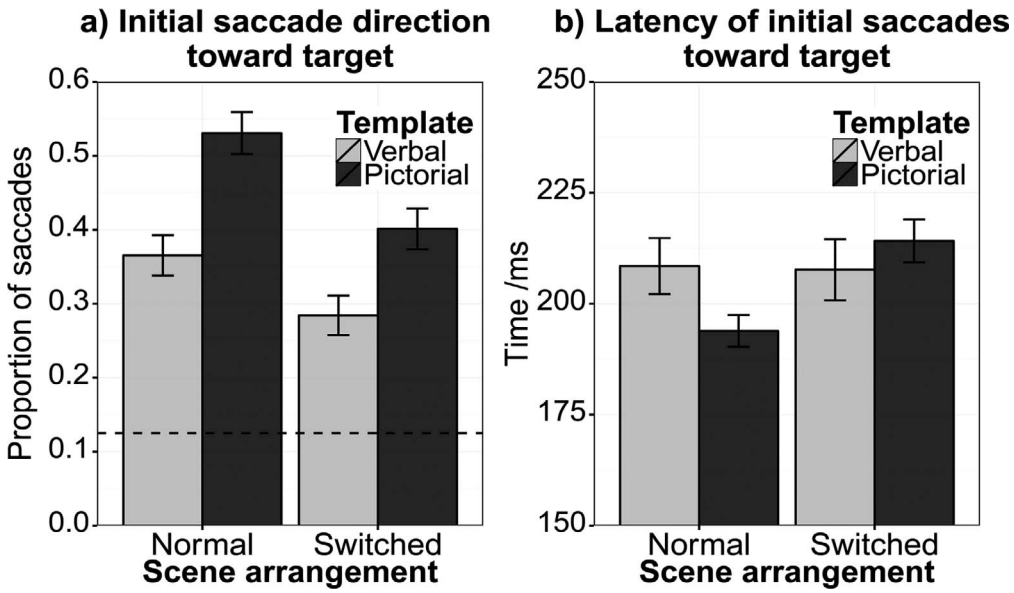

Figure 4. Eye movement measures during search initiation, as a function of type of template and type of scene arrangement. Bars show condition means ± 1 *SE*. (a) Probability of directing the first saccade toward the target object (dashed line indicates chance level). (b) Latency of the first saccades directed toward the target object.

saccade in the scene; (b) scene scanning, taking into account the number of fixations before first fixating the target and their total duration (scanning time, in milliseconds); the initial fixation in the scene was removed from these computations as it was always in the center of the scene following the calibration check; and (c) target verification (in milliseconds), measuring the time needed, from its first fixation, to accept the currently inspected object as being the target, and to execute the manual response.

Five measures (Figures 4 and 5) were examined in this set of analyses. Following inspection of the distribution and residuals, latency of the initial saccades directed toward the target, scanning time and verification time were log-transformed in order to meet LMM assumptions.

Search initiation

Proportion of first saccades directed toward the target object: The proportion of first saccades directed toward the target object was influenced by the type of template, $\beta = 0.663$, $SE = 0.124$, $z = 5.34$, $p < 0.001$, being greater following a picture cue (0.46) than a word cue (0.33),

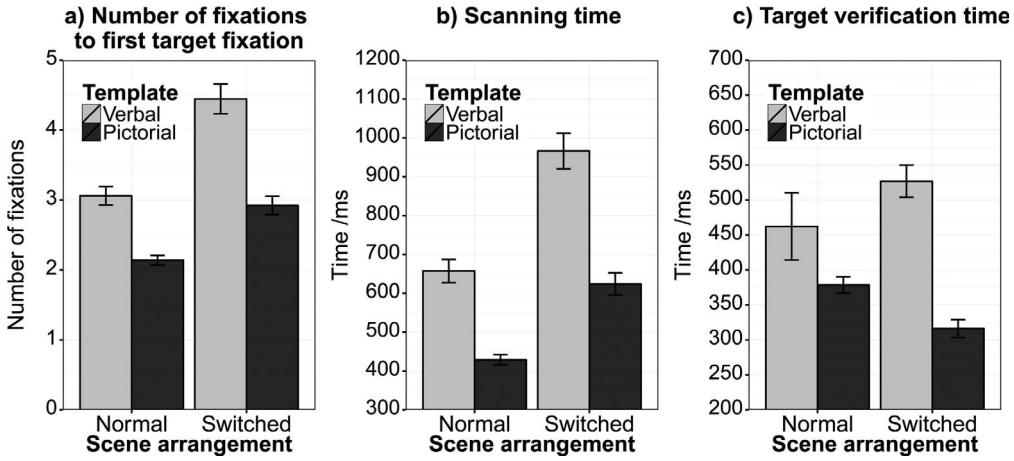

Figure 5. Eye movement measures during scene scanning and target verification phases, as a function of type of template and type of scene arrangement. Bars show condition means ± 1 *SE*. Scene scanning: (a) number of fixations until first fixation on the target object (the first central fixation, at scene onset, has been excluded from this count); (b) time from the end of the first saccade in the scene to the first fixation on the target object. Target verification: (c) time from the beginning of first fixation on the target to button pressure.

and by the scene arrangement, $\beta = 0.519$, $SE = 0.124$, $z = 4.18$, $p < 0.001$, with a higher proportion of first saccades directed toward the target when it was plausibly placed (0.45) than when it was placed in an implausible region (0.36). The two-way interaction was not significant, $\beta = -0.163$, $SE = 0.332$, $z < 1$, $p = 0.622$ (Figure 4, left panel).

One-sample t tests (one tailed) showed that, in all conditions, the probability of initiating search toward the target was above chance (corresponding to the ±22.5 degree criterion for considering a saccade as directed toward target's ROI), all ts ≥ 5.95, all ps < 0.001.

Latency of the initial saccades directed toward the target: There was a main effect of the type of scene arrangement, $\beta = -0.022$, $SE = 0.009$, $t = -2.36$, $p = 0.018$, as search initiation toward the target object was faster when the target was plausibly placed (200 ms) than when it was in an implausible location (212 ms). Latency of correct search initiation was not influenced significantly by either the type of template, $\beta = -0.013$, $SE = 0.009$, $t = -1.45$, $p = 0.148$, or the two-way interaction, $\beta = 0.030$, $SE = 0.042$, $t < 1$, $p = 0.473$ (Figure 4, right panel).

Scene scanning

Number of fixations before first target fixation: There was a significant main effect of the type of template, $\beta = -1.257$, $SE = 0.132$, $t = -9.49$, $p < 0.001$, and a significant main effect of the scene arrangement, $\beta = -1.140$, $SE = 0.132$, $t = -8.64$, $p < 0.001$. Fewer fixations were necessary to locate the target following a picture cue (2.53) than following a word cue (3.72), and with a normal scene arrangement (2.60) than with a switched arrangement (3.65). The two-way interaction was not significant, $\beta = -0.063$, $SE = 0.398$, $t = -1.59$, $p = 0.111$ (Figure 5, left panel).

Scanning time: The time to first fixate the target object was affected by the type of the template, $\beta = -0.156$, $SE = 0.013$, $t = -11.78$, $p < 0.001$, being significantly shorter following a picture cue (527 ms) than following a word cue (808 ms), and by the type of scene arrangement, $\beta = -0.129$, $SE = 0.013$, $t = -9.77$, $p < 0.001$, being significantly shorter in a normal arrangement (544 ms) than in a switched arrangement (787 ms). The effect of the two-way interaction was not significant, $\beta = -0.024$, $SE = 0.046$, $t < 1$, $p = 0.591$ (Figure 5, central panel).

