The Elephant in the Room: Inconsistency in Scene Viewing and Representation

Sara Spotorno

Institute of Neuroscience and Psychology, University of Glasgow

Benjamin W. Tatler

School of Psychology, University of Aberdeen

Author Note

The authors contributed equally to all aspects of this manuscript and the work contained within it.

We are very grateful to Andrew McKechnie for conducting a pilot study that informed the design and analysis of the data presented in Experiment 3 of the present study. For assistance with data collection for Experiments 1 and 2 we thank Ilaria Dal Lago, Anais Leroy, Teodor Nikolov, Hannah Nsiah-Amoako, Hanane Ramzaoui, Lara Wilson and Radia Zaghari. We also thank Monica Castelhano, Carrick Williams and an anonymous reviewer for thorough and helpful suggestions to improve this manuscript.

Correspondence concerning this article should be addressed to Benjamin W. Tatler, School of Psychology, University of Aberdeen, Aberdeen AB24 3FX, Scotland, UK

Contact: b.w.tatler@abdn.ac.uk

Abstract

We examined the extent to which semantic informativeness, consistency with expectations and perceptual salience contribute to object prioritisation in scene viewing and representation. In scene viewing (Experiments 1-2), semantic guidance overshadowed perceptual guidance in determining fixation order, with the greatest prioritisation for objects that were diagnostic of the scene's depicted event. Perceptual properties affected selection of consistent objects (regardless of their informativeness) but not of inconsistent objects. Semantic and perceptual properties also interacted in influencing foveal inspection, as inconsistent objects were fixated longer than low but not high salience diagnostic objects. While not studied in direct competition with each other (each studied in competition with diagnostic objects), we found that inconsistent objects were fixated earlier and for longer than consistent but marginally informative objects. In change detection (Experiment 3), perceptual guidance overshadowed semantic guidance, promoting detection of highly salient changes. A residual advantage for diagnosticity over inconsistency emerged only when selection prioritisation could not be based on low-level features. Overall these findings show that semantic inconsistency is not prioritised within a scene when competing with other relevant information that is essential to scene understanding and respects observers' expectations. Moreover, they reveal that the relative dominance of semantic or perceptual properties during selection depends on ongoing task requirements.

Keywords:

Semantic Consistency, Perceptual Salience, Scene Viewing, Change Detection, Eye Movements

Public significance statement

There has been long-standing debate about whether we look sooner at objects that are unexpected for the scene (i.e., semantically inconsistent objects). The present study shows that they are not prioritised over the most expected and informative objects for the event depicted in the scene (i.e., diagnostic objects), especially when these objects also stand out visually from their surroundings. Unexpected objects are looked at later, and when they are perceptually salient they slow down how quickly we select the most expected and informative objects in the scene. This study also shows that the impacts of object perceptual salience and object-scene semantic associations are task-dependent: semantics appear the most important source of guidance when viewers explore the scene for memorisation, whereas perceptual salience has a greater impact when changes have to be found. These findings provide new key insights into the roles of high- and low-level factors in how we view and remember scenes.

The Elephant in the Room: Inconsistency in Scene Viewing and Representation

Information selection during both online processing and memory representation is one of our fundamental and most striking abilities. Fundamental, because selection of a few, key aspects enables us to adapt and act flexibly in everyday life, where we have to face rich, noisy and changing environments. Striking, because it deals with an overwhelming flow of information (estimated in 10 billion bits per second, only considering the rate of data transmission from the retina, see Koch et al., 2006), where we are exposed to multiple high-level and low-level information sources at the same time. If we are to understand how we select information from the world around us to serve ongoing behaviour, it is therefore central to understand what dimensions contribute to prioritising particular information amongst competing sources.

There remains a long-standing debate about how prioritisation of objects for inclusion, maintenance and availability in visual representations takes place. In particular, while most previous research suggests that perceptual and semantic factors interact in affecting our viewing behaviour and memory (but see Kollmorgen, Nortmann, Schröder & Köning, 2010), it is still unclear how this interplay takes place.

One view is that perceptual salience has a predominant influence: selection in scenes proceeds sequentially from the most salient point, following a hierarchy of salience weights (e.g., Itti & Koch, 2000; Koch & Ullman, 1985; Parkhurst, Law & Niebur, 2002). While consensus has almost been reached about the inappropriateness of models strictly based on pixel-to-pixel differences in low-level features like luminance, colour or orientation to account for human behaviour (Tatler, Hayhoe, Land & Ballard, 2011, for review; but see Borji & Itti, 2013, and Latif, Gehmacher & Castelhano, 2014), one of the main unresolved issues is about what happens when the inclusion of features within objects is considered (e.g., Borji, Sihite & Itti, 2013a; Einhäuser, Spain & Perona, 2008a, Nuthmann & Henderson, 2010; Pajak & Nuthmann, 2013; Xu, Jiang, Wang, Kankanhalli & Zhao, 2014). In this framing, it might still be that processing of objects follows the sequential rule of perceptual priority (e.g., Underwood, Humphreys & van Loon, 2011). Some studies that manipulated perceptual and semantic dimensions orthogonally (Coco, Malcolm & Keller, 2013; Pringle, Irwin, Kramer & Atchley, 2001; Spotorno & Faure, 2011; Underwood & Foulsham, 2006) indicated that when an object is highly salient the effect of semantics does not emerge or, at least, is strongly reduced.

A second view suggests, on the contrary, that semantic factors predominate and may completely override perceptual salience. Semantically relevant objects may indeed be systematically prioritised even when of lower salience than semantically marginal ones (e.g., Chen & Zelinsky, 2006; Einhäuser, Rutishauser & Koch, 2008b; Stirk & Underwood, 2007; Tatler et al., 2011; Underwood, Templeman, Lamming & Foulsham, 2008), and this may be the case from the early moments of scene inspection (e.g., Henderson, Malcolm & Schandl, 2009; Nyström & Holmqvist, 2008, but see Dombrowe, Olivers & Donk, 2010). However, it has been controversial what type of semantic informativeness of an object is the most influential on information gathering from a scene. Since the pioneering work by Loftus and Mackworth (1978), in particular, we do know that attentional allocation to an object during scene viewing and representation may depend on its semantic relationship with the scene context in which it is included. However, previous literature has been contradictory on whether strong agreement with expectations about object occurrence in the scene (e.g., Coco et al., 2013; O'Regan, Deubel, Clark & Rensink, 2000; Pringle et al., 2001; Rensink, O'Regan & Clark, 1997, 2000; Spotorno & Faure, 2011) or, on the contrary, strong disagreement (inconsistency) with expectations about object occurrence in the scene (e.g., Bonitz & Gordon, 2008; Brockmole & Henderson, 2008; Cornelissen & Võ, 2017; Hollingworth & Henderson, 2000, 2003; Loftus & Mackworth, 1978; Stirk & Underwood, 2007; Underwood et al., 2007, 2008) results in greater object prioritisation in terms of earlier selection. The studies reviewed above suggest that some semantic processing must be going on prior to foveal inspection of the object in order for that object to be selected sooner than others. The idea that semantic information may be processed extrafoveally, prior to fixation, and may inform fixation selection has been suggested previously (see Tatler, Brockmole and Carpenter, 2017).

Objects can have a powerful role in our understanding of the scene's core conceptual content (i.e., the scene's gist: see Biederman, 1972; Oliva, 2005; Potter, 1975) either by being "diagnostic" (Schyns, 1998) in constructing and confirming scene understanding or by being inconsistent, thus violating scene expectations and challenging the initial, potentially incorrect, interpretation of the whole image (e.g., Friedman, 1979; Palmer, 1975; Spotorno, Tatler & Faure, 2013). While both may be highly informative for the semantic understanding the scene, the underlying factors that confer this informativeness are fundamentally different.

A diagnostic object is one that is not only consistent with the meaning of the scene, but is essential for that meaning – conveying and catalyzing the scene's gist through its relationship with the co-occurring objects. In every scene representing a real-world situation with agent(s) involved in an action, diagnostic objects become semantically important in conveying the sense of the depicted event, providing the essential description of the situation (e.g., "a girl looking at a flower she has picked up", "a man trying to fix a broken tap", see images A and C in Figure 1); rather than merely informing a categorical description of the scene (e.g., "a meadow", "a kitchen" for images A and C in Figure 1). This functional understanding of a scene, centred on what we may call the gist for action, is what is crucial in everyday life, and recently Greene, Baldassano, Esteva, Beck, & Fei-Fei (2016) suggested that it may be the primary principle of scene categorisation. Object diagnosticity contributes to choosing among distinctive (main) event interpretations of the scene, and arises largely from an object's relationship with other objects in the scene, which are part of the depicted event. It is the result of a specific semantic network within the scene and is tied to the placement of the object. For instance, the diagnostic flower in image A (Figure 1) would be no more important than all the other flowers in the image if it were placed in the meadow instead of in the girl's hand. It is the pairing of the diagnostic object's identity with its location that confers the particular meaning to the scene's event; without a flower in that location, the core meaning of the scene would be more about a happy girl on a sunny day. In a similar way, the spanner in image C would be less important for the scene if placed somewhere on the worktops, and the scene would not offer any strong hint about the intention of the man to repair the tap if the spanner were not in his hand. It is clear from these examples that both the identity of a diagnostic object and the location at which it occurs are informative in their own right – the object maintains some informativeness elsewhere in the scene and the location remains important to the scene even when empty – but neither is sufficient to provide the diagnostic understanding of the scene's depiction. Diagnostic objects may therefore be prioritised in scene perception due to this crucial role that they play in defining the semantic interpretations of the scene: a role that arises from the unique combination of the identity of the object and its placement in the scene.

FIGURE 1 ABOUT HERE

Informativeness of the inconsistent object arises, on the contrary, from the absence of plausible relationships with the co-occurring objects and with the scene context, and contributes to the attribution of a plausible category to the scene by challenging the interpretation offered by the rest of the scene. To continue with some concrete examples, the microphone, the balloon, the rubber ring and the painting in Figure 1 are violating what we would expect in those types of scene categories, even before any understanding of the represented events. Therefore, because of its semantic isolation – both from other objects and from depicted events – the possibility of prioritisation of an inconsistent object appears independent of its specific placement within the scene. Indeed, unlike the situation for diagnostic objects, the locations at which inconsistent objects occur in scenes would likely be indeed rather uninteresting if empty (as can be seen for the images shown in Figure 1).

Any effects of (diagnostic or inconsistent) object informativeness on the allocation of attention may emerge not only in terms of quicker and/or more probable selection of an object, but also in longer inspection once the object has been selected. Several studies have shown that inconsistency with the scene's gist leads to longer fixations, probably reflecting the greater effort required to identify the object itself (e.g., Biederman, Mezzanotte & Rabinowitz, 1982; Friedman, 1979; Gordon, 2004; Mudrik, Lamy & Deouell, 2010; but see Gareze & Findlay, 2007) or to solve the conflict of meaning with respect to context (e.g., De Graef, Christiaens & d'Ydewalle, 1990; Ganis & Kutas, 2003; Henderson, Weeks & Hollingworth, 1999; Hollingworth & Henderson, 1998, 1999; Mudrik et al., 2010). Other studies have instead reported that diagnosticity/high predictability in the scene results in a higher proportion of fixations and longer fixation duration (Einhäuser et al., 2008a; 't Hart et al., 2013). No research, however, has considered diagnosticity versus inconsistency when examining effects on attention disengagement. We might expect either shorter inspection of diagnostic than inconsistent objects, because there is not the requirement to solve the conflict between the object and the scene, or even longer inspection, reflecting greater information uptake concerning the whole semantics of the scene.

Comparisons in previous studies have potentially confounded informativeness and consistency, as relevant objects for the scene, either in terms of high consistency/diagnosticity or high inconsistency have typically been compared to consistent but semantically marginal objects, whose presence does no contribute to determining the gist. To our knowledge, only one study has compared high consistency/diagnosticity and inconsistency systematically within the same task (Spotorno et al., 2013). Using a change detection paradigm with coloured drawings, some advantage was found for diagnosticity, however this was in the context of a complex pattern of findings, which appeared affected by the type of change (addition or deletion), and without directly contrasting diagnosticity and inconsistency within the same scene in the same trial.

We should also consider that the semantic relationship between an object and the scene's gist, in terms of both scene category and depicted event, may influence any effects of perceptual salience on object prioritisation in scenes. Two opposite possibilities appear plausible according to the (scarce) literature. Spotorno and colleagues (2013) suggested that salience acts as a preferential filter, promoting further processing of the most salient objects that are also diagnostic for the gist. The work of Itti and Baldi (e.g., 2009) suggested that high salience would favour, instead, selection of semantically inconsistent objects, as it would foster processing of what is surprising and violates expectations.

The main aim of the present study was to determine whether semantic or perceptual information has a predominant influence on prioritisation in scene viewing and memory, when objects compete within a scene. A secondary aim was to consider whether there is any support for previous claims that inconsistency is prioritised in scene viewing and memory. This work, therefore, provides new insights into the ongoing and unresolved debate about the relative influence of semantic and perceptual properties on prioritisation in scenes. We examined object prioritisation within scenes by manipulating both the perceptual and semantic relationships between the objects and the scene in which they occur. More specifically, we manipulated the perceptual relationship by varying the visual conspicuity of objects within the scene defined in terms of salience (see Experiment 1 Method). We manipulated the semantic relationship by including in each scene an object that was diagnostic for scene's gist (in order to define the depicted event) and another that was highly inconsistent with the gist (considered in terms of both scene category and depicted event). To date, no research has analysed direct competition between diagnostic and inconsistent objects within the same scene and has considered how it could be modulated by the objects' perceptual salience. These manipulations may provide new insights into the relative impacts of high-level and low-level factors in competition for selection and how these factors interact in modulating attentional allocation within the online scene and its working memory representation.

In Experiment 1, we examined the role of the competition between the diagnostic object and the inconsistent object, and how this varies depending on the perceptual salience of each of the two objects, during the course of scene viewing, in a task requiring free exploration of the scene for a subsequent memory test. By analysing eye movements, we were able to isolate the impacts of context-object semantic associations in scenes, salience and their interplay, on both initial selection and further ocular inspection of these two objects. We were also able to compare selection and further inspection of either the diagnostic or inconsistent objects in the case of direct competition (both critical objects included in the same scene context) to the case of no competition (only one critical object present). Finally, the analysis of the same scenes but without the inclusion of one or both of the two critical objects allowed us to consider the impact of object presence *per se* and disentangle any influence due to other properties of the areas of the scene (see Region of Interest definition in Method) in which the objects were placed in the other experimental conditions, in terms of either residual semantic importance of the empty location for the scene, or physical aspects such as the eccentricity or size of these regions.

To isolate any effects of inconsistency *per se* suggested by Experiment 1, Experiment 2 compared competition between the diagnostic object and the inconsistent object to that between the diagnostic object and a consistent but not diagnostic (thus of marginal importance to scene understanding) object. To control for the placement and perceptual properties of the non-diagnostic objects, we created a version of each scene in which the inconsistent object was replaced by a consistent, but only marginally informative, object of the same salience, placed in the same location. The analysis of selection and inspection of consistent, marginally informative objects provides an opportunity to assess prioritisation due to semantic informativeness within the scene.

In Experiment 3, we considered the impact of semantic and perceptual factors when objects compete for both information selection from the scene and inclusion in scene representations in visual working memory. For this, we used a flicker change detection task (e.g., Rensink et al., 1997), in which the diagnostic and the inconsistent objects simultaneously appeared and disappeared. Successful change detection, when the associated transient signals are disrupted, requires not only attentional allocation but also effective comparison of pre/post-change memory representations (see Rensink, 2002, for review). Failures to detect changes may arise from either filtering at encoding, leading to sparse representations where only objects on which attention had been focused are included (e.g., Beck & Levin, 2003; Becker & Pashler, 2002; Rensink et al., 2000), or from later selection within relatively rich representations, leading to difficulties in retrieval or in comparison processes (e.g., Busch, 2013; Fernandez-Duque & Thornton, 2000; Varakin & Levin, 2006). The change detection task, therefore, enables us to consider how the semantic consistency of the two competing critical objects and their respective perceptual salience influence the different processes that are necessary in order to utilise memory traces effectively in behaviour guidance, which is the aim of adaptive information prioritisation.

Experiment 1

Method

Participants. One-hundred and eighteen students of the University of Dundee, UK (aged 18-40) participated for course credits. They all had normal or corrected-to-normal vision. The experiment was approved by the local Ethics Committee of the School of Psychology and carried out in accordance with the provisions of the World Medical Association Declaration of Helsinki. All participants provided written informed consent prior to participating.

Apparatus. The experiment was generated in Experiment Builder (SR Research, Canada). It was conducted on a Dell Optiplex 755 computer running OS Windows XP. Stimuli were shown on a ViewSonic G90f-4 19-inch CRT monitor, with a resolution of 1024 x 768 pixels, and a refresh rate of 100 Hz. A chin rest stabilised the eyes about 63 cm away from the display. Eye movements were recorded using an EyeLink 1000 at a sampling rate of 1000 Hz

(SR Research, Canada). Viewing was binocular, but only the dominant eye was tracked. Eye dominance was determined using a variant of the Miles test (Miles, 1928).

Materials. Sixteen coloured cartoon scenes (1024 x 768 pixels, 31.9 x 23.9 deg), plus one for the practice trial, were used and presented filling the screen on which they appeared. They were modified versions of images used in Spotorno and Faure (2011), originally derived from a children's language test (PFLI: Bortolini, 1995). Since we were interested in the competition between diagnostic and inconsistent objects, we did not only consider the situation in which just one of these critical objects were present (as was the case in Spotorno et al., 2013) but also cases in which both or neither was present. Thus, we created four versions of each scene using Adobe Photoshop CS (Adobe, San Jose, CA): (1) both critical objects present, (2) only the diagnostic object present, (3) only the inconsistent object present, (4) both critical objects absent.

The perceptual salience (high or low) of each critical object was manipulated and counterbalanced across scenes in order to have four groups of four scenes with critical objects belonging to one of the experimental conditions (i.e., both of low salience, both of high salience, high salience-inconsistent object paired with low salience-diagnostic target, low salience-inconsistent object paired with high salience-diagnostic target; see Figure 1).

Pre-testing of materials. Ten independent judges (five males, age: $M = 31.5$, $SD = 4.35$, none took part in the main experiment) evaluated the semantic consistency and the perceptual salience of the critical objects, and the visual complexity of each experimental scene, following a procedure similar to that in Spotorno et al. (2013). The scenes had the same size as in Experiment 1 and were presented only once, in random order, against a medium-grey background. Five judges evaluated first consistency and then salience, and the others did the opposite. All evaluated scene complexity last, before being presented with a new scene. As scores were not normally distributed, we report here median and interquartile range values, with results of Wilcoxon signed-ranks tests (two tailed).

