Journal of Experimental Psychology:
Human Perception and Performance
2016, Vol. 42, No. 10, 1477–1481

© 2016 American Psychological Association
0096-1523/16/$12.00 http://dx.doi.org/10.1037/xhp0000264

OBSERVATION

You Think You Know Where You Looked? You Better Look Again

Melissa L.-H. Võ
Goethe University Frankfurt

Avigael M. Aizenman and Jeremy M. Wolfe
Brigham and Women's Hospital/Harvard Medical School

People are surprisingly bad at knowing where they have looked in a scene. We tested participants' ability to recall their own eye movements in 2 experiments using natural or artificial scenes. In each experiment, participants performed a change-detection (Exp.1) or search (Exp.2) task. On 25% of trials, after 3 seconds of viewing the scene, participants were asked to indicate where they thought they had just fixated. They responded by making mouse clicks on 12 locations in the unchanged scene. After 135 trials, observers saw 10 new scenes and were asked to put 12 clicks where they thought someone else would have looked. Although observers located their own fixations more successfully than a random model, their performance was no better than when they were guessing someone else's fixations. Performance with artificial scenes was worse, though judging one's own fixations was slightly superior. Even after repeating the fixation-location task on 30 scenes immediately after scene viewing, performance was far from the prediction of an ideal observer. Memory for our own fixation locations appears to add next to nothing beyond what common sense tells us about the likely fixations of others. These results have important implications for socially important visual search tasks. For example, a radiologist might think he has looked at "everything" in an image, but eye tracking data suggest that this is not so. Such shortcomings might be avoided by providing observers with better insights of where they have looked.

Keywords: eye movements, fixation recall, memory, change detection, visual search

You might stop looking for your keys on your messy desk out of the conviction that you have "looked everywhere." But how much do we really know about where we have looked? We know that our eyes do not always go where we want them to go (Bridgeman & Stark, 1991; Theeuwes, Kramer, Hahn, & Irwin, 1998), and people sometimes report eye movements they actually never made (Marti, Bayet, & Dehaene, 2015). Usually we do not need or care to know where we looked. However, in some cases, like a radiologist who thinks he has 'looked at the whole image' before moving on to the next image (Kundel, Nodine, & Carmody, 1978), it might be useful to know how accurately we can monitor the last few fixations. Shedding more light on these questions is the main aim of this study.

Since 2013, two interesting papers have addressed this question with different methods and different stimuli. Foulsham and Kingstone (2013) asked participants to memorize photographs of real-world indoor environments. At the end of the experiment, they presented the observers with either their own fixations, random fixations, or the fixations of others. Observers were quite good at distinguishing their fixations from random fixations; probably, the authors suggest, because the distributions of human fixations are not random and observers know that they would not typically fixate, for example, in the corners of an image. They were markedly poorer at discriminating between their fixations and those of another observer. Here the authors speculate that observers were performing above chance because a few individual fixations were informative ("I don't remember seeing that pillow and it is marked. Must be someone else's fixations.").

In contrast, Marti et al. (2015) asked observers to perform a visual search task in simple displays of Ts and Ls and to mark the locations of their own fixations after each search. They also tested observers on the search task alone. Their observers could report on their eye movements, albeit imperfectly. For instance, they clearly knew that they made more fixations when search took longer.

These studies show that observers know something about their eye movements. However, it is not clear how much they know about their specific scan paths, as opposed to having somewhat extraneous information about the overall length of a trial or the standard distribution of human fixations in a photograph. In the present study, we wished to gain more clarity about the degree to which observers have memory for their own fixations in a complex scene. Our practical interest in this question comes from real-world expert search tasks like those in radiology. A radiologist wants to move on to the next image only after she has looked at "every-

Melissa L.-H. Võ, Scene Grammar Lab, Department of Cognitive Psychology, Goethe University Frankfurt; Avigael M. Aizenman and Jeremy M. Wolfe, Visual Attention Lab, Brigham and Women's Hospital/Harvard Medical School.