Target verification time

There was a main effect of the type of template, $\beta = -0.121$, $SE = 0.013$, $t = -9.35$, $p < 0.001$, while there was a strong trend toward a main effect of scene arrangement, $\beta = -0.024$, $SE = 0.013$, $t = -1.89$, $p = 0.059$. The two-way interaction was significant, $\beta = -0.185$, $SE = 0.089$, $t = -2.08$, $p = 0.038$. When the target location was unexpected, verification time was quicker following a picture cue than following a word cue, $t = -4.57$, $p < 0.001$, while no significant differences were found depending on the cue when the target was in an expected location, $t < 1$, $p = 0.543$. Moreover, verification was quicker with a normal arrangement than with a switched arrangement only when target template was abstract, $t = -2.41$, $p = 0.016$, while with a pictorial template the time to verify the target was not influenced by the plausibility of target placement, $t = 1.54$, $p = 0.125$ (Figure 5, right panel).

Visual search within unreliable scene contexts

The above results showed clearly that oculomotor efficiency, throughout all the search process, was reduced when the target was either defined by an abstract template or placed in an implausible location. However, the above analyses could not distinguish the relative contribution of the three different sources of high-level information (target representation, spatial expectation about the target, online processing of object-scene associations) to guiding the eyes in the scene. In order to tease them apart, we limited the following analyses specifically to the switched scene arrangement. In this arrangement (see Method and Figure 1), the target and a distractor were presented each in the scene region where the other should be located.

First (section Comparing inconsistent targets with inconsistent distractors during search initiation), by comparing guidance toward the inconsistently placed target to guidance toward the inconsistently placed distractor, we were able to tease apart the relative contribution of the target template (that directs the gaze toward the highest local matching in the scene with target appearance) and spatial expectation guidance (that directs the eyes toward the scene region in which the inconsistent distractor is), while controlling for any effect of online detection of object-scene inconsistency (as both objects were inconsistent). Second, by comparing the inconsistently placed distractor with the consistently placed distractors in the same scene region, we were able to examine whether there was any effect of spatial inconsistency in either extrafoveal processing (see the section Extrafoveal processing of object-scene inconsistencies) or foveal processing (section Foveal processing of object-scene inconsistencies) for objects that were comparable in terms of their mismatch to both the template and spatial expectations for the target object.

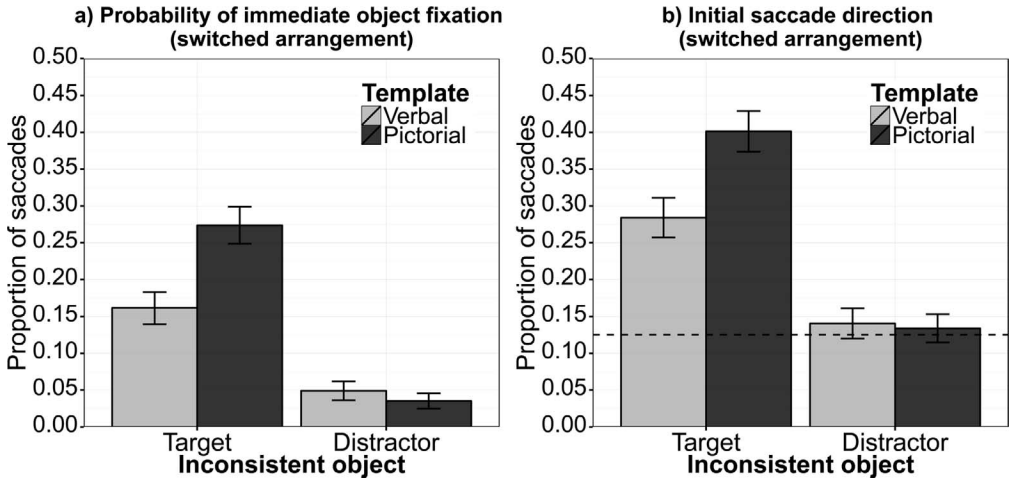

Figure 6. Search initiation in the switched scene arrangement, as a function of type of template and type of object (target object vs. inconsistent distractor). Bars show condition means ± 1 *SE*. (a) Probability that the first saccade within the scene landed on either the target or the inconsistent distractor; (b) Probability that the first saccade within the scene was directed toward either the target or the inconsistent distractor (dashed line indicates chance level).

Comparing inconsistent targets with inconsistent distractors during search initiation

In these analyses, we focused on early guidance during inspection, taking into account the first saccade in the scenes with a switched arrangement (Figure 6). Type of template, type of object (inconsistent target vs. inconsistent distractor), and their interactions were entered as fixed factors in these GLMMs that considered the probability that the first saccade landed on the object and the probability that it was launched toward it. Because of the nature of the task, we believe that any other measure aimed to assess either relatively later preferential selection or greater foveal processing comparing these two types of objects would be misleading. Indeed, once the target has been found the search process must be terminated as quickly as possible, with enough verification to be reasonably sure that the current object matches the template. In cases when the inconsistent distractor is fixated, the viewer is obviously in a fundamentally opposite situation: Scene inspection must be further pursued and any foveal process of the current object has the aim of rejecting it as a target. *Probability of immediate object fixation*: This measure considers the landing point of the initial saccade in the scene. The type of object had a significant main effect, $\beta = 2.010$, $SE = 0.244$, $z = 8.25$, $p < 0.001$, while the main effect of the type of template was not significant, $\beta = 0.225$, $SE = 0.268$, $z < 1$, $p = 0.402$. The significant two-way interaction, $\beta = -1.167$, $SE = 0.485$, $z = -2.41$, $p = 0.016$, revealed that a pictorial template, compared to a verbal template, enhanced the probability that the first saccade landed on the (inconsistently placed) target object, $z = 4.37$, $p < 0.001$, while the type of template did not have an influence for the inconsistent distractor, $z < 1$, $p = 0.457$. Overall, the inconsistent distractor had a very low probability of being fixated immediately, considerably lower than the target object with either a pictorial template, $z = -8.38$, $p < 0.001$, or a verbal template, $z = -4.37$, $p < 0.001$ (Figure 6, left panel). *Probability of directing the first saccade toward the target*: We also examined whether a more important contribution of expectations concerning target placement in the scene might be reported when considering this less stringent measure of early guidance. The findings, however, were very similar compared to those reported for probability of immediate fixation. There was a significant main effect of the type of object, $\beta = 1.201$, $SE = 0.149$, $z = 8.07$, $p < 0.001$, while the type of template had only a weak tendency toward significance, $\beta = 0.247$, $SE = 0.149$, $z = 1.66$, $p = 0.098$. The two factors interacted significantly, $\beta = -0.593$, $SE = 0.297$, $z = -1.99$, $p = 0.046$, but a significantly higher proportion of first saccades was directed toward the target object than toward the inconsistent distractor, with either a pictorial template, $z = 7.76$, $p < 0.001$, or a verbal template, $z = 4.11$, $p < 0.001$. The interaction depended on the fact that a picture cue, compared to a word cue, enhanced the probability of directing the first saccade toward the inconsistent target, $z = 3.67$, $p < 0.001$, whereas the inconsistent distractor was saccaded to in the same proportion of trials regardless of the search template, $z < 1$, $p = 0.659$ (Figure 6, right panel).

One-sample t tests (one tailed) showed that the probability of initiating search toward the target was above chance (1/8 of the scene), following either a word, $t(284) = 5.95$, $p < 0.001$, or a picture cue, $t(313) = 9.97$, $p < 0.001$.The probability of directing the initial saccade toward the inconsistent distractor was at chance following either a word, $t < 1$, $p = 0.229$, or a picture cue, $t < 1$, $p = 0.325$.