With respect to semantic informativeness, judges first indicated up to three objects considered as the most important for understanding the scene (like if they had to describe briefly the situation depicted) or the most inconsistent with the scene's meaning; then they rated on a Likert scale the probability of occurrence related to the depicted situation in the scene (from 1, minimum, to 6, maximum) of the two objects pre-selected as targets by the experimenters. All the diagnostic objects in the main experiment were included by at least eight judges within the objects indicated as the most strongly contributing to scene meaning, and all the inconsistent objects in the main experiment were indicated by all judges as the only inconsistent object in the scene. Moreover, diagnostic and inconsistent objects were scored on the Likert scale as having on average high (Mdn = 5, IQR = 1) or low probability of occurrence (Mdn = 1, IQR = 1), respectively; this difference was highly significant, $Z = 11.15$, $p < .001$.

Perceptual salience was defined as "something highly visible, which captures attention and stands out for its physical properties, regardless of its semantics". Judges indicated up to three objects they considered as the most salient. Subsequently, they rated, on the same six-point Likert scale, the salience of the two preselected critical objects. The appropriateness of explicit judgment of salience given by human observers and their correlations with the output of classic bottom-up saliency algorithms have been demonstrated by several studies (e.g., Borji, Sihite & Itti, 2013b; Spotorno & Faure, 2011). However, in order to obtain an objective measure, we also created salience map for each scene using an implementation (Ezvision: Itti, 2004) of Itti and

Koch's algorithm (e.g., 2000), so that luminance, colour and orientations were taken into account as low-level properties in each scene. Highly salient critical objects in the main experiment were selected by the model among the first three salient points, indicated by at least eight judges to be among the three most (subjectively) salient objects and each scored on average more than 4 (*Mdn* = 5, *IQR* = 1). Critical objects of low salience in the main experiment were not selected by the model among the first seven points, never indicated by the judges and each scored on average less than 3 (*Mdn* = 2, *IQR* = 0). The difference in salience scores between high salience and low salience objects was highly significant, $Z = 11.12$, $p < .001$.

Importantly, diagnostic and inconsistent objects did not differ significantly in their score of perceptual salience, in both the cases of high salience, $Z = 1.19$, $p = .233$, and low salience, $Z < 1$, $p = .891$, and high and low salient objects did not differ significantly in their score of consistency, in both the cases of diagnosticity, $Z < 1$, $p = .656$ or high inconsistency, $Z = - 1.30$, $p = .192$.

Judges were asked to rate on the six-point Likert scale the visual complexity of the scenes by considering the number and spatial organisation of objects. The range of ratings considering the average evaluation for each scene was 2.30-3.80 (*Mdn* = 3, *IQR* = 1). In addition, the four consistency x salience conditions were comparable as for visual complexity of the scene, all *Zs* < 1, all *ps* > .322).

Mirror-reversed versions of each scene, with and without the two critical objects, were created in order to counterbalance between participants the side of the scene in which a given critical object was presented. It has indeed been shown that viewers have a leftward bias in starting scene inspection (Dickinson & Intraub, 2009; Foulsham & Kingstone, 2010; Nuthmann & Matthias, 2014; Ossandón, Onat & König, 2014). In the "original" (evaluated) orientation, two scenes had both objects in the left half, two had both objects in the right half, while the remaining had one object in each side.

Procedure. The experiment was conducted individually in a dimly illuminated room. Participants were seated in front of the computer screen. They were told that they would see a sequence of scenes, and would have to look at them and try to remember as much information as possible for a subsequent memory test (that never took place). Prior to the experimental task, each participant underwent a randomized nine-point calibration and validation procedure. Recalibrations were performed during the task if necessary. Before each trial a single-point calibration check was applied as the participant fixated a dot in the centre of a medium grey (127, 127, 127) background. This was followed by the scene, which was presented for 5 s, and then by a 1-s blank (mid-grey) screen. Thereafter, the experiment automatically proceeded into the single-point calibration check for the next trial.

The experimental trials were presented in random order and each scene appeared only once for each participant. Eighteen versions of the experiment were created in order to counterbalance across participants the salience and the side of occurrence of the diagnostic and the inconsistent objects (either when they competed within the same image or were presented alone), and also scene orientation when both critical objects were absent. Eight versions presented scenes with either no critical object or with only one critical object, eight versions presented scenes with either only one or both critical objects and two versions presented scenes where both critical objects were always present.

ROI Definition and Data Analysis. The regions of interest (ROIs) for scoring eye movements were defined in Matlab 2012b (MathWorks,Inc., Natick, MA, USA) as a rectangle that encompassed the critical object and included a margin of about one degree around it (see Orquin, Ashby & Clarke, 2016). A fixation was considered as being on a specific ROI if the centre of gaze indicated by the eye tracker fell within the boundary of the ROI.

Raw data were parsed into saccades and fixations using the SR Research algorithm, with a minimum fixation duration of 50 ms. We discarded from analyses trials in which the average calibration error was ≥ 0.5 deg (101 trials), or the maximum error in one of the calibration points was ≥ 1 deg (0 further trials), or the error in the single point calibration check before trial start was ≥ 1 deg (a further 34 trials). In total, 135 trials, corresponding to the 7.3% of total data, were removed prior to analysis, leaving 1708 trials from 113 participants for analysis.

Paired sample *t*-tests showed that diagnostic and inconsistent ROIs did not differ in the percentage of screen area that they covered (diagnostic: $M = 7.9\%$, $SD = 4.5$, inconsistent: $M = 6.3\%$, $SD = 2.7$), $t\ (15) = 1.34$, $p = .201$, or the eccentricity of their centres from the centre of the screen (diagnostic: $M = 9.1$ deg, $SD = 4.7$, inconsistent: $M = 9.0$ deg, $SD = 4.4$), $t\ (15) < 1$, $p = .940$. Independent samples *t*-tests were performed to test for any differences in size and eccentricity of ROIs containing low and high salience critical objects in order to check that our salience dimension was not confounded by these factors. Low and high salience critical objects did not differ in terms of size, $t(30) < 1$, $p = .591$, or eccentricity, $t(14) < 1$, $p = .777$.

Analyses were run using the lmer() function of the lme4 package (Bates, Mächler, Bolker & Walker, 2015) in the R programming environment (The R Foundation for Statistical Computing, Version 3.0.3, 2014). We ran linear mixed models (LMMs) with fixed effects describing the variables of interest (these varied depending on the comparison under test and are described in the relevant sections of the Results that follow) as predictors, and participants and scenes specified as random factors. Where possible, random slope models were used with maximal random effects structure (Barr, Levy, Scheepers & Tily, 2013). In cases where the maximal model did not converge, we simplified the model in stepwise fashion. First, we removed correlations between random-slopes and intercepts. After this, we started by removing the slope of the highest order interaction between the fixed effects, and gradually reduced the complexity of the model until it converged. Where possible we tried to maintain slopes for the interactions that were theoretically important to our analyses – typically those involving the interaction between the perceptual properties of the two objects or between perceptual and semantic properties. We simplified the items (scenes) term before simplifying the subject term. In all cases, we report the most complex model that converged. A full list of model structures used in the present study can be found in Appendix 1.

LMMs have many advantages over traditional ANOVA models. Crucially, they optimise power of the experimental design by performing item analysis and 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 exclusions (see Kliegl, Masson & Richter, 2010).

For each model, we report the predictors' coefficients (β-values), the *SE*-values, the *t*-values, and the associated *p*-values for all significant effects ($p < .05$). Where we find effects that approach but do not reach significance ($.05 < p < .1$) we report *t*- and *p*-values, but not report β or *SE*. We also place no interpretive weight on such effects. We do not report effects that fail to reach $p = .1$. *P*-values are not directly supplied by lme4 package, but were generated using the lmerTest library (Kuznetsova, Bruun Brockhoff & Haubo Bojesen Christensen, 2016). When an interaction was significant, we ran follow-up models to explore it. Graphics were created using the ggplot2 package (Wickham, 2009).

Results

Extrafoveal effects of competition between diagnostic and inconsistent objects. Figure 2 shows the cumulative probability of fixation for each of the diagnostic and inconsistent objects over the first 16 fixations of viewing, for each of the four possible combinations of perceptual salience across the two critical objects.

FIGURE 2 ABOUT HERE

The trends in these plots were explored in a Linear Mixed Model to predict the ordinal fixation number of the first fixation on the critical object[1] with object semantics (inconsistent, diagnostic), the salience of the inconsistent object (low, high), the salience of the diagnostic object (low, high) as categorical fixed effects. In order to examine and control for any possible effect of differences depending upon the side of the scene in which the object appeared, we also included object side (left, right) as a categorical fixed effect in the model. Fixed effects were coded using sum coding, and all possible interactions between the fixed effects were included in the LMM (Model 1.1). Only trials on which both critical objects were fixated at some point in viewing (93.6% of trials) were included in this analysis.

While there was an overall effect of the side of the scene in which the object appeared, with objects on the left fixated sooner (M = 4.3 fixations, SD = 3.1) than objects on the right (M = 5.6 fixations, SD = 3.3), $\beta = .158$, $SE = .023$, $t = 6.98$, $p < .001$, this factor did not interact with any of the other fixed effects in the model. There was a significant effect of object semantics, β = .274, *SE* = .031, t = 8.76, $p < .001$, with later fixation of the inconsistent object (M = 6.1 fixations, *SD* = 3.3) than of the diagnostic object (M = 3.9 fixations, *SD* = 2.9). There was also an effect of the salience of the diagnostic object, β = .132, *SE* = .053, t = 2.49, p = .029, and significant two-way interactions between object semantics and diagnostic object salience, β = .059, *SE* = .028, t = 2.13, p = .037, and between object semantics and inconsistent object salience, β = .142, *SE* = .028, t = 5.16, $p < .001$. These two-way interactions and effects of semantics and salience were qualified by a three-way interaction between the semantics of the object, the salience of the diagnostic object and the salience of the inconsistent object, β = .047, *SE* = .022, t = 2.10, p = .037, suggesting that competition for selection between the critical objects depended upon the interplay between the semantics and salience of both objects.

To explore this three-way interaction we ran follow up LMMs for the inconsistent and diagnostic objects separately. For the inconsistent object (Model 1.2), the only significant effect was that of the side of the screen on which the inconsistent object appeared, β = .161, *SE* = .034, t = 4.70, $p < .001$, with fewer fixations required to first fixate the inconsistent object when it was on the left, than when it was on the right. For the diagnostic object (Model 1.3), we found the expected effect of side, β = .165, *SE* = .036, t = 4.53, $p < .001$. There was a significant effect of the salience of the diagnostic object, β = .192, *SE* = .078, t = 2.45, $p < .030$, with high salience diagnostic objects being selected sooner than low salience diagnostic objects. The salience of the inconsistent object also influenced how soon the diagnostic object was selected, β = .178, *SE* = .079, t = 2.25, p = .043, with diagnostic objects being fixated sooner when the inconsistent object was of low salience than when it was of high salience.

The unconditional probability of fixating the critical objects for each of the first 12 fixations (Figure 3) confirmed the selection advantage of the diagnostic object over the inconsistent object. We ran paired sample *t*-tests with Bonferroni correction (corrected α = .05/12) to test for differences in the probability of selection between the diagnostic and inconsistent objects. The advantage for the diagnostic object was especially evident during the first four fixations when the inconsistent object was of low salience (Figure 3, left panels). In contrast, when the inconsistent object was highly salient (Figure 3, right panels) neither object was more likely to be fixated than the other, except in the case of the first two fixations when both objects were high in salience: the diagnostic was more likely to be selected than the inconsistent in these first two fixations.

FIGURE 3 ABOUT HERE

Foveal effects of competition between diagnostic and inconsistent objects. In order to consider how the two critical objects were viewed once selected, we ran a Linear Mixed Model to predict the total time spent fixating the object during the trial, with the semantics of the object (diagnostic, inconsistent), the salience of the diagnostic object (low, high) and the salience of the inconsistent object (low, high) as sum coded categorical fixed effects (Model 2.1).

We found an overall effect of the semantics of the critical object, $\beta = .028$, $SE = .009$, $t = 3.04$, $p = .003$, with more time spent fixating the inconsistent critical object ($M = 962$ ms, $SD = 551$) than spent fixating the diagnostic critical object ($M = 862$ ms, $SD = 562$). There were also effects of the salience of the diagnostic object, $\beta = .067$, $SE = .025$, $t = 2.64$, $p = .022$, but these were qualified by an interaction between these two predictors, $\beta = .035$, $SE = .008$, $t = 4.44$, $p <$

.001 (Figure 4). A follow-up model (Model 2.2) was run to break down the significant two-way interaction using simple effects: diagnostic object salience influenced the time spent fixating the diagnostic object, $\beta = .203$, $SE = .052$, $t = 3.94$, $p = .001$, but did not influence how long was spent fixating the inconsistent object, $\beta = .065$, $SE = .052$, $t = 1.26$, $p = .226$. A frequently reported finding in previous studies (see Introduction) is longer total fixation duration on inconsistent than consistent objects, reflecting more difficult attentional disengagement in the case of inconsistency, interpreted as an indicator of difficulties in identify the object or of the attempt to solve the semantic conflict the object engenders in the scene. For this reason, we also ran a follow-up model of the simple effects of semantic consistency (Model 2.3). We found that the inconsistent object was inspected for longer than the critical diagnostic object only when this diagnostic competitor was low in salience, $\beta = .126$, $SE = .024$, $t = 5.18$, $p < .001$, but not when it was high in salience, $\beta = .012$, $SE = .024$, $t < 1$, $p = .603$.

FIGURE 4 ABOUT HERE

Selection in the presence and absence of competition between the critical objects. An alternative approach to consider the influence of the competition between the two critical objects is to compare situations in which they compete with each other to situations in which they do not. However, while we can consider selection of the diagnostic object when it had no inconsistent competitor, we cannot consider selection of the inconsistent object when it had no highly consistent and informative competitor. Indeed, when the diagnostic critical object was not present, the inconsistent was still presented in the context of other objects that could be highly meaningful for the scene's gist. We, therefore, restricted our analyses to considering the diagnostic object when it was or was not presented with an inconsistent competitor (and only for cases in which this object was fixated during viewing).

We ran an LMM to predict the ordinal first fixation number on the diagnostic object (Model 3.1) with side of the scene in which the diagnostic object appeared, the salience of the diagnostic object, and competition type as categorical fixed effects. The fixed effect of competition type had three levels: none, competing with a low salience inconsistent object, competing with a high salience inconsistent object. Since our focus here was to compare fixation behaviour in the presence and absence of competition between the critical objects, we used contrast coding of the fixed effect competition type to compare (1) no competitor vs. a low salience competitor and (2) no competitor vs. a high salience competitor.

The diagnostic object was fixated after fewer fixations when present without a competitor (M = 3.7 fixations, SD = 3.2) than when competing with a low salience inconsistent competitor (M = 4.3 fixations, SD = 4.0), β = .299, SE = .078, t = 3.84, p < .001, or when competing with a high salience inconsistent competitor (M = 6.1 fixations, SD = 4.3), β = .461, SE = .074, t = 6.24, p < .001. However, competition type did not interact with any of the other fixed effects. Consistent with earlier analyses, there was an effect of object side, β = .129, SE = .023, t = 5.48, p < .001, with earlier selection of diagnostic objects on the left. There was a tendency toward earlier selection of the diagnostic object when it was high in salience, t = 1.85, p = .086.

A final way of considering the factors that influence selection of the diagnostic and inconsistent objects is to compare selection of a given ROI when it includes the object to when the object is absent. We can also compare selection of the ROI when there is a competitor critical object elsewhere in the scene to that when there is no other critical object to compete for selection. Thus we ran a GLMM to predict the probability that the ROI would be fixated during viewing with semantics of the ROI (that is whether a diagnostic or inconsistent object would be in the ROI if the object was present), presence of the diagnostic object and presence of the inconsistent object as categorical fixed effects (Model 4.1). The probability of fixating the ROI differed between the ROIs in which diagnostic or inconsistent objects would appear, $\beta = .833$, $SE = .113$, $z = 7.36$, $p < .001$, and was affected by the presence of the diagnostic object, $\beta = .832$, $SE = .118$, $z = 7.06$, $p < .001$, and the presence of the inconsistent object, $\beta = 1.07$, $SE = .096$, $z = 11.13$, $p < .001$. These effects were qualified by two significant two-way interactions: between the semantics of the ROI and the presence of the diagnostic object, $\beta = .763$, $SE = .111$, $z = 6.89$, $p < .001$, and between the semantics of the ROI and the presence of the inconsistent object, $\beta = 1.46$, $SE = .092$, $z = 15.89$, $p < .001$. To better visualise the pattern in the data, and to illustrate the two follow-up models, we plot the data across all conditions of ROI semantics, diagnostic presence and inconsistent presence in Figure 5.

FIGURE 5 ABOUT HERE

We ran follow-up GLMMs to consider the diagnostic and inconsistent ROIs separately. For the inconsistent ROI (Model 4.2), the only effect was of the presence of the inconsistent object, $\beta = 3.66$, $SE = .326$, $z = 11.22$, $p < .001$, with far higher likelihood that the ROI would be fixated when it contained an inconsistent object than when it was empty. For the diagnostic ROI (Model 4.3), there was an effect of the presence of the diagnostic object, $\beta = 1.51$, $SE = .132$, $z = 11.40$, $p < .001$, with a higher probability of fixating the diagnostic ROI when it contained a diagnostic object than when it was empty. There was also an effect of the presence of the inconsistent object, $\beta = .354$, $SE = .127$, $z = 1.99$, $p = .047$, with a lower probability of fixating the diagnostic ROI when an inconsistent object was present in the scene than when no inconsistent object was present.

The above analyses show that the probability of fixating the diagnostic ROI was high even when it was empty. However, the probability of fixating this region was still higher when it was full than when it was empty. In order to consider the impact of presence of the diagnostic object in this ROI in more detail we ran a LMM to predict the ordinal fixation number in which this ROI was first fixated, with the side of the scene, the presence of the diagnostic object and the presence of the inconsistent object as categorical fixed effects (Model 5.1). We found a significant effect of the side of the scene, $\beta = .088$, $SE = .023$, $t = 3.92$, $p < .001$, with earlier selection when the diagnostic object was on the left than when it was on the right (as expected given previous analyses). There was an effect of the presence of the diagnostic object, $\beta = .093$, $SE = .027$, $t = 3.46$, $p < .001$, with the diagnostic ROI fixated earlier in viewing when it was full than when it was empty. This effect was qualified by a two-way interaction between the presence of the diagnostic object and the presence of the inconsistent object, $\beta = .052$, $SE = .024$, $t = 2.19$, $p = .029$. A follow up LMM (Model 5.2) coded to explore simple effects within this interaction showed that the presence of the inconsistent object in the scene influenced how soon the diagnostic ROI was fixated when it was full, $\beta = .178$, $SE = .063$, $t = 2.84$, $p = .005$, but not when the diagnostic ROI was empty.