This work was supported by National Institutes of Health Grant EY017001 to Jeremy M. Wolfe as well as Deutsche Forschungsgemeinschaft (DFG) Grant VO 1683/2-1 to Melissa L.-H. Võ.

Correspondence concerning this article should be addressed to Melissa L.-H. Võ, Scene Grammar Lab, Department of Cognitive Psychology, Goethe University Frankfurt, Theodor-W.-Adorno-Platz 6, 60323 Frankfurt am Main, Germany. E-mail: mlvo@psych.uni-frankfurt.de

thing" relevant in the current image. This implies knowledge of where they have focused the eyes. Following Foulsham and Kingstone (2013), we had observers search in complex scenes. Following Marti et al. (2015), we had them attempt to locate their own fixations. The Marti et al. observers were probably encoding fixations during the task. In any case, response times were longer in blocks when observers knew that they would be asked to mark fixations on every trial. To reduce the inclination to deliberately encode fixations, we only asked about fixations on a minority of trials. We compared observers' guesses about their own fixations with their guesses about someone else's. There is little or no evidence that observers have access to their own history of fixations. As Foulsham and Kingstone (2013) suggested, they merely have access to intuitions about where a sensible person would fixate.

Method

The study consisted of two experiments. In Experiment 1, observers viewed images in preparation of a change detection task, whereas in Experiment 2 a different group of observers previewed visual displays in preparation for a search task. We chose these tasks because they closely mimic socially highly relevant types of visual inspection. Radiologists, for instance, regularly scrutinize medical images to detect changes between different scans and are usually searching for signs of cancer.

Participants

Eight observers participated in Experiment 1 (mean age = 29, *SD* = 10, 3 female), and another 8 in Experiment 2 (mean age = 30, *SD* = 12, 5 female). All observers passed the Ishihara test for color blindness (Ishihara, 1987) and reported normal to corrected-to-normal vision. The Partners Health care Corporation Institutional Review Board approved all experimental procedures and observers gave informed consent and were compensated for their time.

Apparatus

Eye movements were recorded with an EyeLink1000 desktop mount system (SR Research, Canada) at a sampling rate of 1000 Hz. Viewing was binocular, but only the position of the right eye was tracked. Saccades and fixations were extracted from raw gaze data during recording, by the EyeLink parser. Velocity and acceleration thresholds were set to the EyeLink default values of 30 degrees/s and 8000 degrees/s^2, respectively. Stimulus presentation and response recording was controlled by MATLAB using Psychophysics Toolbox (Brainard, 1997; Pelli, 1997).

Stimuli

Photographs. In Experiment 1, colored images of real-world indoor scenes (size: 1024 × 768 pixels) were used for a change detection task (Figure 1A). There were two instances of each image. One object in the scene could change either its identity or its location from the first to the second presentation of an image in the change blindness test. Images were not created by post hoc insertion of objects into scenes. Rather, both instances of an image were photographs of a scene that was physically modified between shots to ensure realistic lighting conditions and minimize Photoshop editing.

Waldo scenes. For Experiment 2, "Where's Waldo" images (Figure 1C) were acquired by scanning the 6 books that are part of the "Where's Waldo Wow Collection" (Martin, 2012). Each book consists of 12 different "Where's Waldo" searches, where a search spans two large pages. Each search is richly detailed and covers a large area. Because of this, four-color scans of each of these images were made at 1200 dpi using a Brother DCP-8065DN scanner. Each scan was resized in Photoshop to be 1024 × 768 pixels large. Images were displayed on a 19-in. computer screen for both experiments. Resolution was 1024 × 768 pixel (refresh rates Exp.1: 100Hz, Exp.2: 65Hz) and the entire image subtended 37° of visual angle horizontally and 30° vertically at a viewing distance of 65 cm.

