Central Lancashire Online Knowledge (CLoK)

Title	Cat and mouse search: the influence of scene and object analysis on eye movements when targets change locations during search
Type	Article
URL	https://clok.uclan.ac.uk/22337/
DOI	https://doi.org/10.1098/rstb.2016.0106
Date	2017
Citation	Hillstrom, Anne P., Segabinazi, Joice D., Godwin, Hayward J., Liversedge, Simon Paul and Benson, Valerie (2017) Cat and mouse search: the influence of scene and object analysis on eye movements when targets change locations during search. Philosophical Transactions of The Royal Society B Biological Sciences, 372 (1714). pp. 1-9. ISSN 0962-8436
Creators	Hillstrom, Anne P., Segabinazi, Joice D., Godwin, Hayward J., Liversedge, Simon Paul and Benson, Valerie

PHILOSOPHICAL
TRANSACTIONS B

Phil. Trans. R. Soc. B.
doi:10.1098/not yet assigned

Cat and mouse search: The influence of scene and object analysis on eye movements when targets change locations during search.

Anne P. Hillstrom[1]*, Joice D. Segabinazi[2], Hayward J. Godwin[1], Simon P. Liversedge[1], Valerie Benson[1]*

[1]*Psychology Department, University of Southampton, Shackleton Building, Highfield Campus, Southampton SO17 1BJ, United Kingdom*
[2]*CAPES Foundation, Ministry of Education of Brazil, Brasília – DF 70040-020, Brazil, Federal University of Rio Grande do Sul*

Keywords: *Scene gist, visual search, eye movement planning*

Note

This is a preprint version of an article published in January 2017. The published version, which has a few small differences from this version, can be found at: http://rstb.royalsocietypublishing.org/content/372/1714/20160106

Main Text

Summary

We explored the influence of early scene analysis and visible object characteristics on eye movements when searching for objects in photographs of scenes. On each trial, participants were shown sequentially either a scene preview or a uniform grey screen (250ms), a visual mask, the name of the target, and the scene, now including the target at a likely location. During the participant's first saccade during search, the target location was changed to: (1) a different likely location, (2) an unlikely but possible location, or (3) a very implausible location. The results showed that the first saccade landed more often on the likely location in which the target re-appeared than on unlikely or implausible locations, and overall the first saccade landed nearer the first target location with a preview than without. Hence, rapid scene analysis influenced initial eye movement planning, but availability of the target rapidly modified that plan. After the target moved, it was found more quickly when it appeared in a likely location than when it appeared in an unlikely or implausible location. The findings show that both scene gist and object properties are extracted rapidly, and are used in conjunction to guide saccadic eye movements during visual search.

Introduction

---

* Authors for correspondence (a.hillstrom@soton.ac.uk, v.benson@soton.ac.uk).
†Present address: Department, Institution, Address, City, Code, Country

Scene perception research helps us in understanding how we interact with the real world. One such interaction is visual search, which is a fundamental process underlying object selection in scene perception [1]. When people search visual scenes, they direct their attention using goals (task-oriented use of knowledge), object salience (involuntary attraction to stimulus properties) and an interaction between the two to guide, or prioritise, where in the scene they look [2]. The role of knowledge has been explored in further research that investigates how search in scenes is guided by the characteristics of the target object and by knowledge about the scene. It is widely agreed that when searching for a specific object, i.e. a cup of coffee at the office, the searcher will primarily look at objects with the characteristics of the target, e.g., objects that look like a coffee cup [3-4]. The role of scene knowledge has been demonstrated by giving foreknowledge about a scene by showing a brief preview of it prior to defining the search target. Doing so reduces the time to find objects [5-8]. Influential factors include the preexisting representations of scene structure and the "fit" of the object to semantics of the scene and to locations in the scene [6-9]. Importantly, the scene gist information provided by a preview can guide initial eye movements but has little impact on targeting of later eye fixations [10].

The clearest evidence about how knowledge about scenes affects search comes from studies that manipulate the plausibility of the target appearing in the scene or the plausibility of the target location in the scene. A seminal paper from Loftus and Mackworth [11] investigated fixations on objects semantically consistent or inconsistent with the scene during free viewing of pictures. Their findings showed that an inconsistent object was fixated on earlier and for a longer duration than a consistent object, though attempts to replicate this finding have had mixed results [12-15]. When the task is to search for objects that might be semantically consistent or inconsistent with scene context, attention is not initially drawn to inconsistent targets [6, 16] but response times are longer. Both of the studies showing this result limited visibility of the scene to a small window during search; we could find no studies that investigated the speed of searching for semantically inconsistent targets in fully visible scenes, and so the influence of semantic consistency on guidance of later stages of search is not yet fully understood.

Knowledge about where objects might appear in scenes, sometimes called syntactic knowledge [17], also has an effect on search. In two studies that manipulated the plausibility of target locations in scenes, objects that appeared in likely locations were found more quickly than targets that appeared in unlikely locations, even though the use of windowed viewing during search meant that detailed scene syntax could only be derived during the preview of the scene [6, 16]. Another study used fully visible scenes, and showed that the first saccade during search landed closer to the likely target location than to the unlikely target location, both when the target was present and when it was absent [9]. No information was reported about the rest of search. The studies summarised so far, then, only speak to the influence of location plausibility on guidance at the start of search.

A few studies have used scenes fully visible during the entirety of search. In one, targets only appeared in likely locations, but the likely location of some objects were constrained by the scene (e.g., a jeep on the ground rather than in the air) and the likely location of other objects were less constrained (e.g., a helicopter could appear either on the ground or in the air) [18]. When the target was present, fixations were mostly limited to locations where the target object was expected to be; when the target was absent, search was less restricted to these locations. In addition, they found search times were on average 19% faster when the target was in expected areas. Another study asked participants to count instances of an object type (e.g., mugs or pictures) that were likely to appear in certain regions of pictures. A model of where they looked was more accurate if location-likelihood information was included in the model than if it was not [19].

In summary, what we have is strong evidence that location-likelihood information influences guidance of the first fixation of search, and that as long as there is no reason to expect objects to appear in unlikely locations, guidance of search will constrain the eyes mostly to likely locations. We also know that when target locations are manipulated so that targets sometimes appear in implausible locations, search is overall less efficient if targets appear in implausible locations than if they appear in plausible locations. But we do not know whether that efficiency is entirely driven by the start of search or whether scene syntactic knowledge has continued power throughout the search, if visible features in the scene are available to compete with syntactic knowledge.