Discussion

We examined how two objects that are both informative for the semantics of the scene, but in opposite directions, are selected when they compete directly for attentional resources within the same scene. We compared diagnostic objects, highly contributing in constructing and confirming the whole meaning of the scene, to inconsistent objects, violating viewers' expectations. Our results reveal insights into the effect of this competition during free viewing, when observers try to memorise as much as they can about the scene, in preparation for a subsequent – here sham – memory test. The insights concern both the selection of the objects for foveal inspection and the foveal inspection that follows overt selection.

There was a clear advantage in terms of earlier selection of the diagnostic object compared to the inconsistent object, with the difference appearing particularly prominent during the first few fixations within the scene and already emerging in the probability of foveating the object with the very first saccade after scene presentation. We can then argue that semantic mismatch between an object and the scene is not as important for information prioritisation as is diagnosticity for categorising the scene and understanding the event depicted by the scene. The relative prioritisation of the diagnostic object, however, was modulated by the perceptual properties of the two objects, and was overridden when a low salience diagnostic object competed with a high salience inconsistent object. The type of interplay between perceptual salience and semantic consistency appears to be of particular interest in our study. Selection of the inconsistent object was independent of the salience of either critical object. On the contrary, selection of the diagnostic object appeared to be influenced by the salience of both critical objects, in opposite directions: enhanced by its own high salience, which resulted in overall earlier fixation during viewing, and impaired, with later fixation, by high salience of the inconsistent counterpart.

Previous research has shown preferential early inspection of the left than of the right half of the scene (Dickinson & Intraub, 2009; Foulsham et al., 2013; Nuthmann & Matthias, 2014; Ossandón, et al., 2014). In our study, even though we reported a leftward bias for the selection of the critical objects, this bias did not modulate in any way the effects of competition between the two objects.

We were also able to consider the role of competition *per se* in object selection by comparing the situation where both critical objects are present to how the same critical object was selected in the absence of this competition. For this consideration, we focused on the diagnostic object because the inconsistent object was a semantic singleton with respect to the scene's gist, whereas the diagnostic object was not, with other objects in the scene contributing to the gist. We showed that having an inconsistent competitor within the same scene delayed selection of the diagnostic object, and this was largely regardless of the salience of either critical object.

Further suggestions about the nature of the informativeness of diagnostic and inconsistent objects can be drawn considering how the region containing the critical object was selected when it was empty, and whether this selection was influenced by the presence of the competitor. Moreover, these are key aspects in order to disentangle any impact due to other intervening physical and semantic factors related to the specific region beyond object's properties. Possible confounding physical effects on selection, linked to size (see Spotorno, Malcolm & Tatler, 2015) or eccentricity, were minimised as the critical regions of interest did not differ for the diagnostic and inconsistent objects in terms of their size or eccentricity in the scene (see Experiment 1 Method). We found that both regions were more likely to be selected when containing the object than when empty, but the difference was much bigger for the region of the inconsistent object, which indeed was fixated only in about 25% of cases when it was empty and almost always, even though usually after the diagnostic region, when it contained the object. The presence of the diagnostic object only gave an overall modest contribution to the selection of the diagnostic region, fixated in slightly more than 70% of the cases even when empty, although later than when it included the object. These findings support the idea of diagnosticity being linked to specific placement that draws attention for its semantic role within the scene rather than its physical properties. Informativeness of the diagnostic object emerges not only from the top-down influence of global scene context (see Trapp & Bar, 2015) but also from a network of co-occurrent semantically related and spatially coherent objects, which observers utilise to guide eye movements effectively during viewing (Davenport, 2007; Hwang et al., 2011; Mack & Eckstein, 2011; Sadeghi et al., 2014; Wu et al., 2014) to the most relevant locations for the current task. In this regard, Pereira and Castelhano (2014) proposed that scene context guides the eyes to relatively broad, potentially meaningful regions while objects mainly affect which placements are selected therein. We may therefore argue that the co-occurring objects in our scenes contributed to defining the depicted event and guiding the eyes to this location even when it was empty.

The low likelihood of selecting the region of the inconsistent object when empty and the finding that the presence of the diagnostic critical object did not affect selection of this region (either empty or not) reinforce the claim that informativeness depending on gist violation is concentrated in the inconsistent object, and arises mainly from the relationship between this object and the whole scene context. Together with this finding, the fact that, on the contrary, the presence of the inconsistent object did influence diagnostic region selection (especially when it contained the object) sheds further light on the direction of the competition according to semantic consistency in the scene, supporting what we have already described discussing the impact of the respective salience of the two critical objects: this competition emerges as an interfering effect of the inconsistent object over the diagnostic, while processing of the inconsistent object appears to be largely insulated from processing of the diagnostic object.

We found that semantics and salience not only influenced the time of object selection but also the duration of object inspection once selected. The literature has provided strong evidence that violations of a scene's semantics leads to greater engagement of attention, resulting in longer dwell time on inconsistent than consistent objects once selected (e.g., Bonitz & Gordon, 2008; Cornelissen & Võ, 2017; De Graef et al., 1990; Gordon, 2004; Friedman, 1979; Henderson et al., 1999; Loftus & Mackworth, 1978; Underwood et al., 2007; Võ & Henderson, 2009). Our study reported this influence, but also showed that it may depend upon the perceptual properties of the diagnostic object (see also Coco et al., 2013, for some indication of an interplay between perceptual and semantic factors in influencing fixation duration). Longer inspection of the inconsistent object than of the diagnostic object was indeed found only when this latter was low in salience, while when it was high in salience viewers spent a similar amount of time on each object. This was because dwell time lengthened on high salience compared to low salience diagnostic objects.

A key question remains unanswered by Experiment 1 concerning prioritisation of the critical objects: how much are the effects of the competition for attentional resources specifically due the semantic identity of the non-diagnostic object, inconsistent with the rest of the scene and with expectations? The effects of the salience of the inconsistent object on selecting and inspecting the diagnostic object that we found in Experiment 1 could arise from the co-occurrence of strong perceptual (salience) and semantic (inconsistency) signals, but could be perceptual effects independent of semantic factors. In order to test this possible explanation of our findings, Experiment 2 juxtaposed the diagnostic object with a consistent but semantically marginally informative object that had the same salience (high or low) as the inconsistent object and replaced it in the same location in half of the trials. This manipulation allowed us to disentangle informativeness from mere scene consistency. We were, therefore, able to consider whether the manner in which competition is mediated between objects depends upon their semantic importance within the scene and whether there is any evidence that inconsistent objects are prioritised or de-prioritised relative to consistent but only marginally informative objects matched for salience and placement in the scene.

Experiment 2

Method

Participants. Seventy-two participants (aged 18-40) took part in this study for course credits or for no remuneration. Data were collected across three laboratories: in Aberdeen (26 participants), Glasgow (22 participants), and Nice (24 participants). The experiment was carried out in accordance with the provisions of the World Medical Association Declaration of Helsinki.

Apparatus. The experiment was generated in Experiment Builder (SR Research, Canada). Stimuli were shown at 32 x 24 degrees in all laboratory setups[2]. Eye movements were recorded using an EyeLink 1000 at a sampling rate of 1000 Hz (SR Research, Canada). Viewing was binocular, but only the dominant eye was tracked.

Materials. These were as used in Experiment 1, but with an additional version of each scene created, in which the inconsistent object was replaced with an object that was consistent with the scene, but of marginal informativeness rather than being diagnostic (we will refer to this object as the marginal object in this study). For each scene, the marginal object that we added was chosen to be of similar size to the inconsistent object it replaced. Furthermore we used the same salience model used for the scenes in Experiment 1 (Ezvision implementation: Itti, 2004) to ensure that the marginal object matched the inconsistent object that it replaced in terms of low-level salience, and we confirmed that it did (as we maintained the same criterion as in Experiment 1 to consider whether an object was of low or high salience): all high salience marginal objects were included within the first three points selected by the model, whereas all low salience marginal objects were not included within the first seven points selected by the model.

Pre-testing of materials. We pre-tested the stimuli in order to confirm that the marginal object was (1) consistent with the scene, but (2) indeed of marginal informativeness for the scene. Subjective evaluations of consistency, defined in terms of probability of occurrence in the scene, and informativeness of the new added object with respect to the scene's general meaning were supplied in a pilot study by ten new judges (4 males, age: $M = 31.5$, $SD = 7.5$) on a six-point Likert scale (1 = minimum, 6 = maximum), in counterbalanced order between participants. The new objects were rated as highly consistent ($Mdn = 6$, $IQR = 1$) but of low informativeness ($Mdn = 1$, $IQR = 1$).

Procedure. As in Experiment 1, except that both critical objects were present in all trials and the semantic consistency of the non-diagnostic object (inconsistent or consistent) was manipulated as a within-participants manipulation. In order to fully counterbalance assignment of images to conditions across participants, and so that each participant viewed each scene only once, four versions of the experiment were created.

ROI Definition and Data Analysis. The regions of interest (ROIs) for scoring eye movements were defined in the same way as for Experiment 1. ROIs for marginal objects did not differ significantly either in size (M = 6.5% of screen area, SD = 2.9) from those for inconsistent objects (M = 6.3% of screen area, SD = 2.7), t (15) = 0.80, p = .438, or in eccentricity, t (15) = 0.24, p = .817.

We discarded from analyses trials in which the average calibration error was ≥ 0.5 deg (77 trials), or the maximum error in one of the calibration points was ≥ 1 deg (15 further trials), or the error in the single point calibration check before trial start was ≥ 1 deg (a further 8 trials). In total, 100 trials, corresponding to the 9.1% of total data, were removed prior to analysis, leaving 1002 trials from 66 participants for analysis.

LMMs were run as in Experiment 1 to explore selection and inspection of the critical objects in the scene. Given the lack of any interaction between the side of the scene and our variables of theoretical interest in Experiment 1, we excluded this fixed effect in the analyses that follow. For each measure, we first modelled the competition between diagnostic and inconsistent objects to test the robustness and replicability of findings in Experiment 1. We then modelled the competition between diagnostic and marginal objects in order to see whether this competition was modulated by the same factors as that between diagnostic and inconsistent objects. Finally, we compared selection of marginal and inconsistent objects (although never in direct competition with each other) in order to consider whether selection of the non-diagnostic critical object was influenced by its consistency with the scene.

Results

Extrafoveal selection of the critical objects. Figure 6 shows the cumulative probability of fixation for each of the diagnostic, marginal and inconsistent objects over the first 16 fixations of viewing, for each of the four possible combinations of perceptual salience across the two critical objects. The patterns evident in these plots were explored by running LMMs to predict ordinal first fixation number on the critical objects.

FIGURE 6 ABOUT HERE

An LMM comparing diagnostic to inconsistent critical objects was run to test whether the findings of Experiment 1 were replicated within Experiment 2 (Model 6.1). We found very similar patterns of significant effects as in Experiment 1. Crucially, we found the same three-way interaction between object semantics, the salience of the diagnostic object and the salience of the inconsistent object as we did in Experiment 1[3], $\beta = .077$, $SE = .027$, $t = 2.86$, $p = .004$. Follow-up LMMs showed that the ordinal fixation number on which the inconsistent object was first fixated (Model 6.2) was not significantly influenced by the salience of either object, although the effect of inconsistent object salience approached significance, $t = 1.91$, $p = .082$. In contrast, diagnostic objects (Model 6.3) were selected sooner when they were high in salience, $\beta = .220$, $SE = .095$, $t = 2.31$, $p = .039$, or when the inconsistent object was low in salience, $\beta = .235$, $SE = .096$, $t = 2.44$, $p = .030$. Thus, selection of diagnostic and inconsistent objects when in competition with each other in Experiment 2 fully replicated that found in Experiment 1.

To consider whether the competition between the diagnostic and marginal objects is mediated by the same factors as the competition between the diagnostic and inconsistent objects we ran an LMM to compare selection of diagnostic and marginal objects (Model 7.1). Diagnostic objects were selected earlier ($M = 4.2$ fixations, $SD = 3.5$) than marginal objects ($M = 8.1$ fixations, $SD = 4.5$), $\beta = .463$, $SE = .039$, $t = 11.9$, $p < .001$, and were selected earlier when high in salience than when low in salience, $\beta = .161$, $SE = .065$, $t = 2.46$, $p = .029$. There was significant two-way interaction between object semantics and diagnostic object salience, $\beta = .095$, $SE = .030$, $t = 3.14$, $p = .002$. A follow-up model to look at simple effects of salience in this interaction (Model 7.2) showed that the diagnostic object was selected earlier when it was high in salience than when it was low in salience, $\beta = .512$, $SE = .144$, $t = 3.57$, $p = .002$, but the salience of the diagnostic object did not affect selection of the marginal object. Looking at simple effects of semantics (Model 7.3), we confirmed the advantage of the diagnostic over the marginal object both when the diagnostic object was low in salience, $\beta = .735$, $SE = .100$, $t = 7.38$, $p < .001$, and when it was high in salience, $\beta = 1.12$, $SE = .099$, $t = 11.5$, $p < .001$.

There was also an interaction between object semantics and marginal object salience, $\beta = .099$, $SE = .030$, $t = 3.28$, $p = .001$. A model of simple effects of salience in this interaction (Model 7.4) showed no significant effects, but an approaching effect of earlier selection of the diagnostic object when the marginal object was low in salience, $t = 1.84$, $p = .083$. Once again, simple effects of semantics within this interaction (Model 7.5) confirmed the advantage of the diagnostic over the marginal object both when the marginal object was low in salience, $\beta = 1.25$, $SE = .101$, $t = 11.1$, $p < .001$, and when it was high in salience, $\beta = .727$, $SE = .095$, $t = 7.62$, $p < .001$.

The three-way interaction between object semantics and the salience of each of the two critical objects that was present in Experiment 1 and for the comparison between diagnostic and inconsistent objects in Experiment 2, was not found for the comparison between diagnostic and marginal objects, $\beta = .008$, $SE = .030$, $t = 0.27$, $p = .788$.

An alternative way to consider the potential influence of inconsistency in the competition between objects is to compare selection of the marginal and inconsistent objects. These objects were never placed in direct competition with each other, but comparing selection of these objects (which were always a competitor for the diagnostic object) allows us to consider whether selection of these objects is influenced by their semantic relationship with the scene. An LMM to predict the ordinal fixation number on which the object was first selected with fixed effects of object semantics (marginal, inconsistent), diagnostic object salience and non-diagnostic (i.e. marginal or inconsistent) object salience (Model 8.1) showed a significant effect of object semantics, with inconsistent objects selected sooner ($M = 6.6$ fixations, $SD = 4.0$) than marginal objects ($M = 8.1$ fixations, $SD = 4.5$), $\beta = .174$, $SE = .030$, $t = 5.81$, $p < .001$. We also found a significant two-way interaction between the semantics and salience of the non-diagnostic object, $\beta = .060$, $SE = .030$, $t = 2.00$, $p = .046$. However, this interaction should be treated with caution as the follow-up model of simple effects in this interaction found no effect of salience for marginal or inconsistent objects (Model 8.2), while it showed the expected earlier selection of inconsistent than marginal objects for both low salience objects, $\beta = .228$, $SE = .089$, $t = 2.56$, $p = .011$, and high salience objects, $\beta = .469$, $SE = .082$, $t = 5.75$, $p < .001$ (Model 8.3).

Figure 7 shows that the result we found in Experiment 1 for the unconditional probability of fixating the diagnostic and inconsistent objects for each of the first 12 fixations was well replicated in Experiment 2. Figure 8 shows that these patterns differed somewhat for trials in which the diagnostic and marginal objects were both present. The selection advantage of the diagnostic object over the marginal object is evident. We ran paired sample *t*-tests with Bonferroni correction (corrected $\alpha = .05/12$) to test for differences in the probability of selection between the diagnostic and marginal objects. The advantage for the diagnostic object was evident during at least the four or first five fixations for all situations except when the diagnostic object was low in salience and the marginal object was high in salience.

FIGURES 7 and 8 ABOUT HERE

Foveal inspection of the critical objects. As for our analysis of extrafoveal selection, we first tested the replicability of our findings from Experiment 1 by modelling the data when the diagnostic and inconsistent objects were present in the scene (Model 9.1). Replicating our findings from Experiment 1, we found longer fixation times on inconsistent objects ($M = 930$ ms, $SD = 513$) than diagnostic objects ($M = 902$ ms, $SD = 596$), $\beta = .018$, $SE = .009$, $t = 2.08$, $p = .038$, and longer total fixation times on critical objects when the diagnostic object was high in salience, $\beta = .080$, $SE = .026$, $t = 3.07$, $p = .010$. These effects were qualified by a two-way interaction between the semantics of the critical object and the salience of the diagnostic object, $\beta = .023$, $SE = .009$, $t = 2.73$, $p = .006$; the nature of this interaction is almost identical to that observed in Experiment 1 (Figure 9, Left). A follow-up model of the simple effects of semantics within this interaction (Model 9.2) showed that the inconsistent object was inspected longer than the critical diagnostic object only when this diagnostic competitor was low in salience, $\beta = .085$, $SE = .027$, $t = 3.20$, $p = .002$, but not when it was high in salience; this result replicates that in Experiment 1. We also looked at simple effects of diagnostic object salience within this interaction (Model 9.3). Inspection time of the diagnostic object was longer when it was high in salience, $\beta = .210$, $SE = .053$, $t = 3.96$, $p = .001$, confirming the result of Experiment 1. However, unlike Experiment 1, inspection times were marginally longer on the inconsistent object when the diagnostic object was high in salience, $\beta = .111$, $SE = .053$, $t = 2.10$, $p = .052$.

FIGURE 9 ABOUT HERE

Unlike in Experiment 1, we also found a two-way interaction between the semantics of the critical object and the salience of the inconsistent object, $\beta = .024$, $SE = .009$, $t = 2.81$, $p = .005$ (Figure 9, Right). A follow-up model of the simple effects of semantics within this interaction (Model 9.4) found that viewing time did not differ between diagnostic and inconsistent objects when the salience of the inconsistent object was low; however, inconsistent objects were viewed for longer than diagnostic objects when the salience of the inconsistent object was high, $\beta = .081$, $SE = .026$, $t = 3.17$, $p = .002$. A model of the simple effects of inconsistent object salience in this interaction (Model 9.5) found no effects of salience on either the diagnostic or inconsistent object.