Procedure

Experiment 1. Each experiment started with a nine-point calibration procedure (see Figure 1A). Calibration was only deemed successful when average fixation error was under 0.5° for all validation points and when no point produced an error larger than 1.0°. After successful calibration, observers viewed a scene for 3 seconds in preparation for a *change detection task* followed by another 3-s presentation of the scene separated by a 500-ms blank screen. Changes occurred in 50% of all change detection trials. Of these changes, 50% were identity and 50% were location changes. Participants were told to indicate whether a change had occurred after the second presentation of the image by pressing one of two keys.

Experiment 2. After a nine-point calibration procedure, observers were shown a 3-s preview of a Waldo scene in preparation for a *search task* (see Figure 1C). Only after the preview, a word was shown for 500 ms indicating what object they should search for (Waldo was never actually the designated target). The same Waldo scene would reappear and observers had another 3 seconds to find the target. After the second presentation of the Waldo scene, observers were asked to respond whether the target was absent or present by key-press.

In both Experiments 1 and 2, there were 135 trials. Crucially, on 30 of these trials (i.e., in about 25%), we omitted the second presentation of the image needed to perform the change detection (Exp.1) or search (Exp.2) task. Instead, immediately after the initial 3-s scene viewing, participants were asked to mark 12 locations where they thought *they* had just looked in the scene. At the end of the experiment, observers were shown 10 new scenes and were asked to mark 12 locations where they thought *someone else* would look. Thus, in total every participant performed 145 trials.

Data Analysis

As a measure of fixation memory, we calculated the overlap between fixations and the 12 clicks by placing a circular region around each actual fixation and each click indicating a remembered fixation. The critical measure was the degree of overlap between those two sets of circles. This overlap was calculated for a range of different radius. Obviously, if the clicks and fixations were in the same location, overlap would be 100%