There is still a dispute in the literature considering the degree to which eye movements during scene perception are influenced by bottom-up image properties such as contrast or color, on one hand [20-23], or by top-down factors such as scene gist or likely object location. In our earlier work [10], when the scene was fully visible, and so bottom-up characteristics were available to influence search, scene gist benefits were short lived in relation to the planning of saccades. It is possible that from the point the target was known and visible, the target's visible features outweighed scene gist information, including knowledge of likely locations, in guiding eye movements.

The experiment presented here was developed to address whether, when both visible target characteristics and location likelihood information are available, guidance toward a target is still influenced by location-likelihood. Participants were initially either shown a preview of the scene or not. The preview showed the scene without the target. When search began, the target was in a likely location. As such, both location likelihood and visual characteristics of the target were expected to guide attention toward the target's location at the start of search, and there was more opportunity for location likelihood to exert an influence when the preview was presented than when it was not. Based on previous research, we might expect that the first saccade would be directed toward the initial likely location, which contained the target, and that the likelihood that the first saccade would be aimed there might be higher and the saccade made faster with a scene preview than without.

However, the main empirical question related to what would happen next if the location of the target object moved as soon as the participant began a saccade. In this experiment, therefore, upon initiation of the first saccade, we changed the target location to one of three other locations: another likely location, an unlikely location or a physically implausible location. We liken this target movement during saccades to a form of "cat and mouse" search where the target (mouse) jumps around with the searcher (cat) rapidly trying to keep up with it. We were interested to know whether participants found the relocated target more quickly when it appeared in a (different) likely location than when it appeared in a (different) less likely location. If location likelihood remained influential in guiding eye movements, then we would expect that targets moved to more likely locations would be found more quickly than targets moved to less likely, or implausible, locations. Alternatively, if the visible characteristics of the target were more influential than location likelihood, then search for the target should be equally efficient regardless of which of the three locations to which it was ultimately moved.

Method

Design

The experiment had a 2 (Preview vs No-Preview) by 3 (likely vs unlikely vs implausible final target location) repeated measures design. There were multiple scenes presented to each participant. For each scene, a preview was present or absent, and the final target location during search was in a likely, unlikely or implausible location.

Participants

Twenty-two undergraduate/postgraduate students at University of Southampton (13 female) ranging in age between 19 and 33 (*M*=23.27, *SD*=2.9) participated in the study for partial course credit or for £6/hour. All participants reported normal or corrected-to-normal vision and were unfamiliar with the stimuli material.

Materials

Apparatus

Eye movements were recorded with an EyeLink1000 tower system (SR Research, Canada), which tracks with a resolution of 0.018 visual angle at a sampling rate of 1000 Hz. The position of the right eye was tracked whilst viewing was binocular. A head restraint and chin rest ensured that the head position was stable during eye movement recording, with the eyes 66 cm from the screen. Stimulus presentation and reaction time recording was controlled by Experimental Builder (SR-

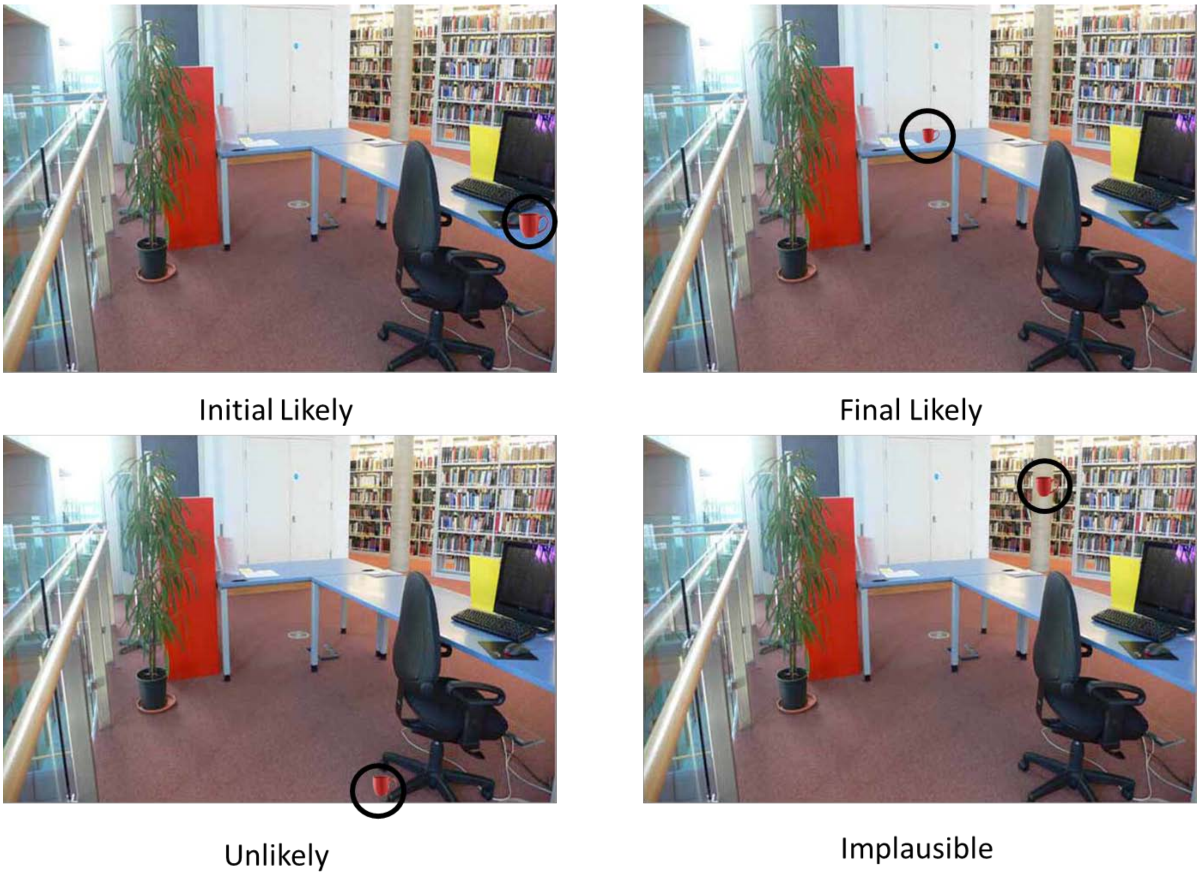
Research Ltd, Canada). Stimuli were presented on a ViewSonic P227f CRT monitor with a 100Hz refresh rate and a resolution of 1280 x 1024 pixels.