When comparing foveal inspection of the diagnostic and marginal objects (Model 10.1) we found an overall longer inspection of diagnostic objects ($M = 942$ ms, $SD = 611$) than marginal objects ($M = 635$ ms, $SD = 401$), $\beta = .088$, $SE = .009$, $t = 10.0$, $p < .001$, two-way interactions between object semantics and diagnostic objects salience and between object semantics and marginal object salience, and a three-way interaction between object semantics, diagnostic objects salience and marginal object salience, $\beta = .024$, $SE = .009$, $t = 2.77$, $p = .006$ (Figure 10). Follow-up models showed that inspection time for the marginal object (Model 10.2) was not significantly influenced by the salience of either object, but there was a non significant trend toward a two-way interaction between diagnostic and marginal object salience, $t = 1.86$, $p = .088$. For the diagnostic object (Model 10.3) inspection time was only influenced by the salience of the diagnostic object, $\beta = .110$, $SE = .040$, $t = 2.77$, $p = .016$, with longer inspection of the diagnostic object when it was of high salience ($M = 1187$ ms, $SD = 668$) than when it was of low salience ($M = 689$ ms, $SD = 416$). We also considered simple effects of semantics within the three-way interaction (Model 10.4) and confirmed longer inspection time on diagnostic than marginal objects when they were of similar salience or when the diagnostic was higher in salience, all ts > 3.98, all ps < .001; when the marginal object was more salient than the diagnostic object, we found longer inspection of the marginal object, $\beta = .072$, $SE = .033$, $t = 2.20$, $p = .028$.

FIGURE 10 ABOUT HERE

Comparing inspection time on the marginal and inconsistent objects (Model 11.1), we found an overall effect of non-diagnostic object semantics, with longer inspection of inconsistent than marginal objects, $\beta = .098$, $SE = .009$, $t = 11.3$, $p < .001$. There were significant two-way interactions between non-diagnostic object semantics and the salience of each critical object in the scene, which were qualified by a three-way interaction between non-diagnostic object semantics, diagnostic objects salience and marginal object salience, $\beta = .038$, $SE = .009$, $t = 4.35$, $p < .001$. A model of simple effects of salience within the three-way interaction for inspection time (Model 11.2) confirmed that inspection time on the inconsistent object was longer than that on the marginal object in all cases, ts > 6.40, ps < .001, except when the non-diagnostic object was high in salience and the diagnostic object was low in salience, where we found no significant difference in inspection times.

Discussion

In Experiment 2, the inclusion of objects that were consistent but semantically marginally informative, placed in the same location and having the same perceptual salience as the inconsistent object, allowed two key comparisons in order to better understand the nature of the competition between objects. Comparing diagnostic and inconsistent objects (as in Experiment 1) allowed us to characterise competition when both objects were highly informative but differed in terms of their consistency with the scene. Comparing diagnostic and marginal objects allowed us to characterise competition when both objects were consistent with the scene but differed in terms of their informativeness about the scene's depicted event. In this way, we could isolate effects arising from semantic informativeness and from semantic consistency within these competitive situations.

The results of Experiment 2 about the competition between the diagnostic object and the inconsistent object for attentional resources during selection replicated exactly what we found in Experiment 1, showing once more that the diagnostic object is selected earlier that the inconsistent counterpart, and that this effect emerges during the initial fixations within the scenes. Moreover, this advantage of the diagnostic object disappeared only when that object was low in salience and its (inconsistent) competitor was high in salience. In that case, as reported in Experiment 1, the time course of fixation selection was similar for both critical objects.

We found – not surprisingly – that the diagnostic object was prioritised over the marginal object, which appears (from Figures 6-8) to be even greater than the advantage of the diagnostic object over the inconsistent one. This finding suggests that consistency and informativeness both contribute to prioritising objects in scene viewing.

Considering how perceptual salience may modulate object selection for competition between consistent (diagnostic or marginal) objects, we found that it only influenced how early the diagnostic object was selected, with earlier selection of highly salient diagnostic objects; the salience of the marginal object did not influence how soon either object was selected. We did also find a tendency for high salience marginal objects to delay selection of the diagnostic object, but not to the point that the diagnostic object was selected later than the marginal one. When comparing competing diagnostic and inconsistent objects, the pattern of findings reproduced that reported in Experiment 1, with the diagnostic object selected earlier when it was highly salient and later when the inconsistent competitor was highly salient. Inconsistent object selection showed only a weak tendency to be influenced by that object's salience (while it appeared totally unaffected in Experiment 1). Overall, we can conclude that the pattern of influence of salience on selection varies according to both the object's consistency and semantic informativeness. More specifically, perceptual properties of consistent objects do not appear to modulate the time course of selection of other objects in the scene substantially, and this seems to be largely independent on their informativeness. Perceptual properties of inconsistent (and, therefore, informative) objects appear instead to act by modulating ocular selection of what conveys the core meaning of the scene.

In our experiment, marginal and inconsistent objects never occurred in direct competition, thus we cannot comment on how their respective perceptual properties might interact. We can only note that, showing no salience effects when comparing selection of these two kinds of objects, our finding differs from that reported by Coco and colleagues (2013), who found quicker selection of high salience than low salience inconsistent objects, with no salience impact on selection of consistent objects. These authors, however, reported this result only for eye movements in an object-naming task and not during free viewing.

A key finding of Experiment 2 is that, ruling out any effect of placement within the scene, we found that inconsistent objects were selected earlier than (consistent) marginal objects, showing thus prioritisation due to the object's semantics (we also found a significant interaction involving semantics and salience, but follow-up models failed to find evidence of salience effects on selection of these objects). Taken together with the overall advantage of diagnosticity for selection in both Experiment 1 and 2, this finding suggests that previous reports of earlier selection of inconsistency (e.g., Bonitz & Gordon, 2008; Brockmole & Henderson, 2008; Cornelissen & Võ, 2017; Loftus & Mackworth, 1978; Underwood et al., 2007, 2008) may have arisen from comparisons with consistent but low informative objects for the scene. They might have arisen, thus, by the confusion between the dimension of informativeness, which should promote selection, with that of simple consistency, which is less important for ongoing scene perception and memory once the scene's meaning has been recognised.

To sum up, from our study we can claim that inconsistency does have an informative value affecting selection during scene inspection. More specifically, we may hypothesise that this is due to signalling potential errors in scene interpretation. We can also conclude that once informativeness for the scene is controlled for, inconsistency does not lead to any preferential attentional allocation for selection, but the contrary. On this point, ours is the first study to supply clear evidence that, with comparable informativeness, highly consistent objects are prioritised over inconsistent ones during a free-viewing task. Indeed, previous research that did not support preferential selection of inconsistency in scenes led to a null result when examining free viewing, finding that inconsistent and (low-informative) consistent objects were selected on similar timescales (Coco et al., 2013; Gareze & Findlay, 2007; Henderson et al., 1999; Võ & Henderson, 2009, 2011). So far, except for Coco et al. (2013), who found a consistency prioritisation in the object naming task, any report of an advantage of consistent objects over inconsistent object selection has originated from visual search, which emphasises the importance of matching with expectations in order to perform efficiently (e.g., Castelhano & Heaven, 2011; Eckstein, Drescher & Shimozaki, 2006; Henderson et al., 1999; Spotorno, Malcolm & Tatler, 2014, 2015; Võ & Henderson, 2011).

Considering the duration of foveal information gathering once the object has been selected, virtually every study on the topic has reported so far longer inspection of inconsistent than of consistent objects (see Introduction and Discussion Experiment 1 for references). Experiment 2 showed this pattern regardless of the informativeness of the consistent object but also confirmed that it was limited to specific conditions of salience. Indeed, when comparing inconsistent to diagnostic objects, longer inspection on the inconsistent objet was reported when either the diagnostic object was low in salience, replicating and extending therefore what was shown in Experiment 1, or the inconsistent object was high in salience (whereas no modulation due to inconsistent object salience had been reported in Experiment 1). When comparing inconsistent to consistent but marginally informative objects, we found longer inspection on the inconsistent object except when the non-diagnostic object (inconsistent or marginal) was high in salience and co-occurred with a low salience diagnostic object. In addition, diagnostic objects were fixated for longer than consistent, marginally informative competitors, suggesting that greater attentional engagement may arise from the attempt of maximising information about the scene's meaning. This effect was true in all salience conditions, except when a low salience diagnostic object competed with a high salience consistent but marginally informative object; in this case, inspection time was longer on the marginal object. There was also a suggestion, but only in Experiment 2, that the presence of a perceptually salient diagnostic object might promote longer inspection of co-occurent inconsistent (therefore highly informative) objects in the scene.

Previous work has found that physical properties influence inspection times: Wang, Hwang and Pomplun (2010), who examined visual conspicuity considering size, showed that inspection time on high predictable objects in scenes was longer when they were large. However, effects of perceptual properties on inspection times in our results did not arise from the sizes of objects, with high and low salience critical objects being of similar size in our scenes (see Experiment 1 Method).

Taken together, the findings of Experiments 1 and 2 clearly revealed that objects contributing the most to a scene's (event) gist definition (i.e., diagnostic) are prioritised during free viewing, and also that informativeness arising from inconsistency within the scene leads to some prioritisation if compared to selection of objects that are consistent with but marginally informative for scene understanding. They also showed clear differences in the impact of perceptual salience on attention allocation according to the semantic consistency and informativeness of the objects.

However, while these patterns emerged with respect to scene exploration, we do not know whether they would persist when the critical informative objects compete for prioritisation within scene representation in visual working memory. In other words, we do not know whether and how the diagnostic object and the inconsistent object would compete in terms of visual awareness and utilisation of representation to guide behaviour. In order to explore this question, Experiment 3 employed a flicker change detection task (see Rensink et al., 1997), where the diagnostic object and the inconsistent object appeared or disappeared simultaneously: successful change detection requires both object selection during viewing and during retrieval and comparisons of the working memory traces of the original and the modified scene (see Rensink, 2002, and Introduction). In Experiment 3, we did not measure eye movements because we were not interested in re-characterising oculomotor behaviour during viewing, but rather at characterising any effects of semantic and perceptual factors on the resultant memory representations that support detection of change. We utilised two different versions of the task, requiring participants to detect either both changing objects or only one of the two changes. In this way, we were able to manipulate the cognitive load imposed by the task, and to analyse its influence on object competition and on interplay between semantic and perceptual factors.

Experiment 3

Method

Participants. Sixty students of the University of Aberdeen, participated in the experiment for no remuneration. They all had normal or corrected-to-normal vision. Thirty were randomly assigned to the version of the experiment requiring detection of both changes (seven males, aged 19-44, $M = 22.37$, $SD = 5.39$) and the other thirty were randomly assigned to the version requiring detection of only one of the two co-occurring changes (ten males, aged 19-36, $M = 22.21$, $SD = 4.99$). One participant in the detection of only one change condition, however, was excluded from all analyses due to too many errors, therefore the participants effectively considered for this condition were 29. A two-tailed, independent-sample t test, $t\ (57) = 1.04$, $p = .301$ showed that the two groups did not differ significantly in terms of their Laterality Quotient measured at the Edinburgh Laterality Inventory (Oldfield, 1971). The experiment was approved by the local Ethic Committee of the School of Psychology and carried out in accordance with the provisions of the World Medical Association Declaration of Helsinki. All participants provided written informed consent prior to participating.

Materials and Apparatus. We used the same scenes used in the Experiment 1, but only for the versions (original and mirror orientations) with both or no critical objects present. We also used four scenes for practice selected from the same image database (see Spotorno et al., 2013). All the scenes were sized 19 ×14 cm, 610 × 440 pixels, and were presented in the centre of the screen on a medium grey background (127, 127, 127) on a DELL CRT screen (37.7 x 30.3 cm, 1024 x 768 pixels, resolution 85 Hz). Responses were provided by clicking on a computer mouse. Viewing distance was not controlled in this experiment but was relatively consistent across participants as all sat in front of the monitor at a comfortable distance for using the computer mouse.

Procedure. The experiment was conducted individually in a dimly illuminated room. Participants were seated in front of the computer screen, at a viewing distance of approximately 60 cm, and had to click with the computer mouse (using the right hand) on a changing object as quickly and accurately as possible. Each trial started with a central fixation square (30 x 30 pixels) displayed on a medium-grey (127, 127, 127) background for 1000 ms and consisted of alternations of the scene versions without or with the two critical objects, each presented for 100 ms and separated by a 900-ms medium-grey blank screen[4] (Figure 11). After detection of one or both changes, according to the experimental condition, or after 60 s elapsed, the full version of the scene (containing both critical objects) was presented again in the centre of the screen until participants reconfirmed their choice by clicking again on the object(s); in the two-change condition, participants were required to indicate the objects in the same order they did during online change detection. This phase was included in the trial to ensure that the order that participants clicked on the changes during the trial reflected the order that they noticed these changes. Thereafter, the trial ended automatically and the experiment proceeded to the next.

FIGURE 11 ABOUT HERE

The experimental trials were presented in random order and the scene orientation was counterbalanced across participants. Each scene was presented only in one trial during the experiment.

Order of detection and accuracy were recorded and analysed for each trial. A mouse click was considered correct if the indicated location was within the ROI of the critical object. The response time for each mouse click was also recorded and were analysed for the first response in the trial.

Data Analysis. We discarded from analyses error trials (indicating the wrong object[5] or detecting only one change when detection of both changes was required: 15 trials) and inversions (where there was a mismatch between the order of selection during the trial and during the confirmation screen for the two-change detection version of the experiment, as we could not be sure of the order in which the changes were noticed: 2 trials). We then excluded trials in which the response time to detect the change (the first change in version of the experiment when participants were asked to detect both changes) was longer than three standard deviations from the mean (24 trials), or were made before the second version of the scene appeared and were therefore too short to be detections of the change (0 trials). For the version when participants detected two changes, we also excluded any trials in which the second click was made immediately after the first as in these cases the order of detection was less clear; for this we employed a minimum gap between the two responses of 0.5 seconds (only 1 trial was excluded because of this). In total, 42 trials, corresponding to the 4.4% of total data, were removed prior to analysis.

In the one-change-detection version of the experiment, there were two changing objects, one diagnostic and one inconsistent, therefore the chance of detecting the inconsistent change was 50%. In the two-change-detection version, there were again two changing objects, and we only considered trials in which both objects were correctly detected; therefore, there was a 50% chance that the first detected change was inconsistent.

As for Experiments 1 and 2, analyses were run using the lmer() function of the lme4 package (Bates et al., 2015) in the R programming environment (The R Foundation for Statistical Computing, Version 3.0.3, 2014). We report the predictors' coefficients (β-values), the *SE*-values, the z-values for GLMMs or the t-values for LMMs, and the associated p-values (generated using the lmerTest library, Kuznetsova et al., 2016, in the case of LMMs). As for Experiments 1 and 2, we report significant effects and tendencies toward significance ($.05 < p < .1$). Significant interactions were further analysed with follow-up models. Participants and scenes were specified as random effects in all the models. Graphics were created using the ggplot2 package (Wickham, 2009).

Results

Prioritising informative objects in scene representations. In order to characterise not only online information selection but also its availability within representations in visual working memory, we considered which of the two changing objects was detected first. Inconsistent object salience (low, high), diagnostic object salience (low, high), and the task of the observer (find one change, find two changes) were entered as predictors in a GLMM that analysed the probability of detecting the inconsistent change first (Model 12.1). Only responses on trials in which both objects were correctly identified (by mouse click) after the trial were analysed here (see Method).

The probability of detecting changes to the inconsistent object first differed between the tasks of the observer, $\beta = 0.272$, $SE = .100$, $z = 2.71$, $p = .007$, with a higher probability of selecting the inconsistent object first when participants were asked to detect only one change ($M = .49$, $SD = .50$) than when participants were asked to detect both changes ($M = .40$, $SD = .49$).

There were also effects of the salience of the inconsistent object, $\beta = 1.20$, $SE = .253$, $z = 4.75$, $p < .001$, and the salience of the diagnostic object, $\beta = .504$, $SE = .249$, $z = 2.02$, $p = .043$. These effects were qualified by a three-way interaction between the salience of the inconsistent object, the salience of the diagnostic object and the task of the observer, $\beta = 0.179$, $SE = .089$, $z = 2.01$, $p = .045$ (Figure 12).

FIGURE 12 ABOUT HERE

To follow up this interaction we first ran one-sample *t*-tests to compare selection of the inconsistent object to chance in each condition (chance value = .50, see Data Analysis). When the salience of both critical objects was low, participants were less likely than chance to detect the inconsistent change first when looking for one change, t (28) = 3.00, p = .006, and when looking for two changes, t (30) = 10.30, p < .001. When the salience of the inconsistent object was high and the salience of the diagnostic object was low, participants were more likely than chance to detect the inconsistent change when looking for one change, t (28) = 11.34, p < .001, or two, t (30) = 5.90, p < .001. When the salience of the inconsistent object was low and the salience of the diagnostic object was high, participants detected the inconsistent less frequently than expected by chance when looking for one change, t (28) = 7.76, p < .001, or two, t (30) = 9.06, p < .001. When both objects were high in salience, participants were equally likely to detect either change as the first.

We ran follow-up GLMMs to break down the three-way interaction identified in the GLMM above. When the task was to detect only one of the changing objects (Model 12.2), there was an effect of the salience of the inconsistent object, β = 1.26, SE = .290, z = 4.35, p < .001, with a higher probability that the participant detected the changing inconsistent object when it was of high salience than when it was of low salience. There was also an effect of the salience of the diagnostic change, β = .709, SE = .272, z = 2.61, p = .009, with a higher probability of detecting the inconsistent change when the changing diagnostic object was low in salience than when it was high in salience.

When the task was to detect both of the changes (Model 12.3), there was an effect of the salience of the inconsistent change, β = 1.15, SE = .262, z = 4.41, p < .001, but no effect of the salience of the diagnostic object. The interaction between the saliences of the inconsistent and diagnostic objects tended toward significance, z = 1.85, p = .065. Because of our theoretical interest in the interplay between the two critical objects, we broke down this marginal interaction using a GLMM to look at simple effects (Model 12.4). When the salience of the inconsistent change was low, the salience of the diagnostic object had no influence on which object was selected first (Figure 12, right panel, light grey bars). However, when the salience of the inconsistent change was high, the probability of detecting it first was modulated by the salience of the diagnostic change, $\beta = 1.74$, $SE = .734$, $z = 2.37$, $p = .018$, with a higher frequency of detecting the inconsistent object first when the diagnostic change was of low salience than when the diagnostic change was of high salience (Figure 12, right panel, dark grey bars).

In order to compare detection of competing inconsistent and diagnostic changes directly in the two versions of the task, we ran Welch two-sample *t*-tests for each consistency x salience condition (Bonferroni corrected $\alpha = .05/4$). When the two changes were both of low salience, we found that detection of diagnostic changes as first was greater when participants had to report both changes than when they had to report only one, $t\ (49.8) = 3.30$, $p = .002$. We found no effects of task version when the consistent and the inconsistent changes were both of high salience or when they differed in salience.