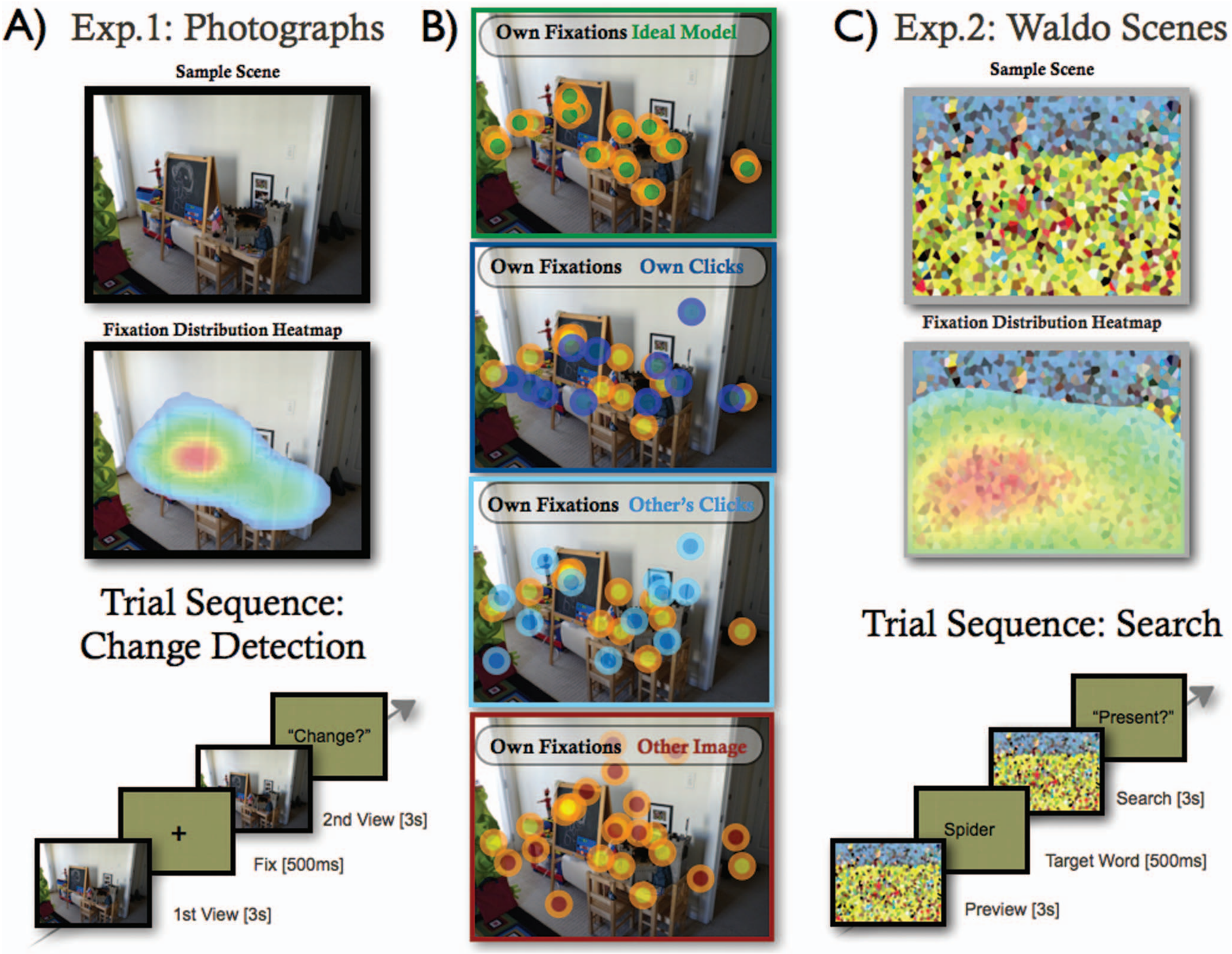

Figure 1. (A) Sample photograph of indoor scene used in Experiment 1 (upper). Heat-map of fixation distribution on indoor scenes show constrained viewing of these types of images during a change detection task (middle). Trial sequence of Experiment 1 using photographs of indoor real-world scenes in a change detection task (lower). (B) Sample scene with "Own Fixations" overlaid together with either "Ideal Model" clicks, the observer's "Own Clicks," "Other's Clicks" on the same image, or clicks made by the observer on an "Other Image." (C) Sample "Where's Waldo" scenes used in Experiment 2 (upper). Heatmap of fixation distribution on Waldo scenes show widespread fixation distributions of these types of images while previewing scenes for a search task (middle). Trial sequence of Experiment 2 using "Where's Waldo" scenes in a visual search task (lower). Note: The images representing the Waldo scenes in this figure are not actual Waldo scenes because the publisher no longer grants permission for reproduction. See the online article for the color version of this figure.

and, obviously, if the radii of the circles are large enough, overlap will approach 100%. In practice, the click and fixation do not fall on the same pixel. Thus, overlap grows as the radius grows from 0.5°, 1°, 2°, 3°, 4°, to 5° visual angle. Small windows produce small overlap while large produce large overlap. Overlap as a function of radius is shown in Figure 2A. Participants made an average of 11.3 fixations during 3 seconds of scene viewing. As shown in Figure 1b, we calculated the overlap of these fixations with four sets of other locations: (a) the clicks the same Participant A made immediately after viewing (dark blue), (b) the clicks Participant B made when guessing Participant A's fixations on the same image (light blue), (c) the clicks Participant A made on a different image within the first 135 trials (lower bound in red), and (d) an ideal observer model that placed a "click" at a location with a 1° jitter around the actual location of each fixation. Thus, the ideal observer had perfect memory for fixations, degraded by modest noise.