Stimuli

The search scenes were colour images, and consisted of 36 photographed scenes collected from the internet combined with 36 objects (one object for each scene) from the Hemera photo-objects library (Focus Multimedia, Ltd). The images varied in dimension, ranging from 550 x 734 pixels to 1024 x 768 pixels. For each scene, we built four variants differing in the location of the target object -- initial likely, final likely, unlikely and implausible – giving rise to 144 different images in total. See Figure 1 for an example of a target object in four different locations. For any one participant, the 36 scenes were divided amongst the 6 conditions (preview or no preview crossed with final target position being likely, unlikely or implausible), and so each participants saw each scene in only one trial. The assignment of scenes to conditions was counterbalanced across participants, but because of the number of participants run and the loss of one participant’s data, full counterbalancing was not completed.

Each target object appeared in only one trial, and the link between target and scene remained fixed across participants. Of the targets, 32 were graspable man-made (manufactured objects or packaged food) and 4 were organic (3 fruits and one plant).

Unlikely locations were locations that were physically possible but unlikely to be used in real life (e.g., a mug placed on a neatly made bed). Implausible locations were locations that would require the object to either float in air or to be attached in a peculiar way to another part of the scene (e.g., a mug seemingly attached to the glass of a mirror). Biederman labeled these position and support violations, respectively [17]. More recently, Võ and Wolfe identified different electrophysiological responses when observers looked at objects exhibiting the two kinds of violations [25]. The size of the object was adjusted to be appropriate to the depth at which it appeared in the scene. For example, if the second location was closer in depth, the object appeared larger. As



Initial Likely

Final Likely

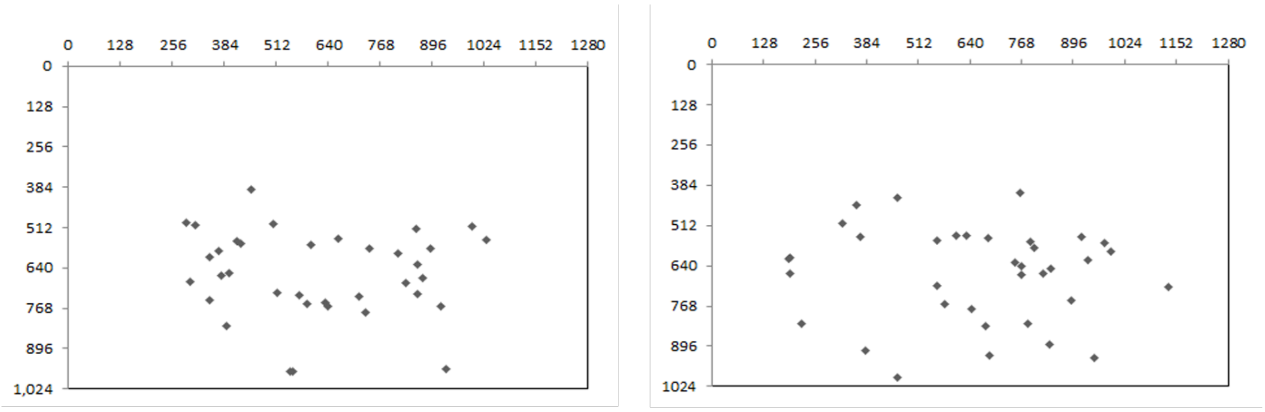
Unlikely

Implausible

shown in Figure 2, likely locations were mostly in the bottom half of the images, implausible objects were mostly in the upper half of images and unlikely locations were fairly distributed across the two.

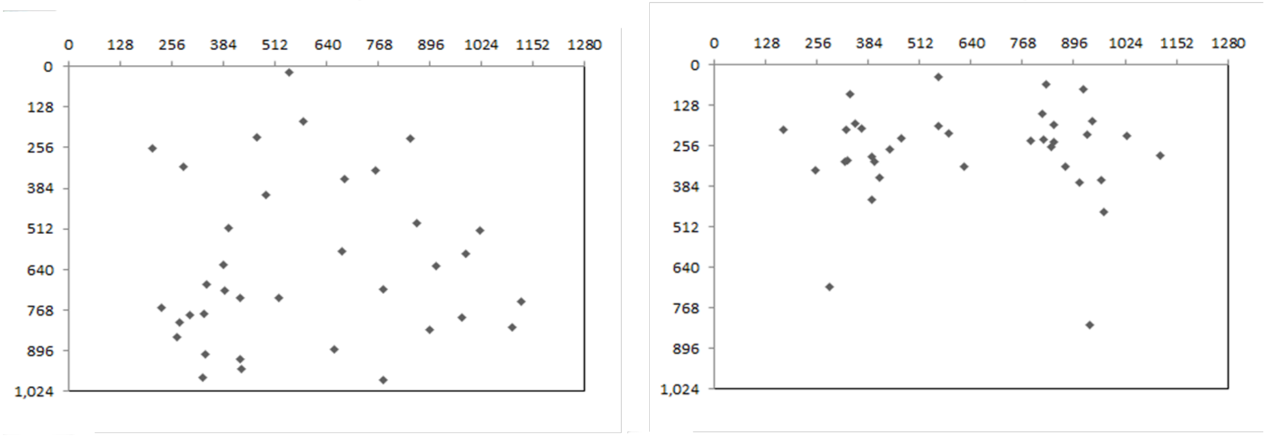
Figure 1. A representative scene with four different locations for the same target (mug). Circles have been added around targets to aid the reader. For copyright reasons, this image is not one of the scenes used in the study.

A 1280 x 1024 image consisting of small, randomly positioned black spots was used as a mask.



Initial Likely

Final Likely



Unlikely

Implausible

Figure 2. Spatial distributions of the initial target locations and, for the three location-plausibility conditions, the final target locations.

Procedures

Eye movement recording: A calibration procedure required participants to fixate nine dots on the screen, each appearing one at a time and covering the dimensions of the screen. Calibration was followed by a validation procedure. A calibration was accepted when the average error during validation was less than 0.6° of visual angle (mean = 0.30° of visual angle) with no one point exceeding 1.3° (mean = 0.59°).

