



DIGITAL ACCESS TO  
SCHOLARSHIP AT HARVARD  
DASH.HARVARD.EDU



HARVARD LIBRARY  
Office for Scholarly Communication

# Even if I showed you where you looked, remembering where you just looked is hard

The Harvard community has made this  
article openly available. [Please share](#) how  
this access benefits you. Your story matters

|                   |  |
|-------------------|--|
| Citation          | Kok, Ellen M., Avi M. Aizenman, Melissa L.-H. Vö, and Jeremy M. Wolfe. 2017. "Even if I showed you where you looked, remembering where you just looked is hard." <i>Journal of Vision</i> 17 (12): 2. doi:10.1167/17.12.2. <a href="http://dx.doi.org/10.1167/17.12.2">http://dx.doi.org/10.1167/17.12.2</a> .                                   |
| Published Version | <a href="https://doi.org/10.1167/17.12.2">doi:10.1167/17.12.2</a>  |
| Citable link      | <a href="http://nrs.harvard.edu/urn-3:HUL.InstRepos:34492265">http://nrs.harvard.edu/urn-3:HUL.InstRepos:34492265</a>  |
| Terms of Use      | This article was downloaded from Harvard University's DASH repository, and is made available under the terms and conditions applicable to Other Posted Material, as set forth at <a href="http://nrs.harvard.edu/urn-3:HUL.InstRepos:dash.current.terms-of-use#LAA">http://nrs.harvard.edu/urn-3:HUL.InstRepos:dash.current.terms-of-use#LAA</a> |

# Even if I showed you where you looked, remembering where you just looked is hard

**Ellen M. Kok**

School of Health Professions Education,  
Maastricht University, Maastricht, the Netherlands  
Brigham and Women's Hospital/Harvard Medical School,  
Boston, MA, USA



**Avi M. Aizenman**

Brigham and Women's Hospital/Harvard Medical School,  
Boston, MA, USA  
University of California, Berkeley, Berkeley, CA, USA



**Melissa L.-H. Võ**

Scene Grammar Lab, Goethe University, Frankfurt,  
Frankfurt, Germany



**Jeremy M. Wolfe**

Brigham and Women's Hospital/Harvard Medical School,  
Boston, MA, USA



People know surprisingly little about their own visual behavior, which can be problematic when learning or executing complex visual tasks such as search of medical images. We investigated whether providing observers with online information about their eye position during search would help them recall their own fixations immediately afterwards. Seventeen observers searched for various objects in “Where’s Waldo” images for 3 s. On two-thirds of trials, observers made target present/absent responses. On the other third (critical trials), they were asked to click twelve locations in the scene where they thought they had just fixated. On half of the trials, a gaze-contingent window showed observers their current eye position as a 7.5° diameter “spotlight.” The spotlight “illuminated” everything fixated, while the rest of the display was still visible but dimmer. Performance was quantified as the overlap of circles centered on the actual fixations and centered on the reported fixations. Replicating prior work, this overlap was quite low (26%), far from ceiling (66%) and quite close to chance performance (21%). Performance was only slightly better in the spotlight condition (28%,  $p = 0.03$ ). Giving observers information about their fixation locations by dimming the periphery improved memory for those fixations modestly, at best.

as to where we have looked. In most situations, this lack of introspective access is not problematic. You do not really want to fill your memory with the exact path, taken by your eyes, as you looked for the perfect apple at the market. However, when learning or executing a complex visual task, knowledge about your own eye movements might improve performance and foster learning (I think I looked at the whole image. Did I actually do so?). In the work reported here, observers were explicitly shown their eye position on-line, as they viewed an image. We asked if this improved their ability to subsequently recall their fixations. To anticipate the results, this online gaze visualization had very little impact on introspective ability during a visual search task.