Results

As can be seen in Figure 2A, performance in estimating both one's own (dark blue lines) and other's (light blue lines) eye movements was better than chance (red lines), but also far from an ideal observer model (green lines). Interestingly, estimates of

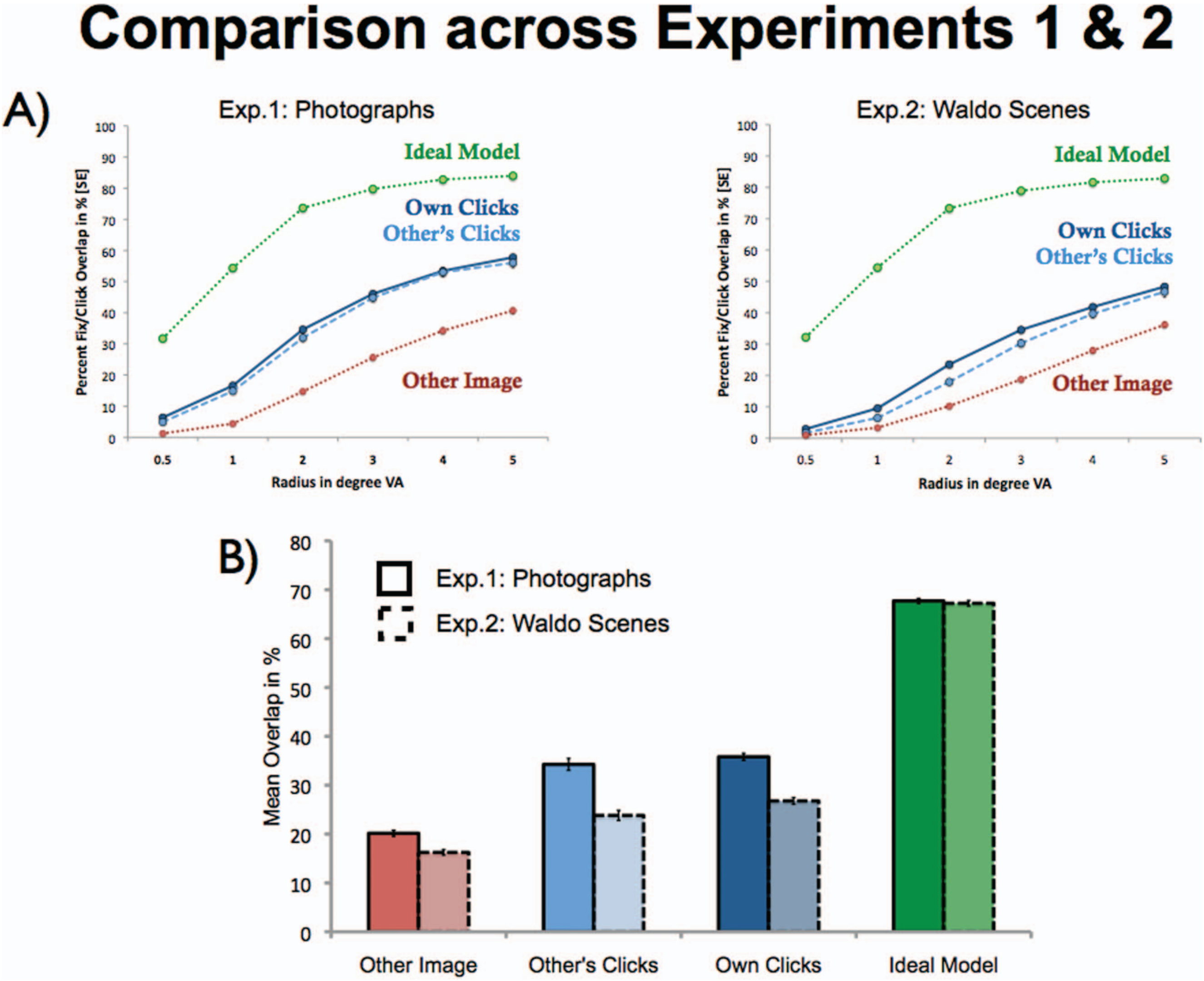

Figure 2. (A) Results of Experiment 1 (left) and Experiment 2 (right) showing the overlap between observers' own fixations with the "Ideal Model" clicks, the observer's "Own Clicks," "Other's Clicks" on the same image, and clicks made by the observer on an "Other Image." Percent overlap increases a function of circle radii (0.5, 1, 2, 3, 4, 5 degree visual angle), but shows no difference between "Own" and "Other's" clicks in Experiment 1 (left) and only a small difference in Experiment 2 (right). (B) Comparison across Experiments 1 and 2 with overlap averaged across all radii as a function of overlap conditions ("Other Image" as lower bound in red [left bars], "Own Clicks" in dark blue [2nd bars], "Other's Clicks" in light blue [3rd bars], and "Ideal Model" as upper bound in green [right bars]) and experiments (Exp.1: solid lines, Exp.2: dashed lines). Performance for search on Waldo scenes in Exp.2 was overall decreased in comparison to performance on photograph images in Exp. 1. Error bars depict ± 1 *SE*. See the online article for the color version of this figure.

one's own fixations were barely distinguishable from guessing the fixations of someone else.

To statistically compare fixation memory performance across conditions, we collapsed overlap values across all circle radii and submitted these averaged values to paired *t* tests. Figure 2B shows that, in *Experiment 1* (solid lines), performance for one's own fixations (M = 36%) was no different from one's guesses about the fixations of a hypothetical observer (M = 34%, $t(7)<1$). Performance was better than chance (M = 20%, $t(7) = 11.01$, $p < .001$), but much worse than our ideal observer (M = 68%, $t(7) = 26.01$, $p < .001$). In *Experiment 2* (Figure 2B, dashed lines), participants were only marginally better at locating their own (M = 27%) as compared to someone else's fixations (M = 24%, $t(7) = 2.09$, $p = .07$), and again far worse than the ideal observer prediction, (M = 67%, $t(7) = 39.31$, $p < .001$). Comparing the Waldo and photograph versions, the overlap of real with proposed fixations (dark blue bar) was significantly worse in Waldo scenes (M = 27%) compared with performance in photographs (M = 36%, $t(14) = 4.88$, $p < .001$). Overlap was also lower between real fixations and guesses of another's fixations (light blue bar), showing that participants had less accurate intuitions about where a reasonable observer would fixate in a Waldo scene. Ideal model performance generated by a 1° jitter around every fixation (green bar) had to be almost identical in both experiments.