Experimental Trials: Each trial sequence was as follows (See Figure 3). First a fixation spot was presented until the participant looked at it steadily. The location of the fixation spot was randomly selected from all positions within 150 pixels width and height from the centre of the scene. Following the fixation spot, either a preview of the scene (identical to the searched scene, except the target object was not present) or a grey screen appeared for 250ms, centred in the display, followed by the mask that filled the display for 50ms. The name of the target, written in Times New Roman 20 font, appeared after the mask, and remained on the screen until the participant looked at it. The location at which the name of the target was shown was randomly chosen from four locations that were not initial or final target locations for the scene. The scene then reappeared showing the target in the initial likely location. It was displayed until the participant started their first saccade, as measured by a velocity of over 30 degrees per second or an acceleration of over 8000 degrees per second squared. During that first saccade, the scene changed to show the target in one of the three other locations: likely, unlikely, or implausible. The scene with the target in the new location then remained visible and unchanged until the participant indicated that they had found the target by pressing a button while looking where they believed the target to be.

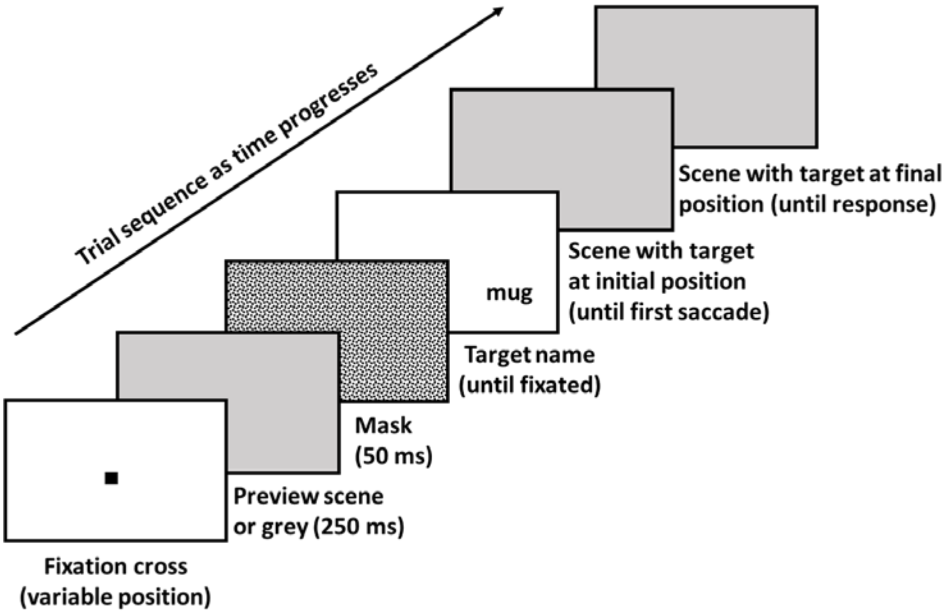


Figure 3. Trial sequence.

The experiment consisted of one session of approximately thirty minutes for each participant to complete the 36 trials.

Results

Manipulation Checks for Stimuli

Table 1 shows target statistics that reflect whether targets were equally detectable in different conditions. The Saliency Toolbox [26], a MATLAB toolbox, was used to produce a salience map for each image based on the luminance, colour and orientation throughout the image, and the value we report for target salience was the value in the salience map at the target location. Because the saliency values produced in the maps are not normally distributed, a non-parametric test was used to compare salience in the implausible, unlikely, and likely location variants of the 36 scenes. A related-samples Friedman's ANOVA by ranks found no significant difference in salience as final-location type varied, $\chi^2 = 2.970$, $p = .227$. The distance in the picture plane from initial-fixation location (first fixation at which target was in its final location) to final target location varied significantly with the final-location type, $F(2,70) = 6.816$, $p=.002$, as did the distance from centre of the image to final target location, $F(2,70) = 5.753$, $p=.005$. Follow-up tests (Bonferroni corrected) of distance from fixation to target showed that targets in the final likely location were somewhat closer to the location of the first fixation in which they were visible than targets in implausible locations, $p=.001$, and marginally closer than targets in unlikely locations, $p=.059$. If size was suitably larger when distance was longer, this distance difference might be offset [27] but there was no significant difference in target size for the different final-location types, $F(2,70)=2.481$, $p=.091$. If anything, the trend was for size to be correlated with distance negatively, not positively. Follow-up tests showed that targets in final likely locations were closer to the centre of the image than targets in both unlikely and implausible locations, p=.046 and p=.013, respectively.

------------- Insert Table 1 about here. --------------------

Data preparation

The data from one participant were excluded because that person did not complete the full session. For the remaining participants, fixations that were shorter than 80ms were removed as outliers. Any saccades that started before the scene to be searched was presented were also removed. The final eye movement dataset comprised approximately 10,000 fixations.

In 3.2% of trials, the target did not change location during the first saccade, usually because the saccade was slower than the velocity trigger used to detect it. These trials were removed.

Search performance

The search-performance measures that we analysed were (1) response accuracy, (2) likelihood of the first saccade landing in the four critical locations (a measure of early scene-gist influence on search), (3) distance of the landing position of the first saccade from the first location of the target (a measure of guidance by both early scene-gist and object), (4) distance of multiple fixations from the final location of the target (a measure of guidance after the target moved), (5) time to the first fixation of the target (a measure of search efficiency, measured from onset of the scene showing the target), and (6) verification time (time from the first fixation of the target to the response). The time measures were included only for trials on which a correct response was made. Medians of all measures (except accuracy) were computed for each condition for each participant. Medians were used rather than means because the number of trials per condition were small, and removing outliers would have made the numbers even smaller. Repeated-measure ANOVAs were run for each of the dependent measures with Scene Preview (Preview or No Preview) and Final Location Type (likely, unlikely or implausible) as factors.

<u>Accuracy</u>. A response was considered accurate if the participant fixated within 2 degrees of the centre of the target when pressing the response key. Accuracy in the different conditions can be seen in Panel A of Figure 4. Mean accuracy was 88.6%. The ANOVA of accuracy rates found no effect of scene preview, location type, or their interaction, $F(1,20) = 1.848$, $p = .189$, $F(2,40) = 1.277$, $p = .290$, and $F(2,40) = 1.217$, $p = .307$, respectively.