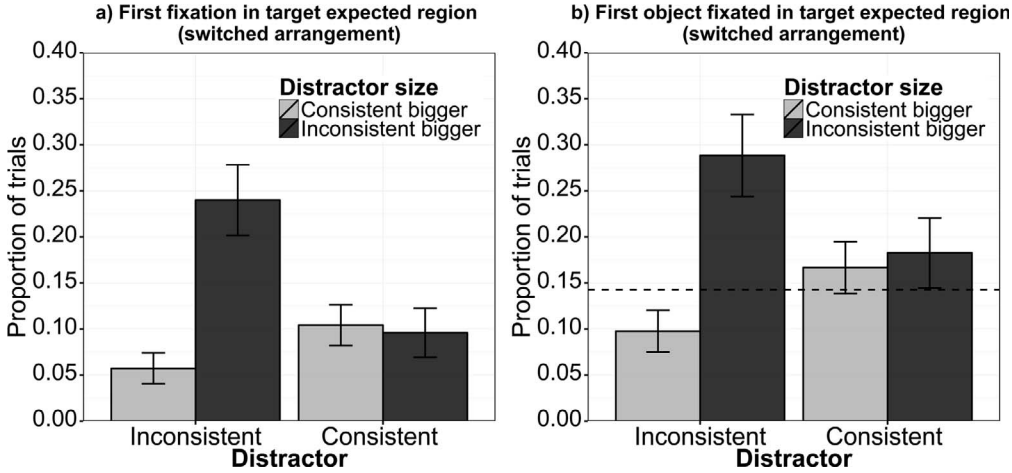

Figure 7. Early eye movements in target expected region in scenes with switched arrangement, as a function of the type of distractor object (inconsistent distractor vs. reference consistent distractor) and its relative size difference (inconsistent distractor bigger than the reference consistent distractor, reference consistent distractor bigger than the inconsistent distractor). Bars show condition means ± 1 *SE*. (a) Proportion of trials in which the first fixation in the target expected region was on the object; (b) Proportion of trials in which the first object fixated in the target expected region was the inconsistent distractor or the consistent distractor (dashed line indicates chance level).

Extrafoveal processing of object-scene inconsistencies

This set of measures takes into account only trials with a switched scene arrangement in which the scene region expected for the target was entered at some point during the search (59% of total trials with this arrangement). In 75% of these trials, this region was saccaded to during search initiation. This indicates that, when directing the eyes, misleading spatial expectations acted mainly within the very first fixation at the scene, whilst participants saccaded into the target expected region rarely after having inspected the actual but implausible region containing the target.

The inconsistent distractor was included in the target expected region with six consistent objects that were relatively smaller in most cases (size of the inconsistent distractor's ROI: $M = 18.5^{\circ 2}$, $SD = 14.1$, range = 2.8–75.2; size of the consistent objects: $M = 10.0^{\circ 2}$, $SD = 5.6$, range = 2.1–26.2). The analyses that follow include a measure of the area of the objects under test in order to account for any possible influence that the size of the object in the target expected region might have upon its selection, given that object size may play an important part in perceptual selection in scenes (Borji, Sihite, & Itti, 2013; Xu, Jiang, Wang, Kankanhalli, & Zhao, 2014). We compared therefore, in each trial, the inconsistent distractor with the biggest consistent object (ROI size: $M = 20.4^{\circ 2}$, $SD = 14.0$, range = 2.8–77.4) that was included in the same region. We computed the difference in size between these two objects, creating a binomial variable that described whether the larger object was the inconsistent distractor (in 41.64% of the cases, $t(126) = 14.00$, $p < 0.001$) or this reference consistent distractor (in the remaining cases, $t(191) = 15.52$, $p < 0.001$) had the greatest size in the region. In following analyses, we entered type of template, type of distractor (inconsistent distractor, reference biggest consistent distractor), size difference (inconsistent bigger, consistent bigger), and their interactions as fixed factors in separate GLMMs. The analyses reported (Figures 7 and 8) considered trials in which the absolute difference between the two objects corresponded to at least $1^{\circ 2}$ (10.2% of the data were removed following this criterion).

We also conducted supplementary analyses to explore whether any disruption to 3-D cues in scenes that arose from the transposition of objects between regions might be driving our findings. We identified 14 scenes for which the inconsistent distractor was not coplanar with the background in which it was inserted, leading to the possibility that this distractor was noticeable also on the basis of dissimilarity with respect to 3-D cues, in addition to syntactic violations of object-scene associations. However, removing these scenes from the analyses did not change substantially the pattern of results (for details, please see footnote).[1]

First fixation in target expected region: We analyzed the probability that the first saccade made into the target expected region landed on the inconsistent distractor or on the consistent distractor that had the greatest size among the six consistent distractors. We took into account all the trials with switched scene arrangement in which the target expected region was entered. A *t* test showed that the difference in size ($M = 15.1^{\circ 2}$, $SD = 13.0$, range = 1.1–62.9) was not unbalanced toward one type of distractor object, $t(266.487) = 1.41$, $p = 0.161$.

There were no main effects of the type of template or the type of distractor, both zs < 1, both ps ≥ 0.338, but

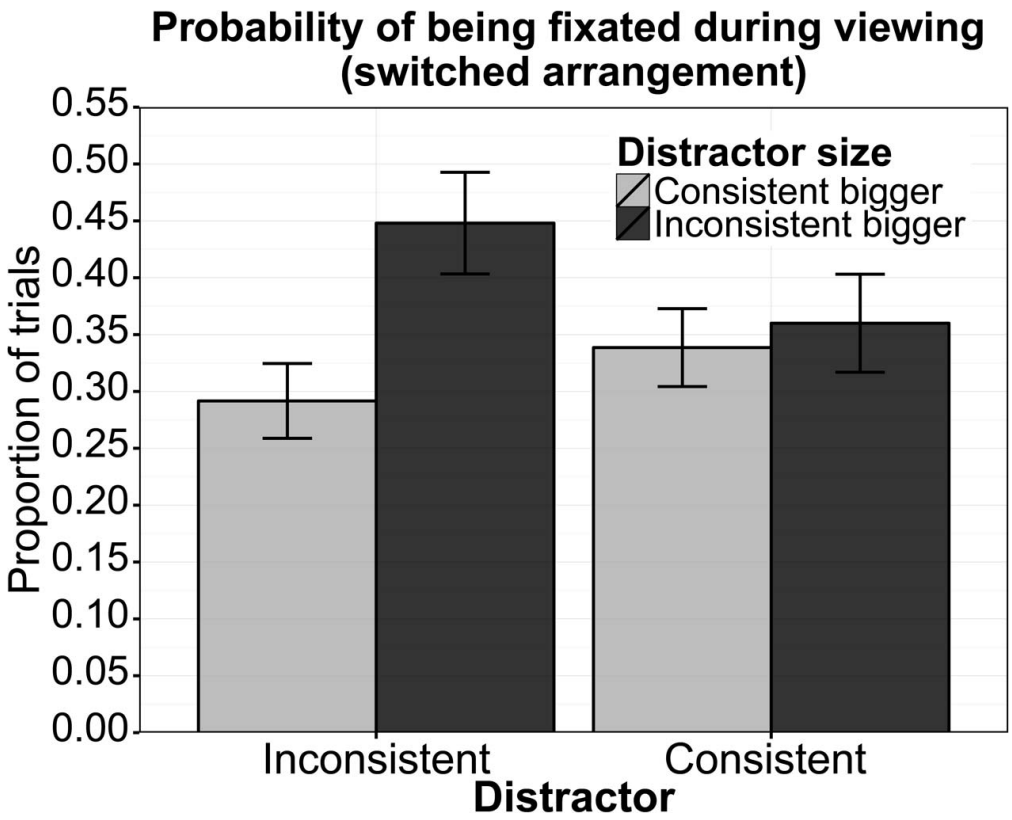

Figure 8. Probability that the object was fixated at least once in scenes with switched arrangement, as a function of the type of distractor object and its relative size difference. Bars show condition means ± 1 *SE*.