Discussion

These results indicate that your 'memory' for where you fixated in a scene is not significantly better than your guess about where someone else would have fixated. The comparison with fixations drawn from another image shows that participants are not simply guessing. The comparison with our "ideal observer" shows that the similarity between 'memory' and estimates of another's fixation is not attributable to a ceiling effect. In principle, one would expect that people would do much better at this task than they did.

These results clarify the interpretation of previous work in this area. Foulsham and Kingstone (2013) showed that observers were better than chance at recognizing their own fixations. Our results indicate that they were correct when they speculated that this performance could have been based on small nuggets of memory ("I remember looking at that chair"). The eye movements of two different observers looking at a real-world scene are likely to be highly correlated simply because the semantics of the scene are very constraining (in Figure 1A everyone will look at the blackboard). Individual idiosyncrasies will produce differences that could be weakly detected in the Foulsham and Kingstone (2013) study. In our data, the guesses about another's fixations must be based on the semantic constraints alone. Memory about one's own fixations seems to add nothing to those semantically driven guesses.

It is possible that our change detection task of Experiment 1 might have directed observers' attention to movable objects. This added semantic constraint could have increased guessing performance, reducing the chance that we would find superior estimates of one's own versus someone else's eye movements in that experiment. We therefore replicated the findings from the first experiment with a search task using Where's Waldo scenes that have many objects to fixate and a less eye movement–constraining structure. An analysis regarding the overlap of fixations made by two different participants on the same images showed that this overlap was modest for real-world scenes (Exp.1: 36%), and was further reduced for the Waldo scenes (Exp.2: 22%). When we minimized structural constraints that could support successful guessing, we found marginally better performance in estimating one's own fixation locations compared with someone else's. Nevertheless, the results of Experiment 2 show only very weak evidence for any privileged access to one's own history of fixation.