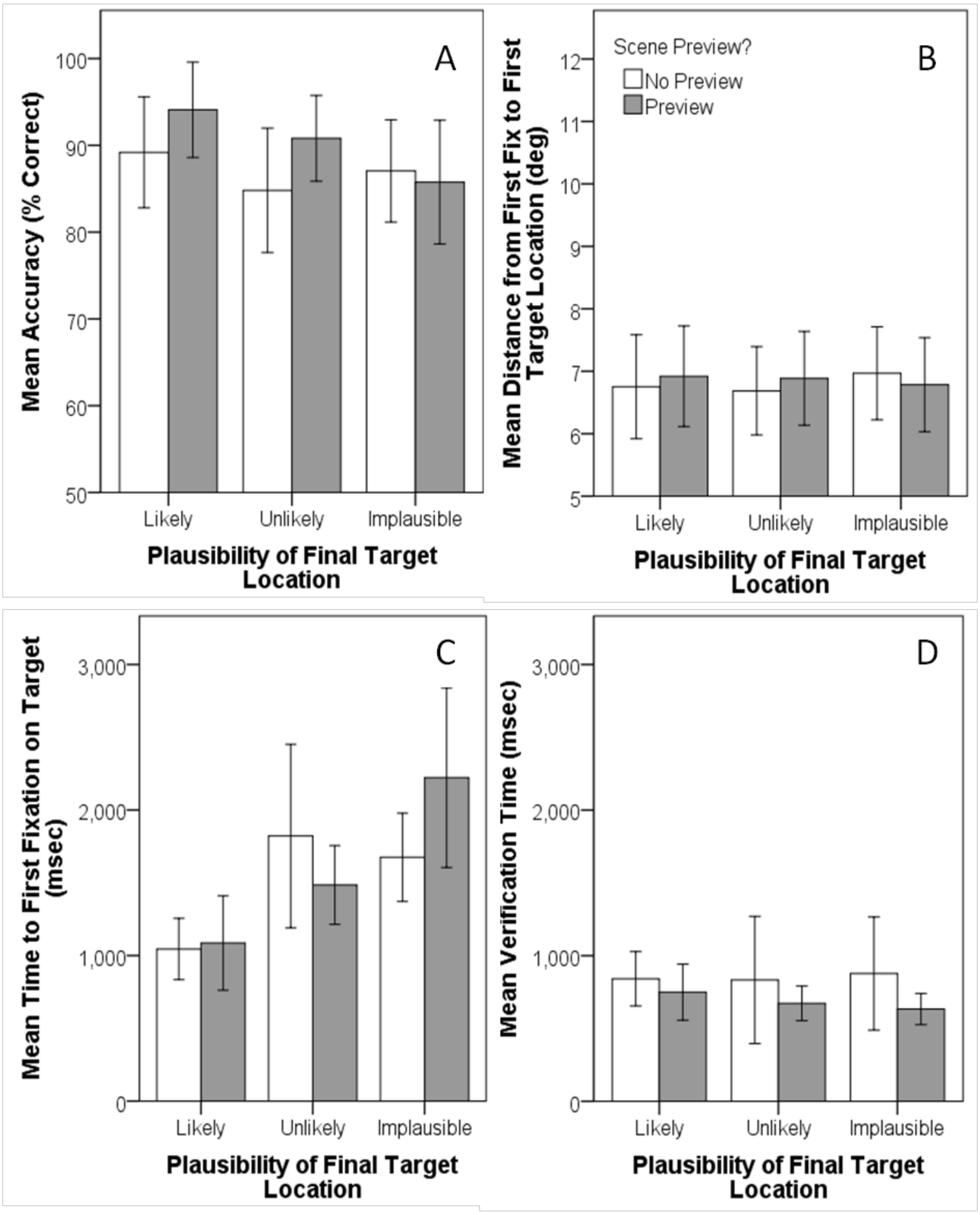


Figure 4: Effect of scene preview and final target location on response measures. A. Accuracy. B. Distance from the first fixation in which the target was at its final location to the centre of the final location. C. Search Time. D. Verification Time. For times and distances, bars are means of each participant's median score, and error bars represent a 95% Confidence Interval.

<u>Likelihood of landing in critical locations.</u> If scene gist influenced where people look for targets, likely locations should be looked at more than unlikely or implausible locations. Of course, if

the target is visible, the evidence of where it actually *is* should also guide search. To identify whether this guidance occurs early in the trial, the landing position of the first saccade was coded according to whether it landed at the first location of the target or any of the three final possible locations of the target. Because the critical regions in the scene were not the only likely, unlikely or implausible locations in the scene, we did not expect most first fixations to land precisely in one of the four critical locations. Accordingly, the results showed that the first saccade most commonly landed away from all four critical locations (on 88.4% trials). The first saccade landed on the initial likely location of the target (which contained the target when the saccade started) on 5.4% trials, the final likely location on 4.5% trials, the unlikely location on 1.3% trials, and the implausible location on 0.2% trials. The difference in frequency between the three final locations was significant by a chi-square test, $\chi(df=2) = 36.17$, $p < .001$, supporting our claim that scene gist had an early effect on guidance of search. It was not possible to use a χ^2 test to compare first-saccade landing positions with and without a scene preview due to insufficient data.

Distance of first fixation from the first target location. If scene gist rapidly influences where people look for targets, the first saccade should land closer to the first target location after a scene preview than after no preview, and the final target location should have no influence on this. Median distances were computed for each participant in each condition, and means across participants are shown in Panel B of Figure 4. A repeated measures ANOVA found no effect of scene preview, location type, or the interaction between preview and location type, $F(1,20) < 1$ for all.

Search Time. Panel C of Figure 4 shows averages across participants' median time to the first target fixation in different conditions. The ANOVA of time to the first target fixation showed that time differed significantly with location type, $F(2,40) = 8.900$, $p = .002$, $\eta^2 = .308$. Pairwise comparisons showed that the time to fixate the target was shorter when the target appeared in the likely location, mean = 1065.8 ms, SD = 89.7 ms, than when it appeared in the unlikely location, mean = 1653.4 ms, SD = 182.6 ms or in the implausible location, mean = 1948.6 ms, SD = 200.7 ms.

There was no difference in time between unlikely and implausible locations. There was no significant effect of scene preview, *F*(1,20) <1, but there was a significant interaction between scene preview and location type, *F*(2,40) = 3.304, p = .047, η^2 = .142. There was no effect of a scene preview on time to the first target fixation in the likely and unlikely locations, *t*(20) = .216, *p* =.831 and *t*(20) = 1.167, *p* = .257, respectively. The first fixation on a target in an implausible location was later when there was a scene preview, mean = 2222.1 msec, SD = 1355.3, than when there was no scene preview, mean = 1675.1, SD = 666.2, *t*(20) = 2.309, *p* = .032.