there was a main effect of size difference between the two distractors, $\beta = 0.704$, $SE = 0.309$, $z = 2.28$, $p = 0.023$, qualified by a significant interaction between size difference and type of distractor, $\beta = -1.865$, $SE = 0.561$, $z = -3.32$, $p < 0.001$ (Figure 7, left panel). All other interactions were not significant, all zs < 1, all ps ≥ 0.382. Follow-up models showed that the inconsistent distractor was more likely to be immediately fixated than a smaller reference consistent distractor, $\beta = 1.455$, $SE = 0.410$, $z = 3.55$, $p < 0.001$, while the reference consistent distractor was not immediately fixated more than a smaller inconsistent distractor, $\beta = 0.548$, $SE = 0.353$, $z = 1.55$, $p = 0.121$. Moreover, the first saccade in the region landed significantly more often on the inconsistent distractor when it was bigger than when it was smaller than the consistent distractor, $\beta = 1.77$, $SE = 0.405$, $z = 4.38$, $p < 0.001$, while the probability of first saccade landing on the consistent distractor was not affected by whether this object was bigger or smaller than the inconsistent distractor, $\beta = 0.124$, $SE = 0.391$, $z < 1$, $p = 0.751$.

We considered whether these results indicating an advantage of inconsistency (even though limited to cases in which the inconsistent distractor was the biggest object in the region) might depend on a difference in eccentricity between the two distractor objects with respect to the center of the scene, rather than being due to an advantage of semantic inconsistency on early object selection. Such effect would be coherent with the central fixation bias usually reported in static images (e.g., Tatler, 2007). We found that this was not the case. The consistent distractor was on average slightly less eccentric ($M = -0.9°$) than the inconsistent distractor, $t(318) = -4.23$, $p < 0.001$. Moreover, we found that this difference in eccentricity was greater ($t(236.631) = 2.16$, $p < 0.031$) when the consistent distractor was smaller ($t(126) = -4.01$, $p < 0.001$, $M = -1.60°$) than when it was bigger ($t(191) = -2.04$, $p = 0.042$, $M = -0.55°$) than the inconsistent distractor.

First object fixated in target expected region: The first saccade into the target expected region may not land on any object, but may select background locations en route to later object fixation. We therefore considered a less stringent measure of object selection: The probability that the first object fixated in the region was the inconsistent distractor or the reference consistent distractor. This measure indicates which object was prioritized in early selection compared to the others included in that region. For this analysis, we considered trials in which at least one object was fixated in the region (87.8% of the cases). A t test showed that the difference in size ($M = 15.1°^2$, $SD = 13.1$, range = 1.0–62.9) was not unbalanced toward one type of distractor object, $t < 1$, $p = 0.326$.

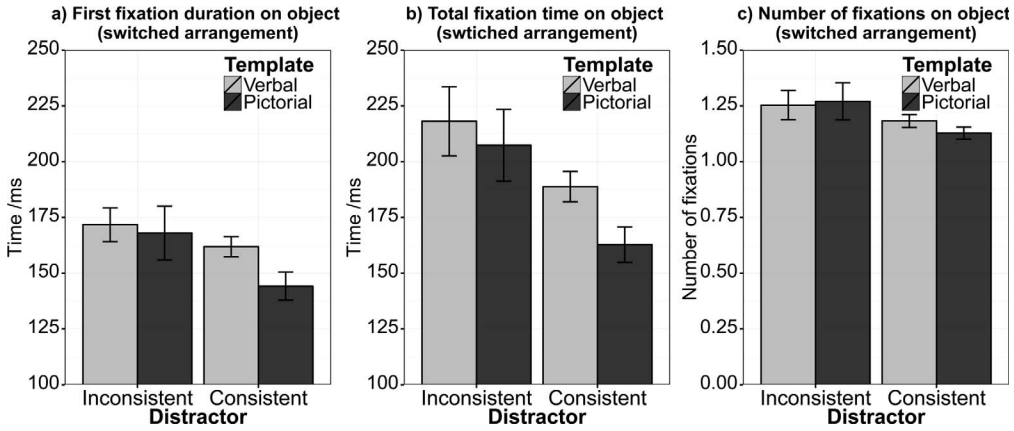

Figure 9. Eye movement measures of foveal processing in scenes with switched arrangement, as a function of the type of template and the type of distractor object. Bars show condition means ± 1 *SE*. (a) Duration of first fixation on the object; (b) Total fixation time on the object (dwell time); (c) Number of fixations on the object.

There were no main effects of the type of template or the type of distractor, both zs < 1, both ps ≥ 0.337, but there was a main effect of size difference between the two objects, $\beta = 0.704$, $SE = 0.309$, $z = 2.28$, $p = 0.023$, which was qualified by a significant interaction between size difference and type of distractor, $\beta = -1.865$, $SE = 0.561$, $z = -3.24$, $p < 0.001$ (Figure 7, right panel). All other two-way interactions and the three-way interaction were not significant, all zs < 1, all ps ≥ 0.372. Follow-up models revealed that the inconsistent distractor had a greater probability of being the first object fixated in the region compared to a smaller consistent distractor, $\beta = 0.750$, $SE = 0.369$, $z = 2.03$, $p = 0.042$, while there was only a tendency to select first a bigger consistent distractor more often than a smaller inconsistent distractor, $\beta = 0.577$, $SE = 0.309$, $z = 1.87$, $p = 0.062$. In addition, while the inconsistent distractor was selected first more often when it was bigger than when it was smaller than the consistent distractor, $\beta = 1.340$, $SE = 0.379$, $z = 3.53$, $p < 0.001$, the probability of first fixating the consistent distractor did not depend on its difference in size with respect to the inconsistent distractor, $\beta = -0.032$, $SE = 0.369$, $z < 1$, $p = 0.932$.

One-sample t tests (one tailed) showed that only inconsistent distractors, when they were the biggest object in the region, were first selected more than chance (1/7 objects included in the same region), $t(105) = 3.19$, $p = 0.001$, while in all the other conditions the probability of selecting the inconsistent or the reference consistent distractor was at chance, all ts < 1, all ps ≥ 0.167.

As before, we found the same differences in eccentricity between the tested objects and these were not in a direction that might have produced our observed pattern of results. It seems therefore unlikely that the advantage found for inconsistent distractors in the selection of the first object in the region was due to differences in eccentricity.

Probability of object fixation during viewing: We explored whether object inconsistency determines a preferential fixation selection also in a later stage of scene viewing, analyzing the probability that the object was fixated at least once during the trial. This measure considered the same set of trials as the previous one, taking into account all the fixations made in the entire scene.

We found a main effect of the type of template, $\beta = 0.988$, $SE = 0.206$, $z = 4.79$, $p < 0.001$, indicating a greater probability of fixating one of the two critical objects during viewing following a word cue ($M = 0.43$, $SE = 0.026$) than following a picture cue ($M = 0.26$, $SE = 0.026$). The main effects of type of distractor and size difference were not significant, both zs ≤ 1.46, both ps ≥ 0.122.