Recall that Marti et al. (2015) asked observers to reproduce their history of fixations on every trial in displays with no semantic structure (Ts and Ls). This task raised the possibility that above chance performance was due to a deliberate effort to memorize at least a few fixations. Visual search RTs were somewhat longer on blocks of trials where observers were asked to reproduce their fixations, supporting the idea that those observers were putting some effort into the fixation memorization task. In contrast, our design provided feedback on every change detection/search trial in an effort to focus attention on that task. We interleaved fixation queries on only 25% of all trials. In this way, we hoped to minimize continued self-monitoring of eye-movements. As evidence that participants were focusing on the detection/search tasks, performance on those tasks did not decline over the course of the experiment. In addition, the participants' fixation estimates did not substantially increase from early to late queries (Exp.1: 36% vs. 37%, Exp.2: 25% vs. 26%, both ts$<$1). As in Marti et al. (2015), our observers were better than chance but, as noted, the similarity to their guesses about the fixations of another, hypothetical viewer, raises the possibility that their performance represents an intelligent guess about where anyone would fixate in a meaningful scenes.

The ability to know where you have looked is probably of very modest value in daily life, though its failure may explain how you can think you have scrutinized your desk and still cannot find that flash drive. In applied settings like medical image perception, the consequences might be more significant. A radiologist, scrolling through a stack of lung CT images wants to look at "everything" important on every slice. Eye tracking data reveals that substantial areas may go unexamined (Drew et al., 2013), suggesting that even expertise and training do not guarantee memory for where one has looked. Given this poor memory, it might be useful to track the eye movements of expert observers and then provide them with feedback about where their eyes have been. Until then, when in doubt you better look again.

References

Brainard, D. H. (1997). The psychophysics toolbox. *Spatial Vision, 10,* 433–436. http://dx.doi.org/10.1163/156856897X00357

Bridgeman, B., & Stark, L. (1991). Ocular proprioception and efference copy in registering visual direction. *Vision Research, 31,* 1903–1913. http://dx.doi.org/10.1016/0042-6989(91)90185-8

Drew, T., Võ, M. L-H., Olwal, A., Jacobson, F., Seltzer, S. E., & Wolfe, J. M. (2013). Scanners and drillers: Characterizing expert visual search through volumetric images. *Journal of Vision, 13,* 3. http://dx.doi.org/10.1167/13.10.3

Foulsham, T., & Kingstone, A. (2013). Where have eye been? Observers can recognise their own fixations. *Perception, 42,* 1085–1089. http://dx.doi.org/10.1068/p7562

Ishihara, I. (1987). *Ishihara's tests for color-blindness* (Concise ed.). Tokyo, Japan: Kanehara & Co.

Kundel, H. L., Nodine, C. F., & Carmody, D. (1978). Visual scanning, pattern recognition and decision-making in pulmonary nodule detection. *Investigative Radiology, 13,* 175–181. http://dx.doi.org/10.1097/00004424-197805000-00001

Marti, S., Bayet, L., & Dehaene, S. (2015). Subjective report of eye fixations during serial search. *Consciousness and Cognition, 33,* 1–15. http://dx.doi.org/10.1016/j.concog.2014.11.007

Martin, H. (2012). *Where's Waldo? The wow collection: Six amazing books and a puzzle*. Somerville, MA: Candlewick Press.

Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: Transforming numbers into movies. *Spatial Vision, 10,* 437–442. http://dx.doi.org/10.1163/156856897X00366

Theeuwes, J., Kramer, A. F., Hahn, S., & Irwin, D. E. (1998). Our eyes do not always go where we want them to go: Capture of the eyes by new objects. *Psychological Science: A Journal of the American Psychological Society/APS, 9,* 379–385. http://doi.org/10.1111/1467-9280.00071

Received March 14, 2016
Revision received June 15, 2016
Accepted June 16, 2016 ■