Response time for first detection. Response times were log-transformed in order to meet LMM assumptions. In an LMM, we entered the consistency of the first changed detected (inconsistent, diagnostic), the salience of the inconsistent change (low, high), the salience of the diagnostic change (low, high) and the observer's task (detect one change, detect two changes) as predictors of the RT of the first detection (Model 13.1).

The salience of the diagnostic change influenced how quickly the first change was detected, $\beta = .027$, $SE = .012$, $t = 2.27$, $p = .040$, with first detections being faster when the diagnostic object was of high salience ($M = 3.86$ s, $SD = 1.67$) than when it was low in salience

(M = 4.30 s, SD = 1.49). This factor did not interact with any other predictors in the model, suggesting that having a high salience diagnostic change in the scene reduced the time needed to find the first change, irrespective of whether participants detected the inconsistent or diagnostic change first. The only other significant effect in the model was an interaction between the semantics of the first detected change and the salience of the inconsistent object, $\beta = .025$, $SE = .005$, $t = 5.34$, $p < .001$. A follow-up LMM (Model 13.2) was run to consider simple effects within this interaction. When the inconsistent object was detected first, the time to detect it was modulated by its salience, $\beta = .070$, $SE = .025$, $t = 2.74$, $p = .013$, with faster detection of the inconsistent change when it was of high salience (M = 3.87 s, SD = 1.57) than when it was of low salience (M = 4.71 s, SD = 1.54). However, when the diagnostic change was detected first, the time to detect it was not influenced by the salience of the inconsistent object.

Discussion

Three key findings in Experiment 3 inform current understanding of how low-level salience and semantic informativeness in terms of consistency with gist influence scene processing and representation.

First, perceptual salience had a dominant influence, in line with some previous change detection studies (Anderson & Donk, 2017; Pringle et al., 2001; Spotorno & Faure, 2011; but see e.g., Stirk & Underwood, 2007). When one of the changing objects was higher in perceptual salience, it had a higher probability of being the first detected, independently of its consistency. Furthermore, when both objects were highly salient, no advantage due to consistency emerged, with both changes having the same probability of being the first detected. Finally, the salience of the change also affected response times, with high salience diagnostic objects generally speeding detection of the first change (irrespective of its semantics), and high salience of the inconsistent object speeding first detection times only when the inconsistent was detected first.

Second, the task of the observer influenced the impact of salience and consistency. When detection of only one change was required, which change was likely to be detected depended upon the perceptual salience of both changing objects. When the two changes both had to be detected, the influence of diagnostic object salience on the type of change detected first was limited to the cases in which the inconsistent object was highly salient, while a low salience, inconsistent object was overshadowed by the presence of a diagnostic object in the scene, regardless of the salience of the diagnostic object. This suggests that a higher cognitive load imposed by the task enhances reliance on the scene's gist. Moreover, the persistent and coherent effect of the salience of the inconsistent object in both tasks is in agreement with the notion that salience acts as surprise signature (e.g., Itti & Baldi, 2009), whose influence is therefore potentiated in association with observations conflicting with viewers' beliefs about the semantic coherence of scene context.

Third, we found an advantage for diagnostic objects compared to inconsistent objects, which emerged specifically when both targets were of low salience and, thus, no low-level critical object prioritisation could occur. This advantage was greater when two changes had to be detected than when only one had to be detected, supporting our suggestion that higher cognitive load enhances reliance on the scene's gist. In previous work, we found better detection of diagnostic than inconsistent changes for highly salient additions or, regardless of salience, for deletions (Spotorno et al., 2013). In contrast, semantic effects of prioritisation of diagnostic over inconsistent objects emerged only for low salience objects in the present study. This apparent discrepancy may stem from several sources that may lead to different perceptual processes. First, the inclusion of two changes rather than one change in each scene, putting diagnostic and inconsistent objects directly in competition, although, as we described above, it should be noted that the advantage due to diagnosticity emerged also when the task was to detect only one of the two changing objects. Second, the use of the flicker paradigm in the present study, that allows for accumulation of information about the scene (e.g., Vierk & Kiesel, 2008), rather than a one-shot change detection paradigm, in which detection is based on representation formed from only very brief and single scene presentation at each trial. Third, differences associated with the type of change in Spotorno et al. (2013), where only an addition or only a deletion was made in each trial, whereas in the present study we alternated these two types of changes within the same trial, as a result of the flickering.

General Discussion

Despite a long history of work on the importance of the match between the semantics of an object and the scene in which it occurs (since the seminal work in the 1970s-80s of Biederman et al., 1982; Boyce, Pollatsek & Rayner, 1989; Friedman, 1979; Loftus & Macworth, 1978; Palmer, 1975), this is the first study to have examined the situation of direct competition between semantically diagnostic and inconsistent objects in the same scene. Moreover, we examined how this competition may be modulated by the perceptual properties of the objects.

We used two different paradigms in order to approach these questions from different perspectives. In the first two experiments, we considered perceptual or semantic influences during scene viewing for memorisation, focusing on any evidence of extrafoveal processing and selection prioritisation of objects' semantic or perceptual information, and on differences in foveal processing. In the third experiment, using a change detection task, we examined how semantic and perceptual properties of informative objects, either diagnostic or inconsistent, may also influence the incorporation into, and availability from, scene representations.

Semantic and perceptual prioritisation in scene processing

Regarding object semantics, we found that, when compared to diagnostic objects, inconsistent objects were looked at later in scene viewing (Experiments 1 and 2) and were less likely to be detected first if changing in a flicker paradigm (Experiment 3). A residual prioritisation benefit, in terms of earlier ocular selection, due to inconsistency was found only when inconsistent objects were compared to objects that were consistent but marginally informative with respect to scene meaning (Experiment 2). Our results concerning object semantic guidance have three main implications. First, they indicate that previous controversial evidence about an inconsistency advantage for selection prioritisation during viewing (e.g., Bonitz & Gordon, 2008; Brockmole & Henderson, 2008; Cornelissen & Võ, 2017; Hollingworth & Henderson, 2003; Loftus & Mackworth, 1978; Stirk & Underwood, 2007; Underwood et al., 2007, 2008) may have arisen from a bias in informativeness *per se,* due to comparisons with poorly informative objects. Second, they corroborate previous research, carried out in the change detection domain, which considered consistent/diagnostic and consistent/low informative objects and showed that informativeness in terms of diagnosticity for scene leads to preferential selection (O'Regan et al., 2000; Pringle et al., 2001; Rensink et al., 1997, 2000; Spotorno & Faure, 2011). We reinforce this earlier claim by demonstrating that diagnosticity prioritisation holds true even when removing the imbalance in informativeness between the compared objects. Third, by showing that the importance of the object for scene meaning acts as a source of guidance for fixation selection (Experiments 1 and 2), we support the claim that objects may be recognised extrafoveally, at least to an extent that enables a first understanding of object-context semantic associations.

Regarding perceptual properties, we showed that objects that were high in perceptual salience were prioritised compared to low salience objects during both scene viewing and when searching for changes in a scene. These findings are broadly in line with accounts that posit an important involvement of low-level guidance in scene processing (e.g., Borji et al., 2013a; Coco et al., 2013; Itti & Koch, 2000; Latif et al., 2014; Parkhurst et al., 2002; Pringle et al., 2011; Spotorno & Faure, 2011; Underwood & Foulsham, 2006). That said, clear differences in how perceptual and semantic factors interacted emerged across the two paradigms we used, highlighting the influence of the type of task, and of the task's cognitive load, on how observers utilise these objects' properties in order to guide attentional allocation and information representation.

Task oriented prioritisations during scene viewing and working memory

When the purpose is to explore the image in order to maximise information gathering about the whole scene (free-viewing for subsequent memorisation test), semantic factors clearly dominate selection of objects and the impact of an object's perceptual salience becomes subordinate to that of the semantic associations between the object and the context of the scene. In particular, observers seem to adopt a strategy favouring overall understanding of the event depicted in the scene despite possible local violations and, therefore, prioritising diagnostic (highly consistent and relevant) information. Secondary to this, we found that semantically inconsistent objects appear to be prioritised relative to perceptually matched objects that were consistent with the scene but of low importance for understanding the depicted event.

When the purpose is to focus on specific local information (i.e., detecting changing objects), observers seem to switch to a strategy that utilises low-level properties as the key dimension guiding behaviour, at least when the scenes are presented quickly as in our flicker task. Indeed, in the absence of a specific template indicating the appearance of the object to look for and under considerable time pressure, selecting objects that are the most visible in the scene may be a useful principle in tasks requiring local focusing and representation. Semantics in this case appear to act as a subsidiary source of guidance when preferential attentional allocation to the changing objects cannot be based on their perceptual salience, and especially when the cognitive load imposed by the task is high.

As an alternative explanation for the differences found between our two paradigms, object prioritisation within scenes might be based mainly on object-context semantic associations during scene viewing and mainly on perceptual salience during the subsequent representational processing in visual working memory required by the change detection task. Previous research, indeed, has demonstrated specific effects of low-level features on working memory representations, increasing memory for high salience objects and decreasing memory for low salience objects (e.g., Fine & Minnery, 2009; Pedale & Santangelo, 2015; Santangelo, 2015). However, these working memory effects appear related mainly to encoding (Santangelo & Macaluso, 2013), that is to say to the phase dependent on selection during viewing, and therefore could indirectly arise from the impact of perceptual features on scene viewing behaviour. In addition, it has been shown that object-scene semantic associations also affect memory contents, with a retrieval advantage for consistent compared to inconsistent objects (Silva, Groeger & Bradshaw, 2006). These previous findings lead us to think that the differential impact of salience and semantic consistency we found in our study, and the modulations of the interplay between these two factors, are largely the genuine result of different task-oriented prioritisations in free-viewing for memorisation versus change detection conditions.

The validity of the scenes

The type of scenes used in this study deserves careful consideration. Cartoon depictions were chosen to manipulate perceptual salience of diagnostic, consistent but marginally informative and inconsistent objects appropriately, within stimuli illustrating meaningful events and maintaining some properties of natural scenes, like realistic organisation, foreground/background relationships, object proportions, gravitational support and colour (all crucial aspects in scene perception, e.g., Oliva & Schyns, 2000). Crucially, the utilisation of drawings allowed us to produce changes in object presence without altering visual contextual properties, like 3D cues or shadows. Drawings have played an important role in the debate about the importance of inconsistency in scene perception, being used in the earliest work in this debate (e.g., Biederman et al., 1982; Boyce et al., 1989; Loftus & Mackworth, 1978; Palmer, 1975) and in key papers over the last few decades (e.g., De Graef et al., 1990; Gareze & Findlay, 2007; Gordon, 2004; Henderson et al., 1999; Hollingworth & Henderson, 1998, 1999, 2003).

Our scenes, however, differ from real-world scenes in important properties, such as perspective and shadows, and are likely to be sparser. We may wonder whether the relative reliance upon the objects' visual features and semantics would differ in more crowded scenes: as crowding does not impair detection, but identification or discrimination (see Whitney & Levi, 2011), in crowded scenes the effects of semantics might be less pronounced as they are based on some access to object identity (although not necessarily on fine identification). The salience influence might be relatively unaffected by the level of crowding, especially if based purely on feature detection (e.g., Felisberti, Solomon & Morgan, 2005). This evidence arises, however, from research utilising simple stimuli, as no study has focused on the effect of crowding on salience and semantic influences in either real-world or realistic scenes. It remains unclear, therefore, whether the properties of our scenes might have modulated the relative magnitude of semantic and salience effects. Despite this eventuality, we argue that these aspects have not affected the validity of our conclusions, in particular about the interaction between high- and low-level factors, for several reasons. First, the nature of visual representations does not seem essentially different between of photographic and non-photographic scenes (Tatler & Melcher, 2007). Moreover, the scenes we utilised in the different conditions of consistency and salience had similar complexity considering their layout and the number of objects, as rated by independent observers (see Experiment 1 Method). These subjective estimates may predict performance to the same extent as objective measures of crowding, like feature congestion or edge density (Neider & Zelinsky, 2011). Finally, our modelling analysis including the scenes as random factor removed interferences from scene idiosyncrasies.

Conclusion

We can conclude that not all informativeness for scene understanding is the same, and that the ongoing task may shape the utilisation of semantic and perceptual properties of the stimulus as well as how these properties interact. High-level and low-level sources of guidance may be utilised flexibly in order to achieve different purposes and optimise behaviour consequently.

What, if any, specific influence of inconsistency on selection behaviour can be found? The present findings show that inconsistency may act as an interfering (and covert) influence on the basic attempt to make sense of the visual world primarily by normalising it according to observers' expectations. This interference seems restricted to when the inconsistent object is also perceptually salient in the scene. Moreover, the conflicting evidence with respect to scene meaning and expectations confers to inconsistency a relative attentional prioritisation during viewing only if compared to consistent information that is not important for interpreting the situation depicted in the scene.

How generalisable are these findings? This remains the open challenge for further research on the topic. Future studies will have to show us – in different visual settings, especially real-world environments, and in different visual tasks – whether and how observers do maximise consistent information that is relevant for current understanding of the scene and for the type of task, and whether and how they are able to use low-level properties in order to achieve cognitive and functional goals. But at least from our findings we can suggest that it is not necessarily the elephant in the room that will be prioritised for inspection and representation, unless the room we are in is part of a zoo or the elephant is visually salient against its background.

References

Anderson, N. C., & Donk, M. (2017). Salient object changes influence overt attentional prioritization and object-based targeting in natural scenes. *PloS One*, *12*(2), e0172132.

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, *68*(3), 255-278.

Bates, D., Mächler, M., Bolker, B. & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software, 67*(1), 1-48.

Beck, M. R., & Levin, D. T. (2003). The role of representational volatility in recognizing pre-and postchange objects. *Perception & Psychophysics*, *65*(3), 458-468.

Becker, M. W., & Pashler, H. (2002). Volatile visual representations: Failing to detect changes in recently processed information. *Psychonomic Bulletin & Review*, *9*(4), 744-750.

Biederman, I. (1972). Perceiving real-world scenes. *Science*, *177*(43), 77-80.

Biederman, I., Mezzanotte, R. J., & Rabinowitz, J. C. (1982). Scene perception: Detecting and judging objects undergoing relational violations. *Cognitive Psychology*, 14, 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. & Itti, L. (2013). State-of-the-Art in Visual Attention Modeling. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, *35*(1), 185–207.

Borji, A., Sihite, D. N., & Itti, L. (2013a). Objects do not predict fixations better than early saliency: A re-analysis of Einhäuser et al.'s data. *Journal of Vision, 13(*10):18, 1-4

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

Bortolini, U. (1995). *Prove per la valutazione fonologica del linguaggio infantile,* Padua: Master.

Boyce, S. J., Pollatsek, A., & Rayner, K. (1989). Effect of background information on object identification. *Journal of Experimental Psychology: Human Perception and Performance*, *15*(3), 556-566.

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.

Busch, N. A. (2013). The fate of object memory traces under change detection and change blindness. *Brain Research*, 1520, 107-115.

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.

Chen, X., & Zelinsky, G. J. (2006). Real-world visual search is dominated by top-down guidance. *Vision Research*, *46*(24), 4118-4133.

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

Cornelissen, T. H., & Võ, M. L. H. (2017). Stuck on semantics: Processing of irrelevant object-scene inconsistencies modulates ongoing gaze behavior. *Attention, Perception, & Psychophysics*, *79*(1), 154-168.

Davenport, J. L. (2007). Consistency effects between objects in scenes. *Memory & Cognition*, *35*(3), 393-401.

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

Dickinson, C. A., & Intraub, H. (2009). Spatial asymmetries in viewing and remembering scenes: Consequences of an attentional bias? *Attention, Perception and Psychophysics, 71*(6), 1251-1262.

Dombrowe, I. C., Olivers, C. N., & Donk, M. (2010). The time course of color-and luminance-based salience effects. *Frontiers in Psychology*, *1*, 189.

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.

Einhäuser, W., Spain, M., & Perona, P. (2008a). Objects predict fixations better than early saliency. *Journal of Vision, 8*(14):18, 1-26.

Einhäuser, W., Rutishauser, U., & Koch, C. (2008b). Task-demands can immediately reverse the effects of sensory-driven saliency in complex visual stimuli. *Journal of Vision*, *8*(2):2, 1-19.

Felisberti, F. M., Solomon, J. A., & Morgan, M. J. (2005). The role of target salience in crowding. *Perception, 34*(7), 823-833.

Fernandez-Duque, D., & Thornton, I. M. (2000). Change detection without awareness: Do explicit reports underestimate the representation of change in the visual system? *Visual Cognition*, *7*(1-3), 323-344.

Fine, M. S., & Minnery, B. S. (2009). Visual salience affects performance in a working memory task. *The Journal of Neuroscience*, *29*(25), 8016-8021.

Foulsham, T., & Kingstone, A. (2010). Asymmetries in the direction of saccades during perception of scenes and fractals: Effects of image type and image features. *Vision Research, 50*(8), 779-795.

Foulsham, T., Gray, A., Nasiopoulos, E., & Kingstone, A. (2013). Leftward biases in picture scanning and line bisection: A gaze-contingent window study. *Vision Research*, *78*, 14-25.

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.

Ganis, G., & Kutas, M. (2003). An electrophysiological study of scene effects on object identification. *Cognitive Brain Research*, *16*(2), 123-144.

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.

Greene, M. R., Baldassano, C., Esteva, A., Beck, D. M., & Fei-Fei, L. (2016). Visual scenes are categorized by function. *Journal of Experimental Psychology: General*, *145*(1), 82-94.

Henderson, J. M., Malcolm, G. L., & Schandl, C. (2009). Searching in the dark: Cognitive relevance versus visual salience during search for non-salient objects in real-world scenes. *Psychonomic Bulletin & 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., & Henderson, J. M. (2000). Semantic informativeness mediates the detection of changes in natural scenes. *Visual Cognition*, *7*(1-3), 213-235.

Hollingworth, A., & Henderson, J. M. (2003). Testing a conceptual locus for the inconsistent object change detection advantage in real-world scenes. *Memory & Cognition*, *31*(6), 930-940.

Hwang, A. D., Wang, H. C., & Pomplun, M. (2011). Semantic guidance of eye movements in real-world scenes. *Vision Research*, *51*(10), 1192-1205.

Itti, L. (2004). The iLab neuromorphic Vision C++ toolkit: Free tools for the next generation of vision algorithms. *The Neuromorphic Engineer, 1*(1), 10.

Itti, L., & Baldi, P. (2009). Bayesian surprise attracts human attention. *Vision Research, 49*(10), 1295-1306

Itti, L., & Koch, C. (2000). A saliency-based search mechanism for overt and covert shifts of visual attention. *Vision Research, 40*(10-12), 1489-1406.

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

Koch, C. & Ullman, S. (1985). Shifts in selective visual attention: towards the underlying neural circuitry. *Human Neurobiology*, *4*(4), 219–227

Koch, K., McLean, J., Segev, R., Freed, M. A., Berry, M. J., Balasubramanian, V., & Sterling, P. (2006). How much the eye tells the brain. *Current Biology*, *16*(14), 1428-1434.