Type of distractor and size difference interacted significantly, $\beta = -0.866$, $SE = 0.388$, $z = -2.23$, $p = 0.026$ (Figure 8). The probability of fixating the object during viewing tended to be greater for the inconsistent distractor, when it was the biggest object in the region, than for the consistent distractor, $\beta = 0.500$, $SE = 0.288$, $z = 1.74$, $p = 0.082$. This probability did not differ significantly however for the consistent distractor, when it was the biggest object in the region, compared to the inconsistent distractor, $\beta = 0.267$, $SE = 0.241$, $z = 1.11$, $p = 0.267$. In addition, while participants were more likely to select the inconsistent distractor during viewing when it was bigger compared to when it was smaller than the consistent distractor, $\beta = 0.796$, $SE = 0.319$, $z = 2.50$, $p = 0.013$, the probability of fixating the consistent distractor did not depend on its difference in size with respect to the inconsistent distractor, $\beta = 0.064$, $SE = 0.313$, $z < 1$, $p = 0.839$.

No other interaction was significant, all zs ≤ 1.17, all ps ≥ 0.242.

Foveal processing of object-scene inconsistencies

In the above analyses we looked for evidence that inconsistency influenced when an object was selected during search. In the analyses that follow we considered whether there was any evidence that spatial inconsistency resulted in different foveal processing of the objects once fixated. Specifically, we examined how long the first fixation on an object lasted, the total time spent fixating an object during search and the number of times an object was fixated as measures of foveal processing. In the following LMMs, type of template and type of distractor (inconsistent vs. consistent) and their interactions were entered as fixed factors. Please note that we considered all the six consistent distractors included in the scene region that was expected for the target, regardless of their size. These models are based on all trials with a switched arrangement, in which at least one object was fixated in the target expected region. Following inspection of the distribution and residuals, total fixation duration on an object was log-transformed in order to meet LMM assumptions.

First fixation duration on object: The duration of the first fixation made on an (distractor) object in the target expected region was marginally influenced by the type of distractor, $\beta = -12.955$, $SE = 6.637$, $t = -1.95$, $p = 0.052$, being shorter on a consistent distractor (160 ms) than on the inconsistent distractor (174 ms). Neither the main effect of the type of template, $\beta = -12.481$, $SE = 11.448$, $t = -1.09$, $p = 0.276$, nor the two-way interaction, $\beta = -5.003$, $SE = 13.115$, $t < 1$, $p = 0.703$, was significant (Figure 9, left panel).

Total fixation duration (dwell time) on object: The total fixation duration on a distractor in the target expected region was longer following a word cue (206 ms) than following a picture cue (176 ms), $\beta = 0.077$, $SE = 0.036$, $t = 2.16$, $p = 0.032$. In addition, significantly more time was spent fixating the inconsistent distractor (218 ms) than a consistent distractor (184 ms), $\beta = 0.046$, $SE = 0.022$, $t = 2.12$, $p = 0.035$. The two-way interaction was not significant, $\beta = 0.042$, $SE = 0.042$, $t < 1$, $p = 0.322$ (Figure 9, central panel).

Number of fixations on object: Neither the type of template, $\beta = 0.057$, $SE = 0.050$, $t = 1.14$, $p = 0.255$, nor the type of distractor, $\beta = 0.063$, $SE = 0.046$, $t = 1.42$, $p = 0.157$, had a significant effect on the mean number of fixations per distractor object in the target expected region. The two-way interaction was not significant, $\beta = -0.041$, $SE = 0.088$, $t < 1$, $p = 0.642$ (Figure 9, right panel).

Discussion

The aim of this study was to tease apart the disruptive effect on visual search related to misleading expectations concerning target location from possible influences on ocular selection and foveal processing of the inconsistency between a given object and the scene context in which it is placed. We also examined whether either of these effects may be modulated by differences in the strength of guidance due to the type of target template.

Searching for implausibly placed targets

Our manipulation of scene arrangement did not introduce any peculiar effect in relation to what is usually reported when comparing search for consistent and inconsistent targets. Indeed, we replicated previous studies showing that a precise template and reliable spatial expectations about the target facilitate search and improve efficiency of oculomotor behavior (e.g., Bravo & Farid, 2009; Castelhano et al., 2008; Castelhano & Heaven, 2010; Eckstein et al., 2006; Malcolm & Henderson, 2009, 2010; Maxfield & Zelinsky, 2012; Neider & Zelinsky, 2006; Schmidt & Zelinsky, 2009, 2011; Võ & Henderson, 2009, 2011; Wolfe et al., 2004). We found this enhancement across all phases of the search process: initiation, scene scanning, and target object verification.

Although placing an object in an implausible location increased the difficulty of searching for it, an implausibly placed target was immediately fixated or immediately saccaded to with a considerably higher probability than the other object (distractor) having an implausible position within the same scene. This was true even though spatial expectations might guide the eyes toward this distractor (because it was always placed where the target might be expected). As both objects violated the normal relationship with scene context, being thus equated in terms of inconsistency, this finding indicates that prior information about target features gives stronger early guidance than predictions about where to find the target. Coherent with the suggestion that the template operates by enhancing selection of local features in the image that match target features, having a precise template increased specifically initial guidance toward the spatially inconsistent target, whilst it did not influence the probability of initially selecting the spatially inconsistent distractor.

Isolating effects of implausibility on search

Selection priorities for inconsistently placed objects

When the scene had a switched arrangement (with the target implausibly placed), the inconsistent distractor was selected with higher probability than the reference (i.e., biggest) consistent distractor in the target

expected region. This selection advantage was found during early inspection of the region, and persisted to some extent later in scene viewing: The inconsistent distractor was more likely to be the first selected object in the region, targeted by the first saccade and fixated at some later point in viewing. However, the advantage for selection, in all the measures considered, was shown only when the inconsistent distractor was bigger than the reference consistent distractor. Note that this also means that, in that case, the inconsistent distractor was the biggest object in the region. Importantly, the same pattern of findings did not coherently emerge for the consistent distractor when it was bigger than the inconsistent distractor (and, thus, when it was the biggest object in the region as well). In this condition, we only found a tendency for this object of being the first one fixated in the region. This probability of first object fixation did not differ from chance, whereas a bigger inconsistent object was first selected significantly above chance. No indication of any prioritization of a bigger consistent distractor over the inconsistent distractor was obtained when taking into account the initial saccade in the region or the overall probability of being selected by the eyes during scene viewing. In addition, the likelihood of selecting the inconsistent distractor increased considerably when it was bigger than the reference consistent distractor compared to when it was smaller, whereas the likelihood of selecting the consistent distractor appeared independent of its size difference with respect to the inconsistent distractor. These differences are unlikely to have arisen from any imbalance in the relative sizes of inconsistent and consistent distractors because, first, the differences in size between the two objects were matched when the inconsistent was the bigger compared to when the consistent distractor was the bigger and, second, because our use of linear mixed effect models allowed us to take into account any variability between scenes in our results, in order to be able to exclude an important role of idiosyncrasies present within our material. Our findings were also not a result of any differences in the eccentricity of the tested objects. It seems reasonable to assume, therefore, that our findings arose from a genuine interaction between perceptual and cognitive factors in fixation selection during scene search, and that this interaction led to a preferential inspection of objects violating scene associations if they were larger than objects co-occurring in the scene region where attention had been oriented.

This study suggests indeed that stimulus-driven/exogenous and cognitive/endogenous aspects concur to engender an integrated priority map (e.g., Awh, Belopolsky, & Theeuwes, 2012; Macaluso & Doricchi, 2013) that guides the visual system during search. Therefore, while highlighting the importance of understanding of object-context relationships for scene inspection, it challenges a rigid interpretation of cognitive relevance theories of eye movement guidance (e.g., Einhäuser, Spain, & Perona, 2008; Henderson, Malcolm, & Schandl, 2009; Hwang, Higgins, & Pomplun, 2009; Nyström & Holmqvist, 2008). We suggest that a first, perceptual filter acts on midlevel processes associated with figure-ground segregation and object-based attention: The biggest object in the region is initially preselected by covert attention and, as a second step, preferential overt selection is made upon cognitive factors, guiding the eyes to inconsistent objects.