Verification Time. Panel D of Figure 4 shows means across participants of median verification time. The ANOVA of time from first target fixation to key press response, that is the verification time, showed that it did not differ significantly with location type, *F*(2,40) < 1, by scene preview, *F*(1,20) = 2.931, p = .102, nor by an interaction between scene preview and location type, *F*(2,40) < 1.

Distance of first four fixations from the final target location. To assess whether likelihood of final target location guided eye movements, we analysed for the first four fixations the distance from fixation to the final target location. Median distances were computed for each participant in each condition, and means across participants are shown in Figure 5. A repeated measures ANOVA found significant main effects for fixation number, F(3,60) = 15.249, p < .001, η^2 = .433, and location type, F(2,40) = 27.223, p < .001, η^2 = .576, and a significant interaction between location type fixation number, F(6, 120) = 6.637, p < .001, η^2 = .249. Figure 5 shows clearly that the eyes moved closer to targets across the first four fixations when the target was in a likely or unlikely location, but not when the target was in an implausible location, and follow up tests of the effect of location for each combination of preview and location support this. Preview did not have a main effect, F(1, 20) = 1.301, p = .268, did not interact significantly with location type, F(2, 40) < 1, or with location type and fixation number, F(6, 120) = 1.105, p = .364, but did interact with fixation number alone, F(3,60) = 2.872, p = .044, η^2 = .126. This interaction appears to be due to an anomalous greater distance of fixation 3 with a preview than without a preview, a result for which we have no explanation.

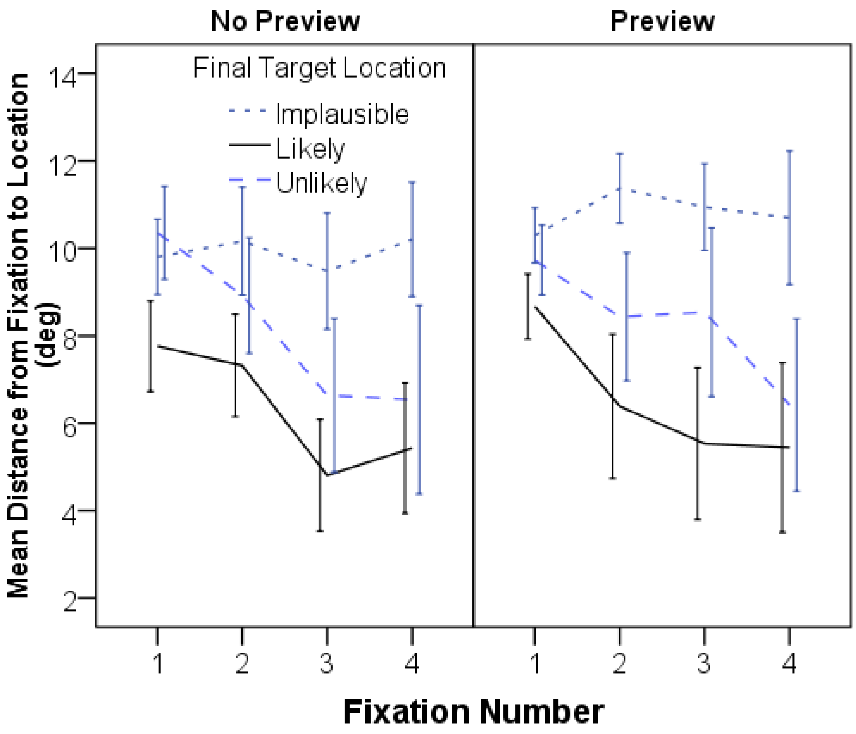


Figure 5. The effect of scene preview and target location on guidance toward the target over the first four fixations after the target moved to the final location. Error bars represent a 95% Confidence Interval.

Additional information concerning the time-course of fixations. Likely locations were predominantly in the lower half of the displays, and implausible locations were predominantly in the upper half of the displays. Consequently, there may have been learning about experiment contingencies in relation to likely locations, which would have added yet another layer of knowledge to the guidance of search [23]. If so, if participants searched likely locations first, they would have searched the lower half of displays before the upper half. Figure 6 shows a difference fixation plot, highlighting which areas were fixated more when searching for likely targets (dark red) and which were fixated more when searching for implausible targets (dark blue). To strengthen the comparison, we removed likely targets above the screen centre and implausible targets below the screen centre. It is apparent that upper regions were searched more when targets were implausible, but that there was less dominance of one location type over the other in the bottom half of the display. In other words, the bottom was almost always searched, and the top was searched only when the target was implausible.

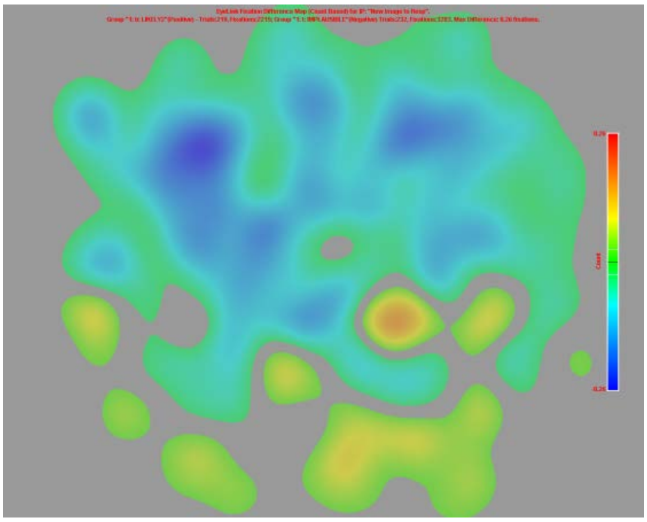


Figure 6. Difference fixation maps comparing search for likely-location targets that appeared in the bottom half and implausible-location targets that appeared in the upper half. A vivid red would indicate more fixations when the target was in a likely location, and a vivid blue indicates more fixations when the target was in an implausible location.

Discussion

The experiment presented here was developed to address whether, when both visible target characteristics and target location likelihood information are available, guidance toward a target is still influenced by location-likelihood. Further, we tested whether location likelihood would have an effect if the location of the target object moved as soon as the participant began a saccade. Earlier research [10] found that a brief scene preview aided guidance toward a target for the first few fixations, more so when the scene was largely masked after the preview than when it was fully visible (see also [5]). When the scene (and target) were fully visible, target characteristics dominated guidance [10].