Kollmorgen, S., Nortmann, N., Schröder, S., & Köning, P. (2010). Influence of low-level stimulus features, task dependent factors, and spatial biases on overt visual attention. *PLoS Computational Biology, 6*(5), 1-20.

Kuznetsova, A., Bruun Brockhoff, P., & Haubo Bojesen Christensen, R. (2016). lmerTest: Tests in Linear Mixed Effects Models. R package version 2.0-32. http://CRAN.R-project.org/package=lmerTest.

Latif, N., Gehmacher, A., Castelhano, M. S., & Munhall, K. G. (2014). The art of gaze guidance. *Journal of experimental psychology: human perception and performance*, *40*(1), 33-39.

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

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.

Miles, W. R. (1928). Ocular dominance: Methods and results. *Psychological Bulletin*, *25*, 155-156.

Mudrik, L., Lamy, D., & Deouell, L. Y. (2010). ERP evidence for context congruity effects during simultaneous object–scene processing. *Neuropsychologia*, *48*(2), 507-517.

Neider, M. B., & Zelinsky, G. J. (2011). Cutting through the clutter: Searching for targets in evolving complex scenes. *Journal of Vision, 11*(14):7, 1-16.

Nuthmann, A. & Henderson, J. M. (2010). Object-based attentional selection in scene viewing. *Journal of Vision*, *10*(8): 20, 1-19.

Nuthmann, A., & Matthias, E. (2014). Time course of pseudoneglect in scene viewing. *Cortex, 52*, 113-119.

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

O'Regan, J. K., Deubel, H., Clark, J. J., & Rensink, R. A. (2000). Picture changes during blinks: Looking without seeing and seeing without looking. *Visual Cognition,7* (1/2/3), 191-211.

Oldfield, R. C. (1971). The assessment and the analysis of handedness: The Edinburgh Inventory. *Neuropsychologia*, *9*, 94-117.

Oliva, A. (2005). Gist of the scene. *Neurobiology of Attention*, *696*(64), 251-258.

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

Orquin, J. L., Ashby, N. J., & Clarke, A. D. (2016). Areas of interest as a signal detection problem in behavioral eye-tracking research. *Journal of Behavioral Decision Making*, *29*(2-3), 103-115.

Ossandón, J. P., Onat, S., & König, P. (2014). Spatial biases in viewing behavior. *Journal of Vision, 14*(2):20, 1-26.

Pajak, M., & Nuthmann, A. (2013). Object-based saccadic selection during scene perception: Evidence from viewing position effects. *Journal of Vision*, *13*(5):2, 1-21.

Palmer, T. E. (1975). The effects of contextual scenes on the identification of objects. *Memory & Cognition*, *3*, 519-526.

Parkhurst, D. J., Law, K., & Niebur, E. (2002). Modeling the role of salience in the allocation of overt visual attention. *Vision Research*, *42*(1), 107-123.

Pedale, T., & Santangelo, V. (2015). Perceptual salience affects the contents of working memory during free-recollection of objects from natural scenes. *Frontiers in Human Neuroscience*, *9*, 60.

Pereira, E. J., & Castelhano, M. S. (2014). Peripheral guidance in scenes: The interaction of scene context and object content. *Journal of Experimental Psychology: Human Perception and Performance*, *40*(5), 2056-2072.

Potter, M. C. (1975). Meaning in visual scenes. *Science*, *187*, 965-966.

Pringle, H. L., Irwin, D. E, Kramer, A. F., & Atchley, P. (2001). The role of attentional breadth in perceptual change detection. *Psychonomic Bulletin and Review*, *8*, 89-95.

Rensink, R. A. (2002). Change detection. *Annual Review of Psychology*, *53*, 245-277.

Rensink, R. A., O'Regan, J. K., & Clark, J. J. (1997). To see or not to see: The need for attention to perceive changes in scenes. *Psychological Science*, *8*(5), 268-373.

Rensink, R. A., O'Regan, J. K., & Clark, J. J. (2000). On the failure to detect changes in scenes across brief interruptions. *Visual Cognition*, 7(1/2/3), 127-145.

Sadeghi, Z., McClelland, J. L., & Hoffman, P. (2015). You shall know an object by the company it keeps: An investigation of semantic representations derived from object co-occurrence in visual scenes. *Neuropsychologia*, *76*, 52-61.

Santangelo, V. (2015). Forced to remember: when memory is biased by salient information. *Behavioural Brain Research*, *283*, 1-10.

Santangelo, V., & Macaluso, E. (2013). Visual salience improves spatial working memory via enhanced parieto-temporal functional connectivity. *The Journal of Neuroscience*, *33*(9), 4110-4117.

Schyns, P. G. (1998). Diagnostic recognition: Task constraints, object information, and their interactions. *Cognition*, *67*(1), 147-179.

Silva, M. M., Groeger, J. A., & Bradshaw, M. F. (2006). Attention–memory interactions in scene perception. *Spatial Vision*, *19*(1), 9-19.

Spotorno, S., & Faure, S. (2011). Change detection in complex scenes: Hemispheric contribution and the role of perceptual and semantic factors. *Perception, 40*, 5-22.

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.

Spotorno, S., Malcolm, G. L., & Tatler, B. W. (2015). Disentangling the effects of spatial inconsistency of targets and distractors when searching in realistic scenes. *Journal of Vision, 15*(2):12, 1-21.

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

t Hart, B. M., Schmidt, H. C. E. F., Roth, C., & Einhauser, W. (2013). Fixations on objects in natural scenes: Dissociating importance from salience. *Frontiers in Psychology*, *4,* 455:1-9.

Tatler, B. W., & Melcher D. (2007). Picture in mind: Initial encoding of object properties varies with the realism of the scene stimulus. *Perception, 36*, 1715-1729.

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.

Tatler, B. W., Brockmole, J. R., & Carpenter, R. H. S. (2017). LATEST: A model of saccadic decisions in space and time. *Psychological Review*, *124*(3), 267-300.

Trapp, S., & Bar, M. (2015). Prediction, context, and competition in visual recognition. *Annals of the New York Academy of Sciences*, *1339*(1),190-198.

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 (Eds). *Eye movements: A window on mind and brain*, 564-579. Amsterdam: Elsevier.

Underwood, G., Humphreys, K., & van Loon, E. (2011). Decisions about objects in real-world scenes are influenced by visual saliency before and during their inspection. *Vision Research, 51*, 2013-2038.

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.

Varakin, D. A., & Levin, D. T. (2006). Change blindness and visual memory: Visual representations get rich and act poor. *British Journal of Psychology*, *97*(1), 51-77.

Vierk, E., & Kiesel, A. (2008). Change detection: Evidence for information accumulation in flicker paradigm. *Acta Psychologica, 127*, 309-323.

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.

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.

Wang, H. C., Hwang, A. D., & Pomplun, M. (2010). Object frequency and predictability effects on eye fixation durations in real-world scene viewing. *Journal of Eye Movement Research*, *3*(3):3, 1-10.

Whitney, D., & Levi, D. M. (2011). Visual crowding: A. fundamental limit on conscious perception and object recognition*. Trends in Cognitive Sciences, 15*(4), 160-168.

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

Wu, C. C., Wick, F. A., & Pomplun, M. (2014). Guidance of visual attention by semantic information in real-world scenes. *Frontiers in Psychology*, *5*, 54.

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.

Notes

1. As an alternative measure of how long it took to first fixate the critical objects, we considered the time from the onset of the scene to the first fixation on the critical object. While fixation number and time are highly correlated, it is possible that, if processing time in fixations differed between our experimental conditions, these two analyses might show differences. However, an LMM to predict time to first fixate the critical object showed the same pattern of significant effects as our reported model of the ordinal fixation number.

2. In Aberdeen, images were displayed at 1440 x 1080 pixels on a monitor with display resolution set to 1920 x 1080, viewed at a distance of 72 cm. In Glasgow, images were displayed at 1142 x 858 pixels on a monitor with display resolution set to 1920 x 1080, viewed at a distance of 57 cm. In Nice, images were displayed at 999 x 756 pixels on a monitor with display resolution set to 1024 x 768, viewed at a distance of 57 cm.

3. Note that we also found the same effect of object semantics, the same two-way interactions between object semantics and diagnostic object salience and between object semantics and inconsistent object semantics as we did in Experiment 1. The only difference in the pattern of significant effects between Experiment 2 and Experiment 1 was that the overall effect of diagnostic salience found in Experiment 1 approached by did not reach significance in Experiment 2 (p = .076).

4. The durations of the scenes and the blank screen separating the scenes were piloted. A previous experiment involving 28 subjects, who did not take part in the present study, utilised the

same procedure but with presentation times for the scenes and the blank screen of 500 ms and 250 ms, respectively. In this case, participants in some trials reported the impression of detecting both changes simultaneously or almost simultaneously. To reduce considerably this potential confound, we conducted further tests on five further subjects, who were not involved in the present study either. We found that shortening scene duration (i.e., the possibility of information gathering at each presentation) to 100 ms and lengthening blank screen duration (i.e., the period of potential decay of the mnestic representation) to 900 ms eliminated subjective reports of simultaneous detection. For a discussion of the impact of scene and interval durations on change detection performance, see Rensink et al. (2000).

5. As for the experimental version requiring detection of both changes, a trial was considered incorrect if each critical object was not selected during the presentation of the A/A' cycle or when the scene reappeared at the end of the trial. As for the experimental version requiring detection of only one change, however, we considered correct all the trials in which participants selected one of the two critical objects at least when the scene reappeared at the end of the trial. This is because in this version of the experiment, in the attempt of maximising their speed, participants sometimes (9.9% of the trials) did not follow the instruction of clicking in the location of the changing object during online detection, and in these cases just clicked on the scene wherever the mouse pointer was positioned while they spotted the change.

Appendix 1: Model structures

Model 1.1 (Diagnostic vs Inconsistent object)
*Ordinal first fixation number on object ~ Side of Screen * Object Semantics * Diagnostic Object Salience * Inconsistent Object Salience* + (1 + *Object Semantics* + *Diagnostic Object Salience* + *Inconsistent Object Salience* + *Object Semantics*:*Diagnostic Object Salience* + *Object Semantics*:*Inconsistent Object Salience* ||*Subject*) + (1|*Scene*)

Model 1.2 (Inconsistent object only) and Model 1.3 (Diagnostic object only)
*Ordinal first fixation number on object ~ Side of Screen * Diagnostic Object Salience * Inconsistent Object Salience* + (1 + *Side of Screen* + *Diagnostic Object Salience*:*Inconsistent Object Salience* ||*Subject*) + (1|*Scene*)

Model 2.1 (Diagnostic vs Inconsistent object)
*Log(Total fixation time) ~ Object Semantics * Diagnostic Object Salience * Inconsistent Object Salience* + (1 + *Object Semantics* + *Diagnostic Object Salience* + *Inconsistent Object Salience* + *Object Semantics: Diagnostic Object Salience* + *Diagnostic Object Salience*:*Inconsistent Object Salience*||*Subject*) + (1|*Scene*)

Model 2.2 (simple effects) and Model 2.3 (simple effects)
*Log(Total fixation time) ~ Object Semantics by Diagnostic Object Salience * Inconsistent Object Salience* + (1 + *Object Semantics* + *Diagnostic Object Salience* + *Inconsistent Object Salience*||*Subject*) + (1|*Scene*)

Where *Object Semantics by Diagnostic Object Salience* is a four-level factor describing this interaction, contrast coded for simple effects

Model 3.1
*Ordinal first fixation number on diagnostic object ~ Side of Screen * Diagnostic Object Salience * Competition Type* + (1 + *Diagnostic Object Salience*: *Competition Type* ||*Subject*) + (1|*Scene*)

Model 4.1 (Diagnostic vs Inconsistent ROIs)
*Probability of fixating ROI ~ ROI Semantics * Diagnostic Object Presence * Inconsistent Object Presence* + (1 + *ROI Semantics:Diagnostic Object Presence* ||*Subject*) + (1|*Scene*)

Model 4.2 (Inconsistent ROI only)
*Probability of fixating ROI ~ Diagnostic Object Presence * Inconsistent Object Presence* + (1 + *Diagnostic Object Presence* + *Inconsistent Object Presence* ||*Subject*) + (1|*Scene*)

Model 4.3 (Diagnostic ROI only)
*Probability of fixating ROI ~ Diagnostic Object Presence * Inconsistent Object Presence* + (1|*Subject*) + (1|*Scene*)

Model 5.1
*Ordinal first fixation in Diagnostic ROI ~ Side of Scene * Diagnostic Object Presence * Inconsistent Object Presence* + (1 + *Diagnostic Object Presence* ||*Subject*) + (1|*Scene*)

Model 5.2 (simple effects)
Ordinal first fixation in Diagnostic ROI ~ *Side of Scene* * *Presence of the Two Objects* + (1|*Subject*) + (1|*Scene*)

Where *Presence of the Two Objects* is a four-level factor describing this interaction, contrast coded for simple effects

Model 6.1 (Diagnostic vs Inconsistent object)
Ordinal first fixation number on object ~ *Object Semantics* * *Diagnostic Object Salience* * *Inconsistent Object Salience* + (1 + *Object Semantics* + *Diagnostic Object Salience* + *Inconsistent Object Salience*||*Subject*) + (1|*Scene*)

Model 6.2 (Inconsistent object only)
Ordinal first fixation number on object ~ *Diagnostic Object Salience* * *Inconsistent Object Salience* + (1 + *Diagnostic Object Salience* + *Inconsistent Object Salience* ||*Subject*) + (1|*Scene*)

Model 6.3 (Diagnostic object only)
Ordinal first fixation number on object ~ *Diagnostic Object Salience* * *Inconsistent Object Salience* + (1 + *Diagnostic Object Salience* + *Inconsistent Object Salience* + *Diagnostic Object Salience*:*Inconsistent Object Salience* ||*Subject*) + (1|*Scene*)

Model 7.1 (Diagnostic vs Marginal object)
Ordinal first fixation number on object ~ *Object Semantics* * *Diagnostic Object Salience* * *Marginal Object Salience* + (1 + *Object Semantics* + *Diagnostic Object Salience* + *Marginal Object Salience*||*Subject*) + (1|*Scene*)

Model 7.2 and 7.3 (simple effects)
Ordinal first fixation number on object ~ *Object Semantics by Diagnostic Object Salience* * *Marginal Object Salience* + (1 + *Object Semantics* + *Diagnostic Object Salience* + *Marginal Object Salience*||*Subject*) + (1|*Scene*)

Where *Object Semantics by Diagnostic Object Salience* is a four-level factor describing this interaction, contrast coded for simple effects of salience (Model 7.2) and semantics (Model 7.3)

Model 7.4 and 7.5 (simple effects)
Ordinal first fixation number on object ~ *Object Semantics by Marginal Object Salience* * *Diagnostic Object Salience* + (1 + *Object Semantics* + *Diagnostic Object Salience* + *Marginal Object Salience*||*Subject*) + (1|*Scene*)

Where *Object Semantics by Marginal Object Salience* is a four-level factor describing this interaction, contrast coded for simple effects of salience (Model 7.4) and semantics (Model 7.5)

Model 8.1 (Marginal vs Inconsistent object)
Ordinal first fixation number on object ~ *Object Semantics* * *Diagnostic Object Salience* * *Non-diagnostic Object Salience* + (1 + *Diagnostic Object Salience* + *Non-diagnostic Object Salience*||*Subject*) + (1|*Scene*)

Model 8.2 (simple effects) and Model 8.3 (simple effects)
*Ordinal first fixation number on object ~ Non-diagnostic Object Semantics by Non-diagnostic Object Salience * Diagnostic Object Salience* + (1 + *Diagnostic Object Salience + Inconsistent Object Salience*||*Subject*) + (1|*Scene*)

Where *Non-diagnostic Object Semantics by Non-diagnostic Object Salience* is a four-level factor describing this interaction, contrast coded for simple effects of salience (Model 8.2) or semantics (Model 8.3)

Model 9.1 (Diagnostic vs Inconsistent object)
*Log(Total fixation time) ~ Object Semantics * Diagnostic Object Salience * Inconsistent Object Salience* + (1 + *Diagnostic Object Salience * Inconsistent Object Salience*||*Subject*) + (1|*Scene*)

Model 9.2 (simple effects) and Model 9.3 (simple effects)
*Log(Total fixation time) ~ Object Semantics by Diagnostic Object Salience * Inconsistent Object Salience* + (1 + *Object Semantics + Diagnostic Object Salience + Inconsistent Object Salience*||*Subject*) + (1|*Scene*)

Where *Object Semantics by Diagnostic Object Salience* is a four-level factor describing this interaction, contrast coded for simple effects of semantics (Model 9.2) or salience (Model 9.3)

Model 9.4 (simple effects) and Model 9.5 (simple effects)
*Log(Total fixation time) ~ Object Semantics by Inconsistent Object Salience * Diagnostic Object Salience* + (1 + *Object Semantics + Diagnostic Object Salience + Inconsistent Object Salience*||*Subject*) + (1|*Scene*)

Where *Object Semantics by Inconsistent Object Salience* is a four-level factor describing this interaction, contrast coded for simple effects of semantics (Model 9.4) or salience (Model 9.5)

Model 10.1 (Diagnostic vs Marginal object)
*Log(Total fixation time) ~ Object Semantics * Diagnostic Object Salience * Marginal Object Salience* + (1 + *Diagnostic Object Salience * Marginal Object Salience*||*Subject*) + (1|*Scene*)

Model 10.2 (Marginal object only)
*Log(Total fixation time) ~ Diagnostic Object Salience * Marginal Object Salience* + (1 + *Diagnostic Object Salience * Marginal Object Salience*||*Subject*) + (1|*Scene*)

Model 10.3 (Diagnostic object only)
*Log(Total fixation time) ~ Diagnostic Object Salience * Marginal Object Salience* + (1 + *Diagnostic Object Salience * Marginal Object Salience*||*Subject*) + (1|*Scene*)

Model 10.4 (simple effects)
Log(Total fixation time) ~ Object Semantics by Diagnostic Object Salience by Marginal Object Salience (1 + *Diagnostic Object Salience + Marginal Object Salience*||*Subject*) + (1|*Scene*)

Where *Object Semantics by Diagnostic Object Salience by Marginal Object Salience* is an eight-level factor describing this interaction, contrast coded for simple effects of semantics

Model 11.1 (Marginal vs Inconsistent object)
*Log(Total fixation time) ~ Non-diagnostic Object Semantics * Diagnostic Object Salience * Non-diagnostic Object Salience* + (1 + *Non-diagnostic Object Salience*||*Subject*) + (1|*Scene*)