This account is based on, and gives support to, two assumptions. First of all, it posits that attention is (covertly and overtly) essentially allocated to objects (e.g., Clarke, Coco, & Keller, 2013; Einhäuser et al., 2008; Nuthmann & Henderson, 2010; Quiles, Wang, Zhao, Romero, & Huang, 2011; Xu et al., 2014), with object size as a guiding attribute for selection during scene viewing (e.g., Borji et al., 2013; Clarke et al., 2013; Einhäuser et al., 2008; Quiles et al., 2011) and visual search (Wolfe & Horowitz, 2004), being especially relevant in early inspection (Xu et al., 2014). Second, it implies that objects in extrafoveal (i.e., blurred) vision are processed to a sufficient extent to allow some understanding of their relationship with the whole scene, as claimed by previous eye movement studies showing an inconsistency advantage in ocular selection (Becker et al., 2007; Bonitz & Gordon, 2008; Brockmole & Henderson, 2008; Loftus & Mackworth, 1978; Underwood et al., 2007, 2008).Concerning the time course of the effect, previous research with different tasks has sometimes reported an early prioritization of inconsistent objects (Brockmole & Henderson, 2008; Gordon, 2004, 2006; Loftus & Mackworth, 1978) as we did. Nevertheless, several studies have also shown that the inconsistency advantage for selection only emerged (e.g., Becker et al., 2007; Underwood & Foulsham, 2006; Underwood et al., 2008) or was however enhanced (Brockmole & Henderson, 2008; Loftus & Mackworth, 1978) after several fixations on the scene. We believe that the principal cause of this discrepancy with our findings is related to the type of comparison examined in our study: The fact that the inconsistent distractor preferentially selected was the biggest object in the region favored its quick selection on a perceptual basis and, therefore, the early emergence of the advantage over smaller consistent objects; indeed when the inconsistent distractor was not the largest in the region, preferential selection was not found.

We also hypothesize that our analysis of the difference in size between the inconsistent object and the cohort of consistent objects, together with our specific experimental design (controlling for the effects of target template and expectations for target place-

ment), is the key factor explaining why we found evidence of extrafoveal processing and preferential ocular selection of inconsistency, while all the other studies on visual scene search did not report such effects. These studies, indeed, usually limited analyses to comparing between consistent and inconsistent target conditions, for which size was equated (Castelhano & Heaven, 2011; Eckstein et al., 2006; Henderson et al., 1999; Malcolm & Henderson, 2010; Võ & Henderson, 2009; 2011), and never considered the impact of size differences between the critical inconsistent objects and other (consistent) distractor objects included in the scene (see also De Graef et al., 1990; Underwood & Foulsham, 2006).

In other paradigms, preferential selection of inconsistency has emerged even when comparing inconsistent and consistent objects matched for size (e.g., memorization tasks: Brockmole & Henderson, 2008; Gordon, 2004; Underwood & Foulsham, 2006; spot-the-difference task: Underwood et al., 2008; free-viewing: Becker et al., 2007; image rating: Bonitz & Gordon, 2008). Size differences seem therefore particularly important in search, where constraints imposed by the task are very high because of the specification of a particular target and, implicitly, of a set of particular task-relevant locations (see Mills, Hollingworth, Van der Stigchel, Hoffman, & Dodd, 2011). Eye attraction by inconsistency in search might be seen as a transitory deviation from task goals, which occurs mainly when the inconsistent object is perceptually preselected according to its (relatively bigger) size. We speculate that this is aimed at reaching better understanding of the whole scene when the eyes are not guided optimally by target appearance and spatial expectations.

It is necessary to keep in mind, indeed, that all the above considerations apply only to about 60% of the trials with a switched arrangement, in which guidance by matching between target template and target features in the scene was not strong enough to guide the eyes immediately toward the region unexpectedly containing the target object. This was more likely to happen with an abstract, verbal template than with a precise, pictorial template (see the section Searching for implausibly placed targets). The fact that following a word cue, compared to a picture cue, we reported higher probability of fixating an object (regardless of its spatial consistency) during inspection of the target expected region in switched arrangements may be another consequence of less effective guidance supplied by abstract target representations.

Differences in foveal processing for inconsistently placed objects

Once an object has been selected for fixation, spatial inconsistency with the local context may result in different foveal engagement with the object. This was indeed what we found: All duration-based measures were higher for the inconsistent distractor than for consistently placed distractors in the same region. An increased duration of object inspection has been reported by almost all previous studies that examined the effects of scene violation on oculomotor behavior (see Coco et al., 2013, for an exception), and has been explained in terms of either object identification difficulties or conflict with respect to scene understanding (see Gordon, 2004, 2006). The point of debate concerns the temporal dynamics of the influence of inconsistency on foveal processing. We found evidence to indicate an early influence of inconsistency on foveal processing (although the effect was marginal), supporting previous findings that inconsistency results in longer first fixations (Bonitz & Gordon, 2008; Castelhano & Heaven, 2011; De Graef et al., 1990; Underwood et al., 2008; Võ & Henderson, 2009). Other studies have argued, instead, that the effect of inconsistency only emerges in measures of foveal processing time across multiple fixations, suggesting a relatively late effect (Becker et al., 2007; Henderson et al., 1999; Rayner et al., 2009). In these studies, however, while the effect was not found for the first fixation on an inconsistent object, it was found for the first inspection of that object comprising several successive fixations on it (the first pass gaze duration), before the eyes were directed elsewhere (see also Friedman, 1979).

Taking into account our findings on extrafoveal processing, our study suggests, overall, that as soon as an object violating scene context has been (partially) detected, it is prioritized during viewing until it is excluded as a potential target. However, we did not show any influence of consistency on the number of fixations on an object. This appears discrepant with most of previous research, which showed more fixations on inconsistent objects (Becker et al., 2007; Bonitz & Gordon, 2008; Henderson et al., 1999; Loftus & Mackworth, 1978; Rayner et al., 2009; Võ & Henderson, 2009; but see Friedman, 1979, for a study that did not report this effect). We speculate that our result may depend on the limited number of fixations made during each trial and on the type of task, which may have discouraged multiple fixations on objects that did not match template features.

Some considerations on the generalizability of our results

When taking into account the generalizability of our findings beyond the present study, the type of material we used deserves careful consideration. The scenes were made by adding photographs of objects into real-

world photographs of scene contexts. While therefore they respected the general organization and many aspects of real-world scenes, object insertions might have altered of some properties characterizing our usual visual experience. The need for equating the number of objects in each scene region may have resulted in deviations from the number and distribution of consistent objects that one may reasonably expect in a real environment. Other alterations may have occurred at a perceptual level (like depth cues, shadows), especially in the switched scene arrangement. This is despite the fact that in the pilot study the insertions of the two critical objects were judged of similarly good quality in both scene arrangements (see Materials).

It is worth considering the implications that the deviations from real-world scenes might have, in particular, on the influence of inconsistency. One might speculate that in more natural images reliance on contextual expectations would be stronger, and consequently any effect related to inconsistency with scene context would be enhanced. However, it has been shown that eye movement behavior may not differ substantially between very dissimilar types of scenes, like when comparing viewing of full-color real-world photographs with viewing of line drawings (Henderson & Hollingworth, 1998). More specifically for what concerns object-scene violations, studies have used scenes that vary greatly in terms of their realism, from line drawings to photographs of real world scenes, with either critical objects inserted a posteriori using an editing software or already placed into the scene when the photograph was taken. Nevertheless, results do not appear to depend upon the scene realism, with evidence for attentional engagement and attentional disengagement being found across all levels of realism. Therefore, although the issue has never been systematically investigated, it seems unlikely that the overall deviation of the type of scenes we used from real-world images had substantial influence on our results concerning object-context violations.