In the current research, the landing position of the first saccade represents a replication of full scene visibility conditions in [10]. Once again, our findings clearly demonstrate that scene gist had an influence on search behaviour, in that the landing position of the first saccade was more likely to be in one of the likely locations than in the unlikely or implausible locations. However, in contrast with what we found in our earlier work, with a scene preview the first saccade landing position was no closer to the initial (likely) target position than when no preview was presented. The failure to replicate might be due to the fact that participants quickly learned that the target always moved and sometimes ended up in an implausible location. These characteristics of the experiment would likely have demotivated participants to use the preview.

The only effects of a scene preview on inspection behaviour throughout the task was that it slowed search time when the target was in an implausible location. Although the direction of this result is consistent with the use of scene gist information to guide search (in this case misdirecting participants to look in likely locations), it is unclear why a preview would have only this effect. Earlier research found an effect of scene preview on time to find the target but not on verification time [10], so again, this work represents a partial replication of the influence of early scene knowledge on search behaviour. We speculate that the weaker effect of scene preview in this experiment may be because the scene was always fully visible; in the previous experiment, where the scene was often masked during search, there was more motivation to purposefully glean as much information from the preview as possible. It may also be that the frequent movement of targets to unlikely and implausible locations in this experiment lessened participants' inclination to deliberately use scene syntax to guide search. If so, then our observed effect of location likelihood on search behaviour is even more striking. Further research will need to explore the role of conscious versus unconscious use of scene syntax in search.

In our search task, we disrupted guidance by visual target information by moving the target to a new location during the first saccade. The continued effect of scene knowledge after the first saccade was probed by manipulating whether the new target location was implausible, plausible but unlikely, or likely, according to the nature of the scene and the object. The influence of scene-gist information on guidance of search was evident in that both the time to find the target and the progression of fixations toward the target were influenced by the plausibility of the new target location; time to find the target was shorter when the target was in a likely location than when in an unlikely or implausible location, and successive fixations in the implausible condition appeared not to be directed toward the target.

The final likely locations, compared to unlikely and implausible locations, tended to be nearer to the first fixation during which they were visible. The nearer distances no doubt make it more likely that targets would be identified to the point where they could influence guidance, but it is important to understand that the differences in distance probably arose because of the guidance of scene knowledge. Previous work, such as Pomplun, Reingold & Shen’s area activation model, has demonstrated that the planning of the first saccade would be influenced by the distribution of plausible target locations, not just a single plausible location [28]. As such, we would *expect* the landing position of the first saccade to be closer to plausible locations than to implausible locations.

Another feature of the design that undoubtedly contributed to faster search times for targets in final likely locations was the fact that final likely locations were closer to the centre of the display than unlikely and implausible locations [29]. Both of these distance effects contribute to the search time results, but they cannot explain why successive fixations in the trial did not approach targets in implausible locations but did approach targets in likely and unlikely locations. The most parsimonious explanation of the fixation distances is that in the implausible condition, observers were actually searching likely locations for the target.

Thus, when we “toyed with the cat” (participants) by “moving the mouse” (the target), participants were more likely to seek the target in locations that were plausible for the target, according to semantic knowledge, than in locations that were unlikely or implausible. So, in answer to our second question, our findings clearly show that location likelihood does have an effect if the location of the target object moves as soon as the participant initiates a saccade. As expected, location plausibility did not affect the time taken to verify the target, only the time to locate the target and the path the eyes took to reach it.

We view the results of this experiment as a conceptual replication that scene syntax gist information influences search for objects in scenes [6, 8, 30], but, we also find limits to the effect of scene previews, as have been reported by other researchers [31]. Most uniquely, however, by moving the target during search, and thereby disrupting guidance by visible target characteristics, we have shown a lasting effect of scene gist. Moreover, this effect masked what would typically be the predicted stronger influence of object characteristics on search. In sum, scene preview is more powerful than object properties in guiding cat and mouse search.

Additional Information

Acknowledgements

We are grateful to Charlotte Riggs, Oliver Tew, and Alex Muhl-Richardson for help with piloting the study and with data collection.

Ethics

The research presented here was reviewed by the Ethics Committee for the Faculty of Social, Human and Mathematical Sciences at University of Southampton.

Data Accessibility

The programme used to run the study and the data resulting from the study are available through University of Southampton’s ePrints data repository. [A link will be added when we have completed this.] Images used in the study are not publicly available, because we had permission to use them but did not have permission to distribute them.

Authors' Contributions

APH participated in early design decisions, analysed the results, and wrote the first draft of the manuscript. JS collected data, contributed to data analyses, and reviewed the manuscript. HG participated in design decisions throughout the project, programmed the experiment, oversaw data collection, and revised the manuscript. SPL and VB participated in design decisions throughout, and revised the manuscript. All authors approved the final version of the manuscript.

Competing Interests

We have no competing interests.

Funding

J. D. S. was supported by the CAPES Foundation.

References

1. Hollingworth, A. (2009). Two forms of scene memory guide visual search: Memory for scene context and memory for the binding of target object to scene location. *Visual Cognition*, *17*(1-2), 273–291. (doi:10.1080/13506280802193367)
2. Wolfe, J. M. (1994). Guided search 2.0 a revised model of visual search. *Psychonomic Bulletin & Review*, *1*(2), 202-238.