Model 11.2 (simple effects)
Log(Total fixation time) ~ Non-diagnostic Object Semantics by Diagnostic Object Salience by Non-diagnostic Object Salience (1 + *Diagnostic Object Salience*||*Subject*) + (1|*Scene*)

Where *Non-diagnostic Object Semantics by Diagnostic Object Salience by Non-diagnostic Object Salience* is an eight-level factor describing this interaction, contrast coded for simple effects of semantics

Model 12.1
*Probability of detecting the inconsistent change first ~ Task * Side of Scene * Diagnostic Object Salience * Inconsistent Object Salience* + (1 + *Diagnostic Object Salience* + *Inconsistent Object Salience* ||*Subject*) + (1 + *Task* + *Side of Scene* ||*Scene*)

Model 12.2 (Detect one change) and Model 12.3 (Detect two changes)
*Probability of detecting the inconsistent change first ~ Side of Scene * Diagnostic Object Salience * Inconsistent Object Salience* + (1 + *Diagnostic Object Salience* + *Inconsistent Object Salience* ||*Subject*) + (1 + *Side of Scene* ||*Scene*)

Model 12.4 (Detect two changes: simple effects)
Probability of detecting the inconsistent change first ~ Salience of the Two Objects + (1 + *Diagnostic Object Salience* + *Inconsistent Object Salience* ||*Subject*) + (1 + *Side of Scene* ||*Scene*)

Where *Salience of the Two Objects* is a four-level factor describing this interaction, contrast coded for simple effects

Model 13.1
*Log(Time to detect first change) ~ Task * Semantics of the First Detected Change * Diagnostic Object Salience * Inconsistent Object Salience* + (1 + *Diagnostic Object Salience* + *Inconsistent Object Salience* ||*Subject*) + (1 + *Task* ||*Scene*)

Model 13.2 (simple effects)
*Log(Time to detect first change) ~ Task * Diagnostic Object Salience * Semantics of the First Detected Change by Inconsistent Object Salience* + (1 + *Diagnostic Object Salience* + *Inconsistent Object Salience* ||*Subject*) + (1 + *Task* ||*Scene*)

Where *Semantics of the First Detected Change by Inconsistent Object Salience* is a four-level factor describing this interaction, contrast coded for simple effects

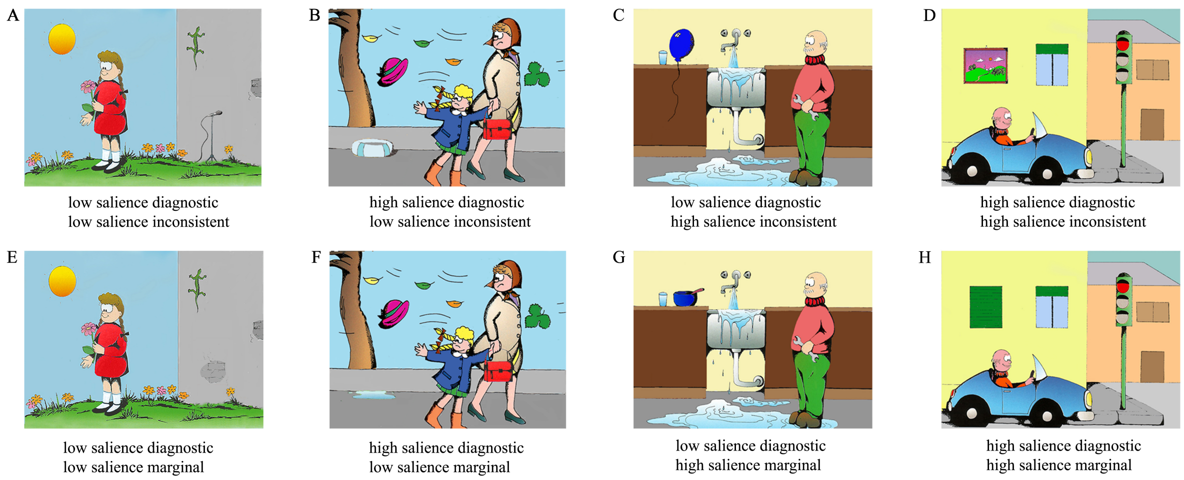

Figure 1. Examples of scenes with the two critical objects present in Experiments 1 (top row) and 2 (bottom row). The four experimental conditions for perceptual salience (low vs. high) are depicted across columns. A-D show scenes from Experiment 1, in which each scene contains a diagnostic and inconsistent object as follows. A: diagnostic low salience object: flower (in hand); inconsistent low salience object: microphone. B: diagnostic high salience object: hat; inconsistent low salience object: rubber ring. C: diagnostic low salience object: spanner; inconsistent high salience object: balloon. D: diagnostic high salience object: traffic light; inconsistent high salience object: painting. E-F show versions of the scenes created for Experiment 2, where inconsistent objects were replaced with objects that were consistent but of low (marginal) informativeness for the scene. Marginal objects were matched for salience, ROI size and placement with the inconsistent objects they replaced.

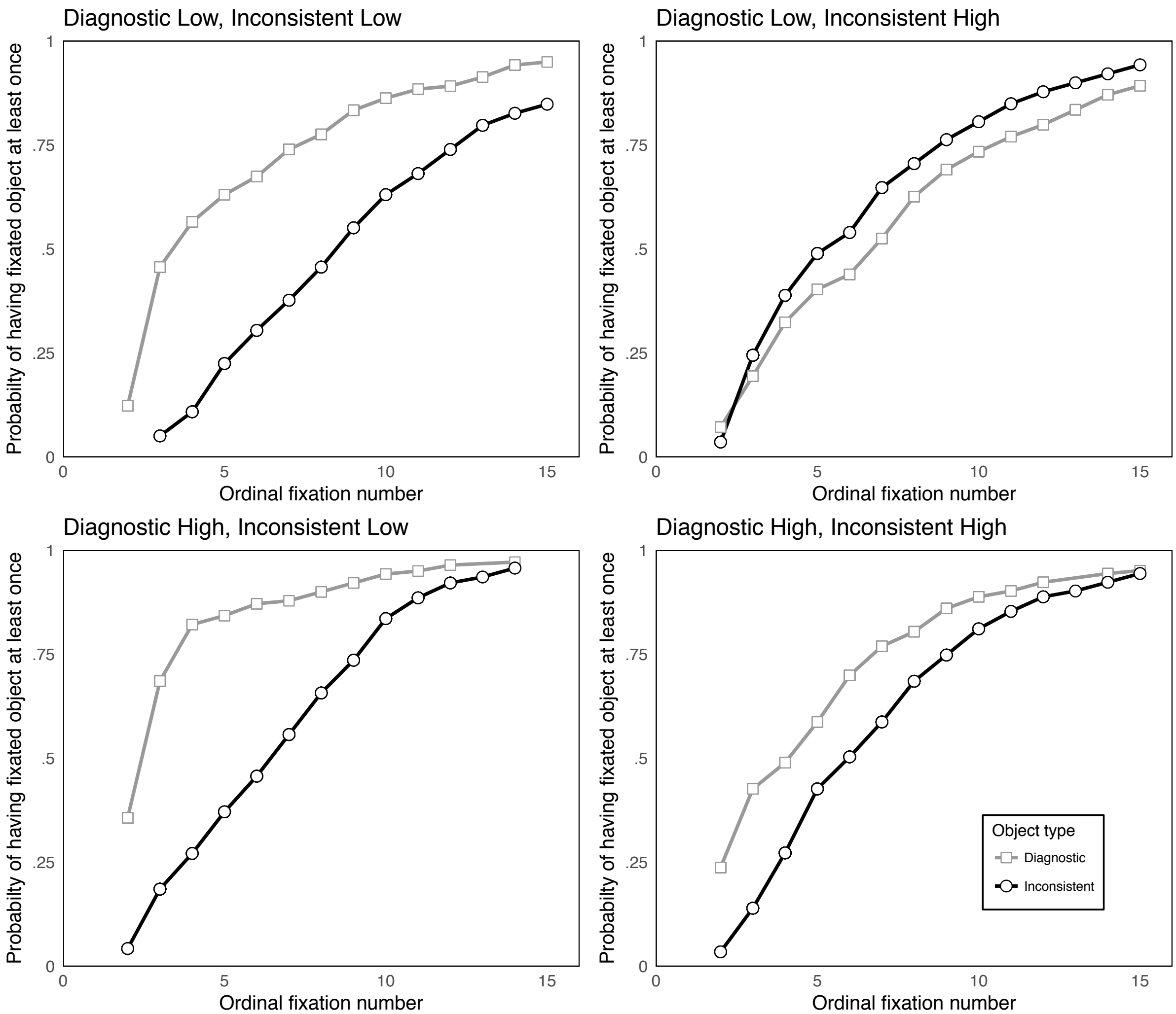

Figure 2. Experiment 1: the conditional (cumulative) probability of fixating each of the two critical objects as a function of ordinal fixation number in viewing. The four panels correspond to the four possible combinations of perceptual properties of the two critical objects. Here data are collapsed across participants and scenes. We plot data for the first 16 fixations for comparability with Loftus and Mackworth (1978).

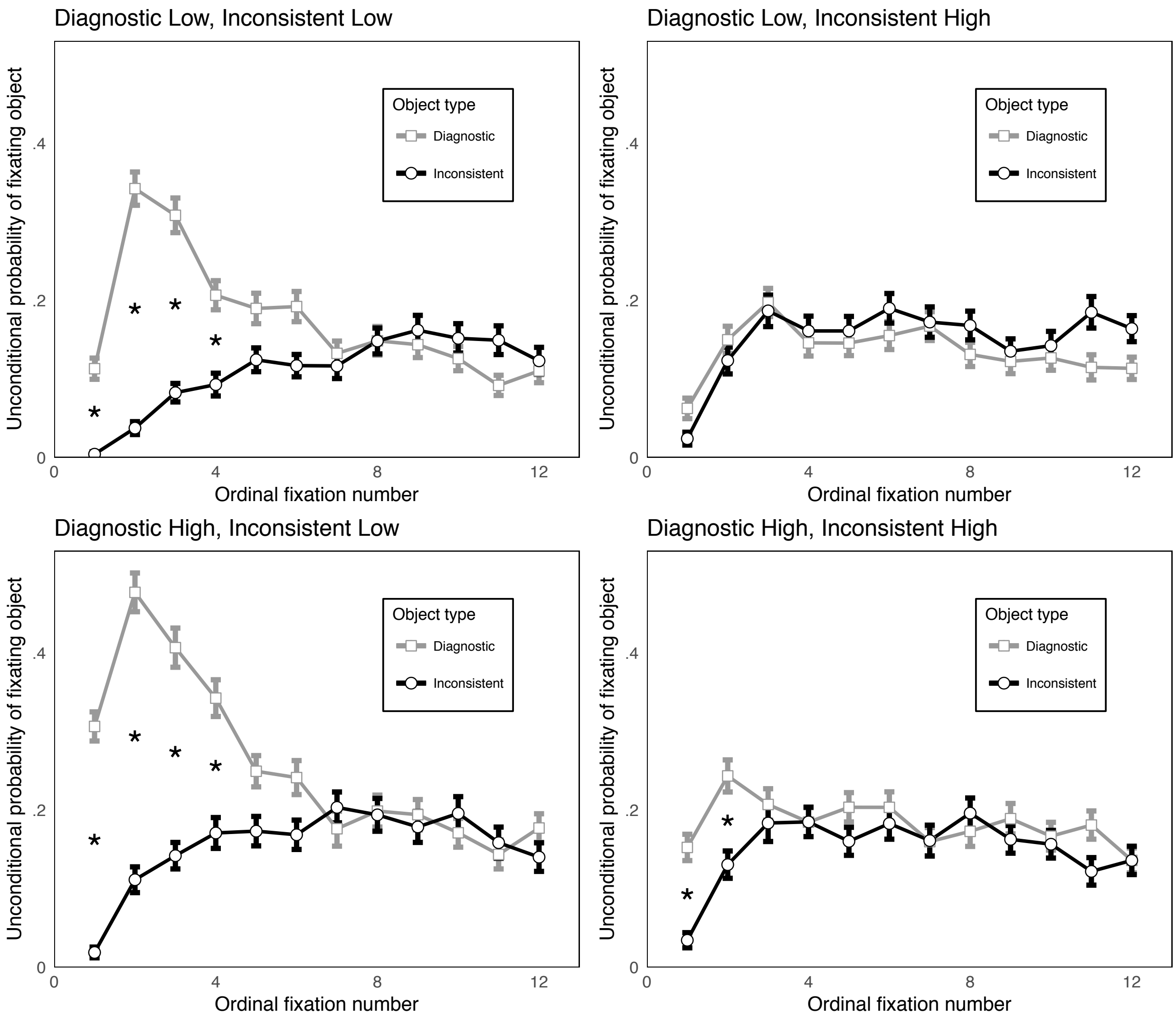

Figure 3. Experiment 1: the unconditional probability of fixating each of the two critical objects as a function of ordinal fixation number in viewing. Panels show the different perceptual relationships between the two critical objects. Data are shown for means calculated for each participant with error bars denoting one standard error of the mean across participants. Stars indicate pairwise comparisons that were significant after Bonferroni correction for multiple comparisons (corrected $\alpha = .05/12$). We plot data for the first 12 fixations for comparability with Loftus and Mackworth (1978).

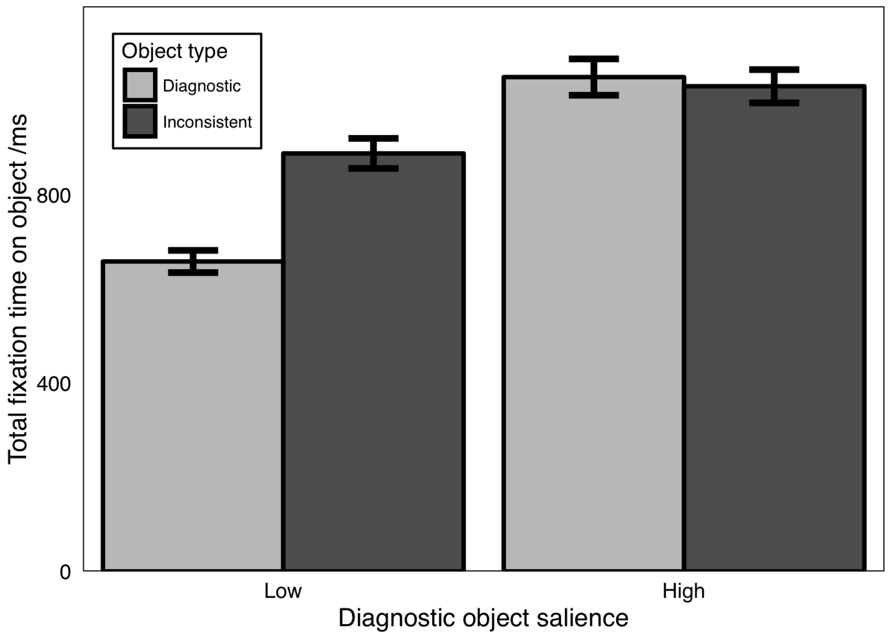

Figure 4. Experiment 1: the effect of the salience of the diagnostic object on the total time spent fixating the critical object during viewing, shown for both the diagnostic and inconsistent objects. Error bars show one standard error around the mean.

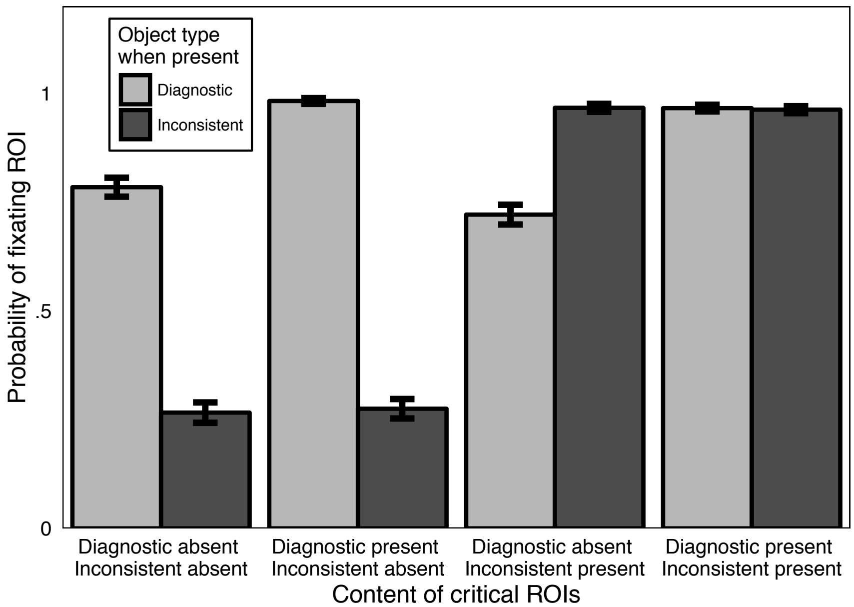

Figure 5. Experiment 1: the probability that each of the two critical regions (diagnostic and inconsistent) of the scene were fixated as a function of whether or not each of the two critical objects was present in these ROIs. Error bars show one standard error around the mean.