We may also consider in particular how the types of violation that involved the inconsistent objects with respect to the local background in which they were placed may have affected our findings, especially regarding prioritization for selection of inconsistently placed distractors. A key role of the violations involving depth cues and perspective, which might lead to detect inconsistent distractors on the basis of physical characteristic, was ruled out by obtaining overall similar patterns of results in the additional analyses conducted only in the subset of scenes that did not present this form of geometrical inconsistency. Two other kinds of local syntactic violations may have been at play: proportionality of size and support (see Biederman, 1977; Biederman et al., 1982). They might have contributed to prioritization of the inconsistent objects in several of our scenes. In any case, both are deviations from rules of our experience with scenes and object-scene relationships, like inconsistency with respect to the probability of object occurrence in a given scene region is. We assume therefore that similar processes may be involved in their detection: As these inconsistencies are defined with respect to our knowledge of the world, they may all be considered, at their core, inherently semantic. Future studies could however be devoted at distinguishing their specific contributions to selection prioritization.

Conclusions

The present study is the first to provide evidence for an influence of spatial inconsistency in object placement that is separable from the effects of template precision and spatial expectations about the target on search guidance. When inconsistency of object placement was controlled for, template and spatial expectations both guided behavior throughout all search processes: When searching for a misplaced target, observers explored the region in which they expected to find it on just over half of the trials, and this was more likely when searching with a verbal template than with a pictorial template. We were able to use these cases, where the region in which the target was expected to occur was searched, to isolate effects of spatial inconsistency, while controlling for guidance by template and spatial expectations. We found that inconsistency of object placement resulted in a relative higher probability that the object was prioritized for fixation selection during early inspection of the region, and this preferential selection tended to persist during viewing. Crucially, this was reported only when the misplaced object was the biggest the region, suggesting that prioritization arises from an initial attentional bias to objects according to their relative size, followed by decision on saccade targeting based upon (extrafoveal) detection of object-context violations. Moreover, our study confirms that inconsistency leads to longer foveal inspection of objects, even during the first fixation on the object. The possibility of isolating an effect of spatial inconsistency during search in realistic scenes enables researchers to have a more precise and reliable measure of the impacts of target template and expectations about target location on oculomotor behavior when observers look for an implausibly placed object.

Keywords: eye movements, visual search, spatial consistency, target template, context information

Acknowledgments

This research was funded by ESRC grant RES-000-22-4098 to BWT.

Commercial relationships: none.
Corresponding author: Sara Spotorno.
Email: s.spotorno@dundee.ac.uk.
Address: School of Psychology, University of Dundee, Park Place, Dundee, Scotland, UK.

Footnote

[1] The additional analyses, carried out without the 14 scenes in which the inconsistent distractor was not coplanar with its background, revealed only three slight differences compared to the main analyses reported for extrafoveal processing of object-scene inconsistencies (see the section Extrafoveal processing of object-scene inconsistencies):
(1) When considering whether the inconsistent or the "reference" consistent distractor was the first object fixated in the target expected region, we found that when the reference consistent distractor was the biggest object in the region, it was selected significantly more (0.18) than the inconsistent distractor (0.09), $\beta = 0.783$, $SE = 0.381$, $z = 2.06$, $p = 0.040$. In the main analysis we did not report such a difference. Nevertheless, this probability of first selecting the (bigger) consistent distractor did not differ from chance level, $t(115) = 1.06$, $p = 0.145$.
(2) As for the probability of object fixation during viewing, we found a greater probability of fixating the inconsistent distractor (0.52) compared to the reference consistent distractor (0.38), when it was the biggest object in the region, $\beta = 0.716$, $SE = 0.348$, $z = 2.06$, $p = 0.040$. In the main analysis, instead, this effect emerged only as a tendency, in the same direction.
(3) We also found, regarding the probability of object fixation during viewing, a tendency toward greater probability of fixating the consistent distractor (0.40), compared to the inconsistent distractor (0.31), when it was the biggest object in the region, $\beta = 0.515$, $SE = 0.293$, $z = 1.76$, $p = 0.078$. In the main analysis we did not report this tendency.

References

Awh, E., Belopolsky, A.V., & Theeuwes, J. (2012). Top-down versus bottom-up attentional control: A failed theoretical dichotomy. *Trends in Cognitive Sciences, 16,* 437–443.

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2014). lme4: Linear mixed-effects models using Eigen and S4 (R package Version 1.1-5). Retrieved from http://cran.r-project.org/package=lme4

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.

Biederman, I. (1977). On processing information from a glance at a scene. Some implications for a syntax and semantics of visual processing. In S. Treu (Ed.), *User-oriented design of interactive graphic systems.* New York: ACM.

Biederman, I., Mezzanotte, R. J., & Rabinowitz, J. C. (1982). Scene perception: Detecting and judging objects undergoing relational violations. *Cognitive Psychology, 14*(2), 143–177.

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

Borji, A., Sihite, D. N., & Itti, L. (2013). What stands out in a scene? A study of human explicit saliency judgment. *Vision Research, 91,* 62–77.

Bravo, M. J., & Farid, H. (2009). The specificity of the search template. *Journal of Vision, 9*(1):34, 1–9, http://www.journalofvision.org/content/9/1/34, doi:10.1167/9.1.34. [PubMed] [Article]

Brockmole, J. R., & Henderson, J. M. (2008). Prioritizing new objects for eye fixation in real-world scenes: Effects of object-scene consistency. *Visual Cognition, 16*(2/3), 375–390.

Castelhano, M. S., & Heaven, C. (2010). The relative contribution of scene context and target features to visual search in real-world scenes. *Attention, Perception, & Psychophysics, 72*(5), 1283–1297.

Castelhano, M. S., & Heaven, C. (2011). Scene context influences without scene gist: Eye movements guided by spatial associations in visual search. *Psychonomic Bulletin & Review, 18*(5), 890–896.

Castelhano, M. S., Pollatsek, A., & Cave, K. R. (2008). Typicality aids search for an unspecified target, but only in identification and not in attentional guidance. *Psychonomic Bulletin & Review, 15*(4), 795–801.

Clarke, A. D. F., Coco, M. I., & Keller, F. (2013). The impact of attentional, linguistic, and visual features during object naming. *Frontiers in Psychology, 4, 297,* 1–12.

Coco, M. I., Malcolm, G. L., & Keller, F. (2013). The interplay of bottom-up and top-down mechanisms in visual guidance during object naming. *Quarterly Journal of Experimental Psychology, 14,* 1–25.

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

Demiral, S. 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,* 1271–1285.

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.

Ehinger, K. A., Hidalgo-Sotelo, B., Torralba, A., & Oliva, A. (2009). Modeling search for people in 900 scenes: A combined source model of eye guidance. *Visual Cognition, 17*(6/7), 945–978.

Einhäuser, W., Spain, M., & Perona, P. (2008). Objects predict fixations better than early saliency. *Journal of Vision, 8*(14):18, 1–26, http://www.journalofvision.org/content/8/14/18, doi:10.1167/8.14.18. [PubMed] [Article]

Friedman, A. (1979). Framing pictures: The role of knowledge in automatized encoding and memory for gist. *Journal of Experimental Psychology: General, 108*(3), 316–355.

Gareze, L., & Findlay, J. M. (2007). Absence of scene context effects in object detection and eye gaze capture. In R. P. G. van Gompel, M. H. Fischer, W. S. Murray, & R. L. Hill (Eds.), *Eye movements: A window on mind and brain.* (pp. 618–637). Amsterdam: Elsevier.

Gordon, R. (2004). Attentional allocation during the perception of scenes. *Journal of Experimental Psychology: Human Perception and Performance, 30*(4), 760–777.