3. Kanan C., Tong M. H., Zhang L., & Cottrell G. W. (2009). SUN: Top-down saliency using natural statistics. Visual Cognition, 17, 979–1003.
4. Williams, L. G. (1966). The effect of target specification on objects fixated during visual search. *Perception & Psychophysics, 1*(9), 315-318.
5. Võ, M. L. H., & Henderson, J. M. (2010). The time course of initial scene processing for eye movement guidance in natural scene search. *Journal of Vision*, *10*(3), 1–13. (doi:10.1167/10.3.14)
6. 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. (doi:10.3758/s13423-011-0107-8
7. Castelhano, M. S., & Henderson, J. M. (2007). Initial scene representations facilitate eye movement guidance in visual search. *Journal of Experimental Psychology: Human Perception and Performance*, *33*(4), 753–763. (doi:10.1037/0096-1523.33.4.753)
8. 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), 24–24. (doi:10.1167/9.3.24
9. Eckstein, M. P., Drescher, B. A., & Shimozaki, S. S. (2006). Attentional Cues in Real Scenes, Saccadic Targeting, and Bayesian Priors. *Psychological Science*, *17*(11), 973–980. (doi:10.1111/j.1467-9280.2006.01815.x
10. Hillstrom, A. P., Scholey, H., Liversedge, S. P., & Benson, V. (2012). The effect of the first glimpse at a scene on eye movements during search. *Psychonomic Bulletin & Review*, *19*(2), 204–210. (doi:10.3758/s13423-011-0205-7
11. Loftus, G. R., & Mackworth, N. H. (1978). Cognitive determinants of fixation location during picture viewing. *Journal of Experimental Psychology: Human Perception and Performance*, *4*(4), 565–572.
12. Henderson, J. M., Weeks, P. A., & Hollingworth, A. (1999). The effects of semantic consistency in eye movement during complex scene viewing. *Journal of Experimental Psychology: Human Perception and Performance*, *25*(1), 210–228.
13. De Graef, P., Christiaens, D., & d'Ydewalle, G. (1990). Perceptual effects of scene context on object identification. *Psychological Research*, *52*(4), 317-329.
14. Gareze, L., & Findlay, J. M. (2007). Absence of scene context effects in object detection and eye gaze capture. In R. van Gompel, M. Fischer, W. Murray, & R. W. Hill (Eds.), *Eye Movements: A Window on Mind and Brain* (pp. 537–562). Amsterdam: Elsevier.
15. Bonitz, V. S., & Gordon, R. D. (2008). Attention to smoking-related and incongruous objects during scene viewing. *Acta Psychologica*, *129*(2), 255-263
16. 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*(6), 1742-1753.
17. Biederman, I., Mezzanotte, R. J., & Rabinowitz, J. C. (1982). Scene perception: Detecting and judging objects undergoing relational violations.*Cognitive Psychology*, *14*(2), 143-177.
18. Neider, M. B., & Zelinsky, G. J. (2006). Scene context guides eye movements during visual search. *Vision Research*, *46*(5), 614–621. (doi:10.1016/j.visres.2005.08.025)
19. Torralba, A., Oliva, A., Castelhano, M. S., & Henderson, J. M. (2006). Contextual guidance of eye movements and attention in real-world scenes: The role of global features in object search. *Psychological Review*, *113*(4), 766–786. (doi:10.1037/0033-295X.113.4.766)
20. Chen, X., & Zelinsky, G. J. (2006). Real-world visual search is dominated by top-down guidance. *Vision Research*, *46*(24), 4118–4133. (doi:10.1016/j.visres.2006.08.008)
21. Glaholt, M. G., Wu, M.-C., & Reingold, E. M. (2010). Evidence for top-down control of eye movements during visual decision making. *Journal of Vision*, *10*(5), 15–15. (doi:10.1167/10.5.15 doi:10.1037/0278-7393.6.2.174
22. 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), 5–5. (doi:10.1167/11.5.5

23. Stainer, M. J., Scott-Brown, K. C., & Tatler, B. W. (2013). Behavioral biases when viewing multiplexed scenes: scene structure and frames of reference for inspection. *Frontiers in Psychology*, *4*. (doi:10.3389/fpsyg.2013.00624
24. Zelinsky, G. J., & Schmidt, J. (2009). An effect of referential scene constraint on search implies scene segmentation. *Visual Cognition*, *17*(6-7), 1004–1028. (doi:10.1080/13506280902764315
25. Võ, M. L. H., & Wolfe, J. M. (2013). Differential electrophysiological signatures of semantic and syntactic scene processing. *Psychological science*, *24*(9), 1816-1823.
26. Walther, D., & Koch, C. (2006). Modeling attention to salient proto-objects. *Neural Networks* **19**, 1395-1407.
27. Anstis, S. M. (1974). A chart demonstrating variations in acuity with retinal position. *Vision Research*, *14*(7), 589-592.
28. Pomplun, M., Reingold, E. M., & Shen, J. (2003). Area activation: A computational model of saccadic selectivity in visual search. *Cognitive Science*, *27*(2), 299-312.
29. Tatler, B. W. (2007). The central fixation bias in scene viewing: Selecting an optimal viewing position independently of motor biases and image feature distributions. *Journal of Vision*, *7*(14), 4.
30.Brockmole, J. R., Castelhano, M. S., & Henderson, J. M. (2006). Contextual cueing in naturalistic scenes: Global and local contexts. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *32*(4), 699–706. (doi:10.1037/0278-7393.32.4.699)
31. Wu, C.-C., Wick, F. A., & Pomplun, M. (2014). Guidance of visual attention by semantic information in real-world scenes. *Frontiers in Psychology*, *5*. (doi:10.3389/fpsyg.2014.00054

Table 1

Descriptive details for the three final location conditions.

	Final Target Location		
	Likely	Unlikely	Implausible
Mean target salience	.09 (.33)	.01 (.03)	.01 (.05)
Mean area of target (squared °)	1.54 (0.94)	1.44 (0.71)	1.43 (0.84)
Mean distance from first fixation (after target moved) to final target location (°)	8.2 (.272)	10.0 (.377)	10.1 (.377)
Mean distance from centre of image to final target location (°)	6.75 (3.22)	8.49 (2.73)	8.70 (1.89)

Note. Standard deviations appear in parentheses.

Figure and table captions

Figure 1. A representative scene with four different locations for the same target (mug). Circles have been added around targets to aid the reader. For copyright reasons, this image is not one of the scenes used in the study.

Figure 2. Spatial distributions of the initial target locations and, for the three location-plausibility conditions, the final target locations.

Figure 3. Trial sequence.

Figure 4: Effect of scene preview and final target location on response measures. A. Accuracy. B. Distance from the first fixation in which the target was at its final location to the centre of the final location. C. Search Time. D. Verification Time. For times and distances, bars are means of each participant's median score, and error bars represent a 95% Confidence Interval.

Figure 5. The effect of scene preview and target location on guidance toward the target over the first four fixations after the target moved to the final location. Error bars represent a 95% Confidence Interval.

Figure 6. Difference fixation maps comparing search for likely-location targets that appeared in the bottom half and implausible-location targets that appeared in the upper half. A vivid red would indicate more fixations when the target was in a likely location, and a vivid blue indicates more fixations when the target was in an implausible location.

Table 1. Descriptive details for the three final location conditions.