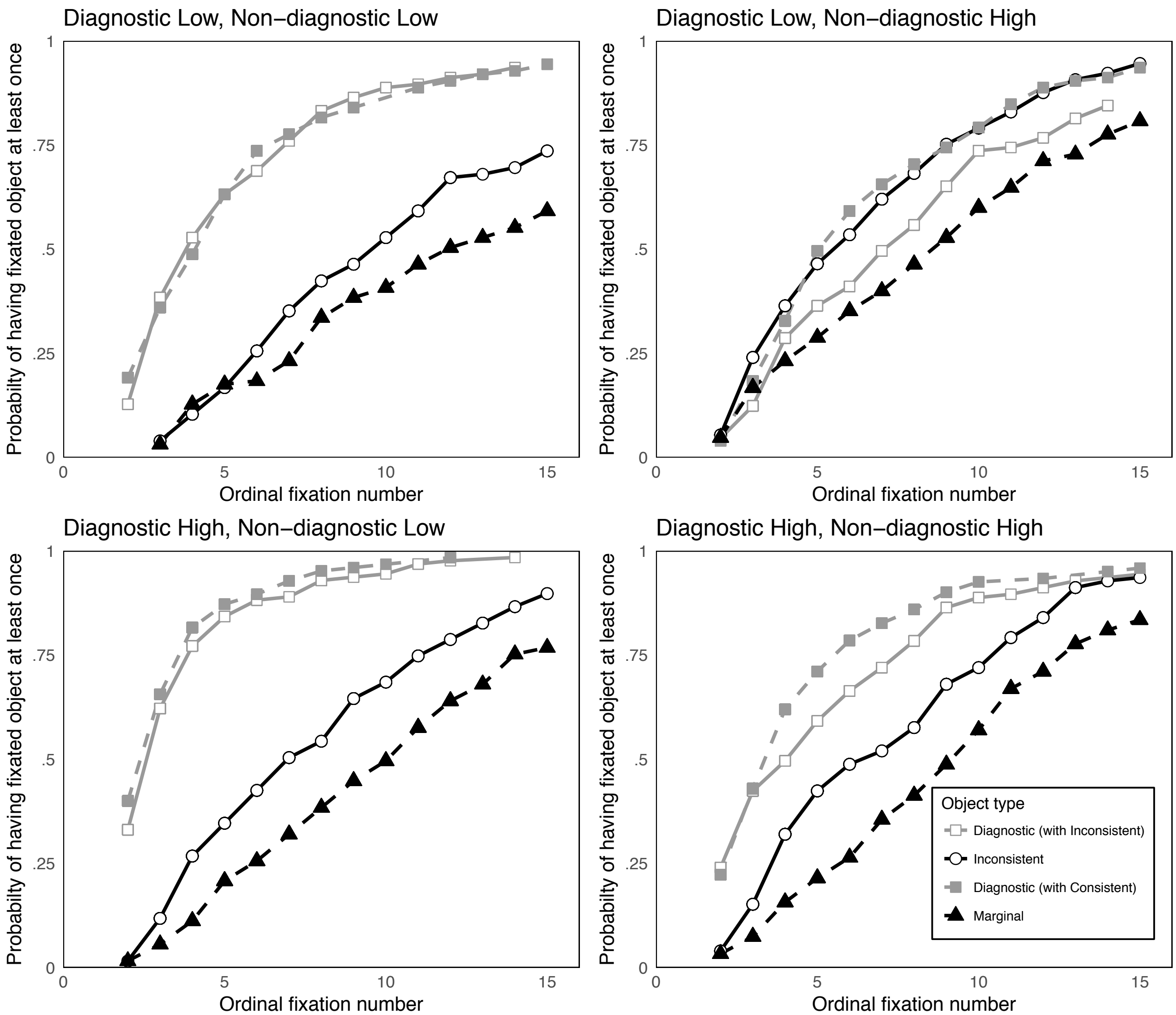

Figure 6. Experiment 2: the conditional (cumulative) probability of fixating each of the two critical objects as a function of ordinal fixation number in viewing. The four panels correspond to the four possible combinations of perceptual properties of the two critical objects. In each panel data are shown for trials with competition between diagnostic and inconsistent objects (open symbols, solid lines) and for trials with competition between diagnostic and marginal objects (filled symbols, dashed lines). Here data are collapsed across participants and scenes. We plot data for the first 16 fixations for comparability with Loftus and Mackworth (1978).

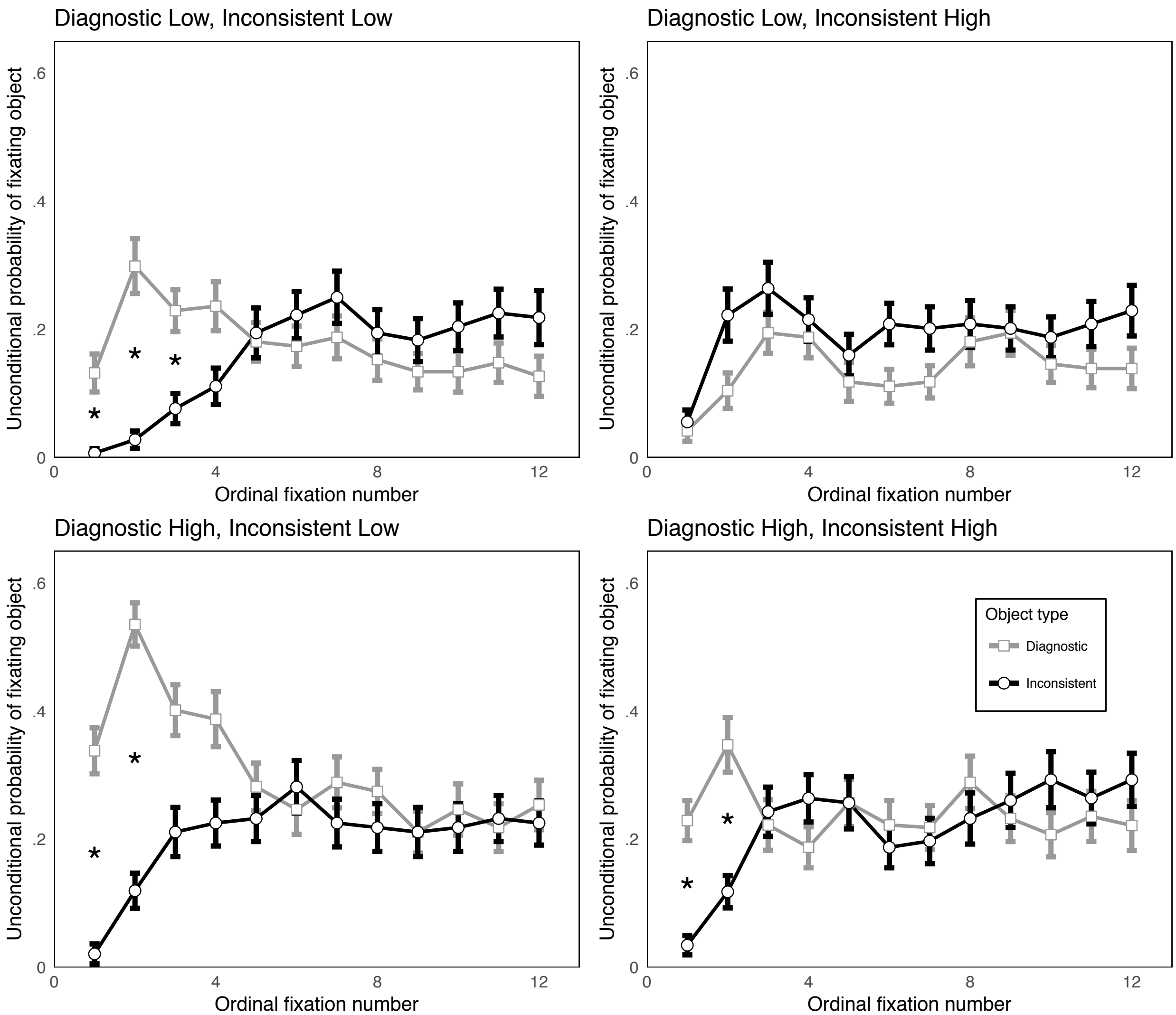

Figure 7. Experiment 2: the unconditional probability of fixating the diagnostic and inconsistent objects in trials where they were in direct competition, as a function of ordinal fixation number in viewing. Data plotted are between-participant means (±1 *SEM*). Stars indicate pairwise comparisons that were significant after Bonferroni correction for multiple comparisons (corrected α = .05/12).

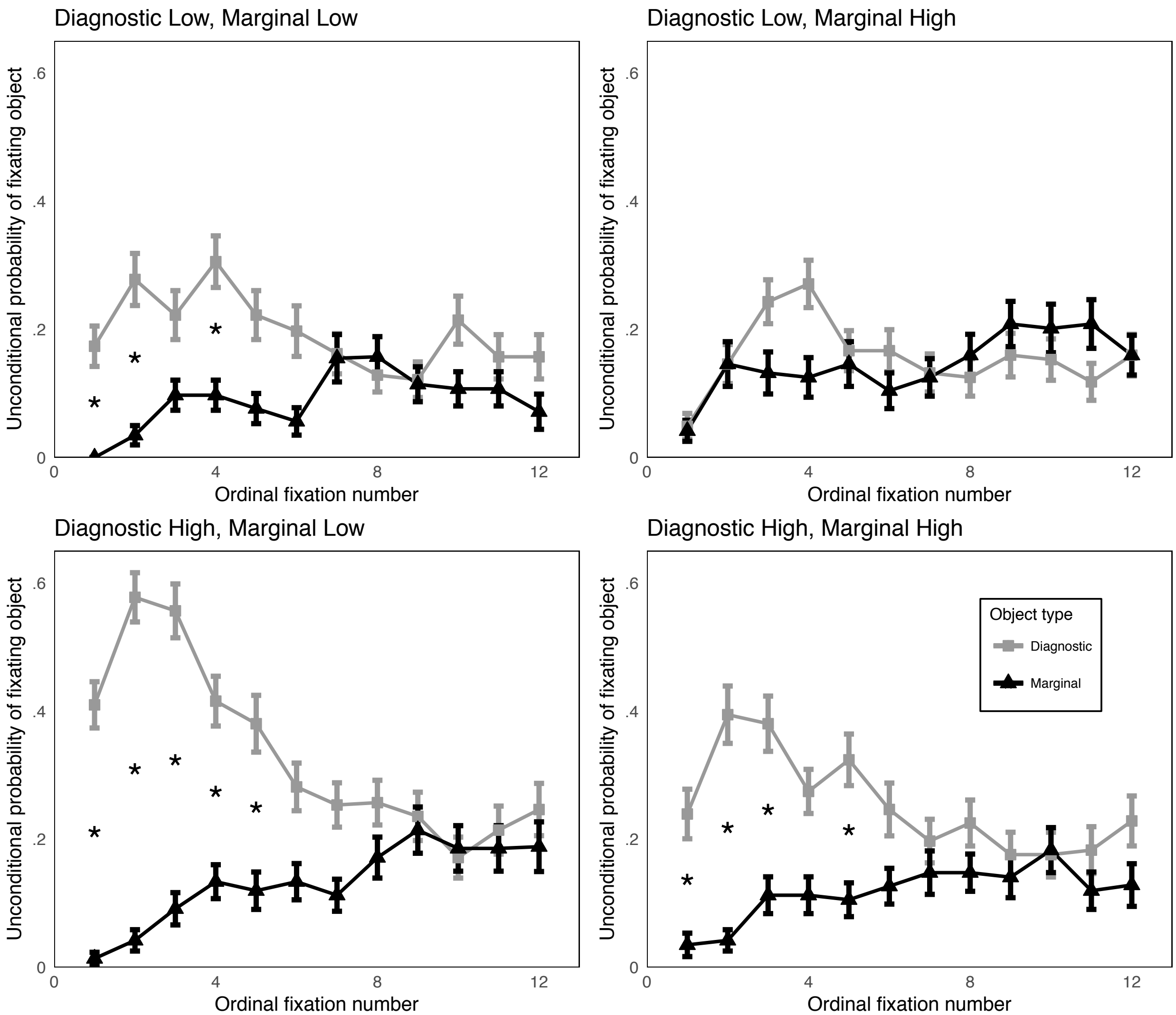

Figure 8. Experiment 2: the unconditional probability of fixating the diagnostic and marginal objects in trials where they were in direct competition, as a function of ordinal fixation number in viewing. Data plotted are between-participant means (±1 *SEM*). Stars indicate pairwise comparisons that were significant after Bonferroni correction for multiple comparisons (corrected α = .05/12).

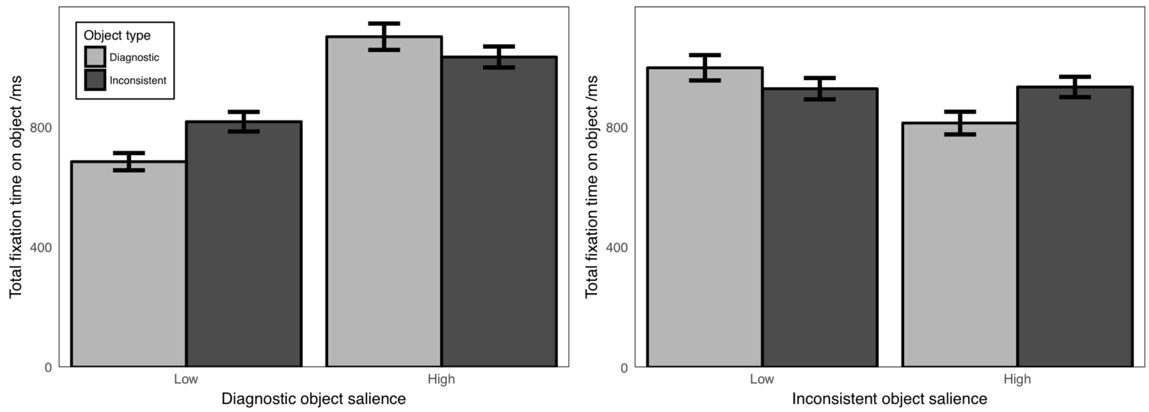

Figure 9. Experiment 2: total inspection time for the diagnostic and inconsistent critical objects. Left panel, the two-way interaction between diagnostic object salience and object semantics. Right panel, the two-way interaction between inconsistent object salience and object semantics. Error bars show one standard error around the mean.

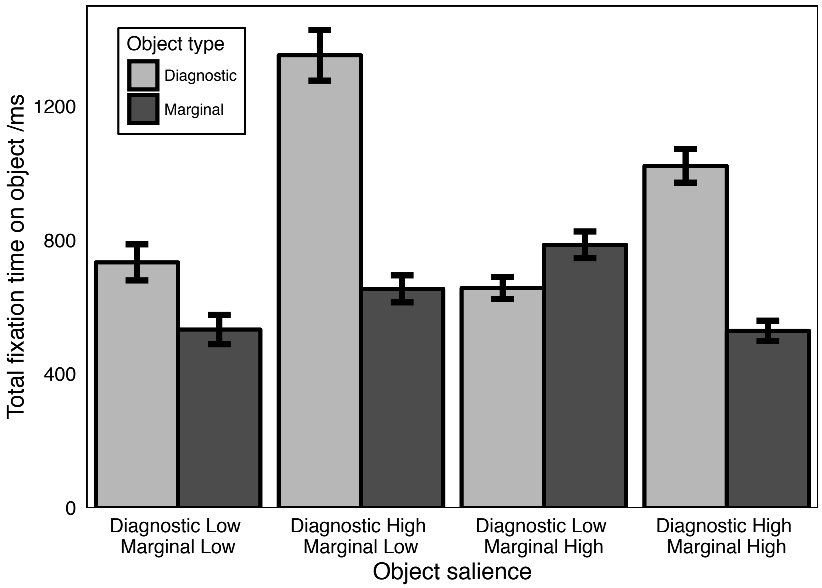

Figure 10. Experiment 2: total inspection time for the diagnostic and marginal objects across the four salience conditions describing the perceptual properties of the two objects. Error bars show one standard error around the mean.

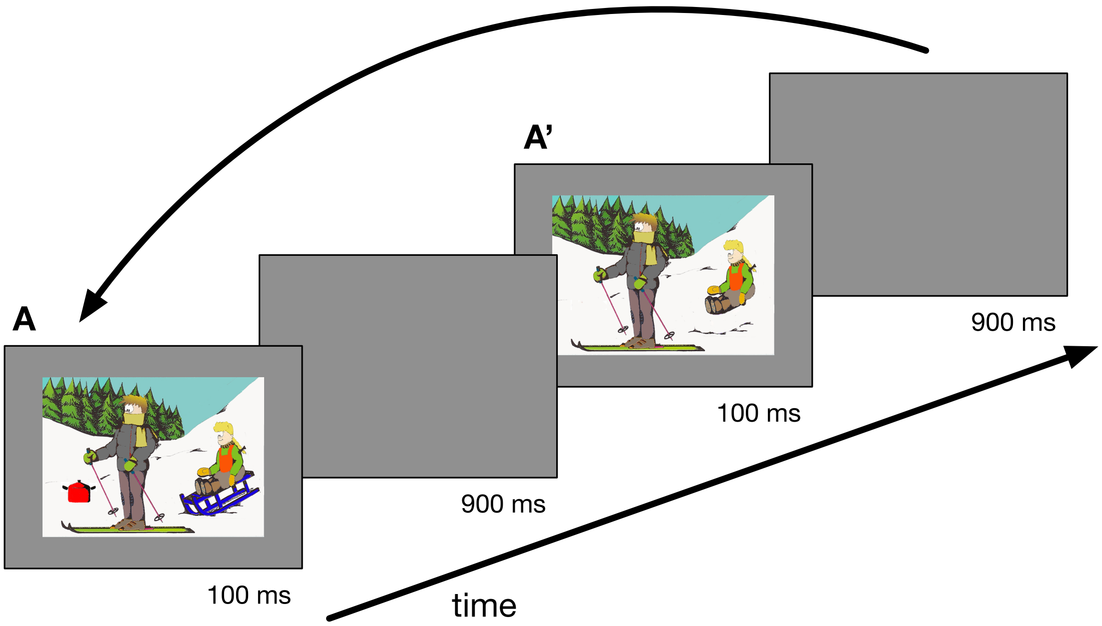

Figure 11. Example of screen shots of trial, Experiment 3 (Change Detection). This example shows the condition in which both the inconsistent (the pot) and the diagnostic (the sleigh) critical objects were highly salient. Cycles of A/A' scenes were presented until when both changes were found or 60 s elapsed. This depiction does not respect the original proportions scene/background, as the scene is here is larger for illustration purpose. Each trial began with a central fixation marker, here not presented. Please refer to the Supplementary Materials for a full-colour version of this figure.

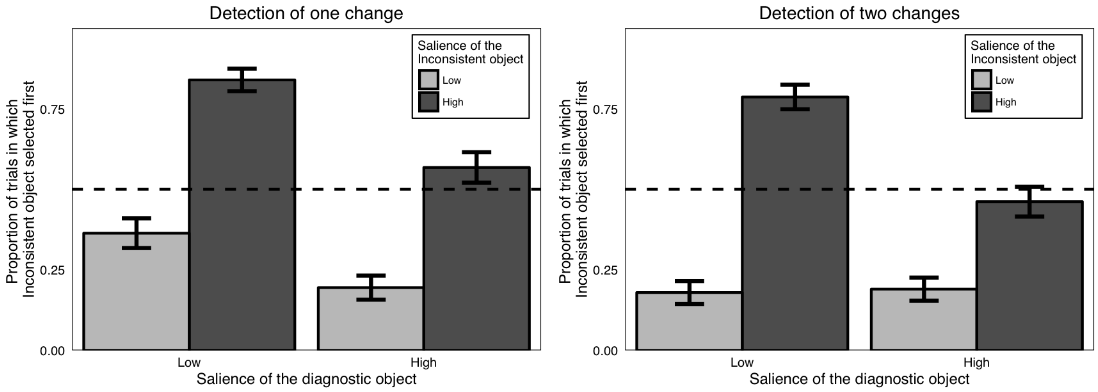

Figure 12. Experiment 3: probability of detecting the inconsistent change first, as a function of the perceptual salience of the diagnostic object and the perceptual salience of the inconsistent object (both presented and changing in each scene at the same time). The left panels shows data when participants were asked to find one change. The right panel shows data when participants were asked to find both changes in the scene. Error bars show one standard error around the mean. The dashed lines indicate chance level.