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

Henderson, J. M., & Hollingworth, A. (1998). Eye movements during scene viewing: An overview. In G. Underwood (Ed.), *Eye guidance in reading and scene perception* (pp. 269–293). Oxford: Elsevier.

Henderson, J. M., Malcolm, G. L., & Schandl, C. (2009). Searching in the dark: Cognitive relevance drives attention in real-world scenes. *Psychonomic Bulletin and Review, 16,* 850–856.

Henderson, J. M., Weeks, P. A., & Hollingworth, A. (1999). The effect of semantic consistency on eye movements during complex scene viewing. *Journal of Experimental Psychology: Human Perception and Performance, 25*(1), 210–228.

Hollingworth, A., & Henderson, J. M. (1998). Does consistent scene context facilitate object perception? *Journal of Experimental Psychology: General, 127*(4), 398–415.

Hollingworth, A., & Henderson, J. M. (1999). Object identification is isolated from scene semantic constraint: Evidence from object type and token discrimination. *Acta Psychologica, 102*(2-3), 319–343.

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.

Hwang, A. D., Higgins, E. C., & Pomplun, M. (2009). A model of top-down attentional control during visual search in complex scenes. *Journal of Vision, 9*(5):25, 1–18, http://www.journalofvision.org/content/9/5/25, doi:10.1167/9.5.25

Kanan, C., Tong, M. H., Zhang, L., & Cottrell, G. W. (2009). SUN: Top-down saliency using natural statistics. *Visual Cognition, 17*(6/7), 979–1003.

Kliegl, R., Masson, M. E. J., & Richter, E. M. (2010). A linear mixed model analysis of masked repetition priming. *Visual Cognition, 18,* 655–681.

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.

Macaluso, E., & Doricchi, F. (2013). Attention and predictions: Control of spatial attention beyond the endogenous-exogenous dichotomy. *Frontiers in Human Neuroscience, 7, 685,* 1–12.

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, 1–16, http://www.journalofvision.org/content/11/9/9, doi:10.1167/11.9.9. [PubMed] [Article]

Malcolm, G. L., & Henderson, J. M. (2009). The effects of target template specificity on visual search in real-world scenes: Evidence from eye movements. *Journal of Vision, 9*(11):8, 1–13, http://www.journalofvision.org/content/9/11/8, doi:10.1167/9.11.8. [PubMed] [Article]

Malcolm, G. L., & Henderson, J. M. (2010). Combining top-down processes to guide eye movements during real-world scene search. *Journal of Vision, 10*(2):4, 1–11, http://www.journalofvision.org/content/10/2/4, doi:10.1167/10.2.4. [PubMed] [Article]

Maxfield, J. T., & Zelinsky, G. J. (2012). Searching through the hierarchy: How level of target catego-

rization affects visual search. *Visual Cognition, 20*(10), 1153–1163.

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, 1–15, http://www.journalofvision.org/content/11/8/17, doi:10.1167/11.8.17. [PubMed] [Article]

Mudrik, L., Deouell, L. Y., & Lamy, D. (2011). Scene congruency biases binocular rivalry. *Consciousness and Cognition, 20*(3), 756–767.

Neider, M. B., & Zelinsky, G. J. (2006). Scene context guides eye movements during visual search. *Vision Research, 46*(5), 614–621.

Nuthmann, A., & Henderson, J. M. (2010). Object-based attentional selection in scene viewing. *Journal of Vision, 10*(8):20, 1–19, http://www.journalofvision.org/content/10/8/20, doi:10.1167/10.8.20. [PubMed] [Article]

Nyström, M., & Holmqvist, K. (2008). Semantic override of low-level features in image viewing—Both initially and overall. *Journal of Eye Movement Research, 2,* 1–11.

Potter, M. C. (1976). Short-term conceptual memory for pictures. *Journal of Experimental Psychology: Human Learning and Memory, 2,* 509–522.

Quiles, M. G., Wang, D., Zhao, L., Romero, R. A. F., & Huang, D. (2011). Selecting salient objects in real scenes. An oscillatory correlation model. *Neural Networks, 24*, 54–64.

Rayner, K., Castelhano, M., & Yang, J. (2009). Eye movements when looking at unusual/weird scenes: Are there cultural differences? *Journal of Experimental Psychology: Learning Memory and Cognition, 35,* 254–259.

Schmidt, J., & Zelinsky, G. J. (2009). Search guidance is proportional to the categorical specificity of a target cue. *Quarterly Journal of Experimental Psychology, 62*(10), 1904–1914.

Schmidt, J., & Zelinsky, G. J. (2011). Visual search guidance is best after a short delay. *Vision Research, 51,* 535–545.

Spotorno, S., Malcolm, G. L., & Tatler, B. W. (2014). How context information and target information guide the eyes from the first epoch of search in real-world scenes. *Journal of Vision, 14*(2):7, 1–21, http://www.journalofvision.org/content/14/2/7, doi:10.1167/14.2.7. [PubMed] [Article]

Spotorno, S., Tatler, B.W., & Faure, S. (2013). Semantic consistency versus perceptual salience in visual scenes: Findings from change detection. *Acta Psychologica, 142*(2), 168–176.

Stirk, J. A., & Underwood, G. (2007). Low-level visual saliency does not predict change detection in natural scenes. *Journal of Vision, 7*(10):3, 1–10, http://www.journalofvision.org/content/7/10/3, doi:10.1167/7.10.3. [PubMed] [Article]

Tatler, B. W. (2007). The central fixation bias in scene viewing: Selecting an optimal viewing position independently of motor biases and image feature distributions. *Journal of Vision, 7*(14):4, 1–17, http://www.journalofvision.org/content/7/14/4, doi:10.1167/7.14.4. [PubMed] [Article]

Tatler, B. W., Hayhoe, M. M., Land, M. F., & Ballard, D. H. (2011). Eye guidance in natural vision: Reinterpreting salience. *Journal of Vision, 11*(5):5, 1–23, http://www.journalofvision.org/content/11/5/5, doi:10.1167/11.5.5. [PubMed] [Article]

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

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. 89–110). Oxford: Elsevier.

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.

Vickery, T. J., King, L. W., & Jiang, Y. (2005). Setting up the target template in visual search. *Journal of Vision, 5*(1):8, 81–92, http://www.journalofvision.org/content/5/1/8, doi:10.1167/5.1.8. [PubMed] [Article]

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://www.journalofvision.org/content/9/3/24, doi:10.1167/9.3.24. [PubMed] [Article]

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). The interplay of episodic and semantic memory in guiding repeated search in scenes. *Cognition, 126,* 198–212.

Wickham, H. (2009). *ggplot2: Elegant graphics for data analysis*. Springer: New York.

Wolfe, J., & Horowitz, T. (2004). What attributes guide

the deployment of visual attention and how do they do it? *Nature Reviews Neuroscience, 5,* 1–7.

Wolfe, J. M., Horowitz, T. S., Kenner, N., Hyle, M., & Vasan, N. (2004). How fast can you change your mind? The speed of top-down guidance in visual search. *Vision Research, 44,* 1411–1426.

Xu, J., Jiang, M., Wang, S., Kankanhalli, M. S., & Zhao, Q. (2014). Predicting human gaze beyond pixels. *Journal of Vision, 14*(1):28, 1–20, http://www.journalofvision.org/content/14/1/28, doi:10.1167/14.1.28. [PubMed] [Article]

Yang, H., & Zelinsky, G. J. (2009). Visual search is guided to categorically-defined targets. *Vision Research, 49,* 2095–2103.

Zelinsky, G. J. (2008). A theory of eye movements during target acquisition. *Psychological Review, 115*(4), 787–835.