## Why might we need to know about our own eye movements?

Literature on self-regulation in (nonvisual) tasks stresses the importance of monitoring your behavior in order to regulate your learning (e.g., Greene & Azevedo, 2010; Nelson & Nahrens, 1990; Winne & Hadwin, 1998; Zimmerman & Schunk, 2001). Introspection or monitoring provides input for the regulation of (learning) processes (Nelson & Nahrens, 1990). Nelson and Nahrens (1990) stress that, while introspection is known to be imperfect, it is still important to

## Introduction

We are constantly moving our eyes to take in relevant information, but we have limited introspection

Citation: Kok, E. M., Aizenman, A. M., Võ, M. L.-H., & Wolfe, J. M. (2017). Even if I showed you where you looked, remembering where you just looked is hard. *Journal of Vision*, 17(12):2, 1–11, doi:10.1167/17.12.2.



understand introspective processes in order to understand what information is used to regulate learning and behavior. For instance, in tasks such as learning key-term definitions, it has been found that suboptimal monitoring of learning behavior impedes optimal regulation and that interventions to improve monitoring promote learning (Dunlosky & Rawson, 2012). Poor introspective abilities could have a negative impact on performance in visual tasks as well. In radiology, for example, the proportion of lung tissue searched (in a CT-scan) was found to correlate with the number of abnormalities found (Rubin et al., 2014). Drew and colleagues (Drew et al., 2013) estimated that radiologists covered on average only 69% of the lungs in a search for lung nodules. Even if this was the “right” amount of the lung to examine, if the radiologist does not know which 69% has been searched, it may be hard to decide whether the search has been successfully completed.

### What do we know about our own viewing behavior?

Several studies have investigated what people know about their own viewing behavior. A basic question asks if people are able to distinguish their own fixation patterns from somebody else’s fixation patterns. Foulsham and Kingstone (2013) found that participants are relatively good at distinguishing their own fixation locations on a scene from random locations or locations fixated in a different image. Participants were less able to tell their own fixations from someone else’s fixations on the same stimulus, but participants’ discrimination accuracy was still modestly above chance level (approximately 55% correct against 50% chance). Similar results were found by Van Wermeskerken, Litchfield, and Van Gog (2017). They found that participants could only discriminate their own eye movements from someone else’s eye movements when a dynamic gaze visualization (i.e., a movie of somebody’s gaze locations on the stimulus) was used and not when a static visualization was used. Further, Clarke and colleagues found that participants could only discriminate their own gaze from another person if this other person was searching for another object (Clarke, Mahon, Irvine, & Hunt, 2016).

At a more detailed level, when asked to recreate their own viewing behavior, what do people know about their previous fixation locations? Clarke et al. (2016) asked participants to view scenes for three seconds each and then to name objects from the scene. Participants were then prompted with objects and asked whether they had fixated that object. Though participants reported having looked at objects that were not fixated and vice versa, their performance was moderately

above chance at this task. Marti, Bayet, and Dehaene (2015) as well as Vö, Aizenman, and Wolfe (2016) both asked participants to click on locations that they thought they had fixated. Marti and colleagues (2015) used a visual conjunction search task, and found that participants reported a subset of fixations: The average number of real fixations was larger than the average number of reported fixations, and participants reported some fixations that could not be matched to actual fixations. Vö et al. (2016) asked participants to report on fixation locations in artificial and natural scenes. They found that participants’ memory for where they looked in a scene was not significantly better than the participants’ guesses as to where somebody else might have looked.

### Making seeing visible

The question of how to improve the monitoring of visual behavior is an important one, yet this problem has not yet been investigated when it comes to learning complex visual tasks. Viewing behavior is typically covert, so making viewing behavior visible might be a promising direction to take. Eye-tracking technology can make viewing behavior visible (Holmqvist et al., 2011; Kok & Jarodzka, 2017). Using eye tracking, Kostons, van Gog, and Paas (2009) replayed participants’ gaze back to them in order help them self-assess their understanding of biology problems. They found that participants with a lower level of expertise remembered their learning process better as a result of the gaze replay, whereas participants with a higher level of expertise were better able to evaluate their task performance. The task to be learned was applying the heredity laws of Mendel, which requires knowledge of the concepts and procedural knowledge of how to solve the tasks, but this was not solely a visual task.

Gaze visualizations have been used in several projects to improve *performance* rather than *monitoring* in complex visual tasks. Gaze-based feedback was found to result in participants missing fewer abnormalities in radiographs (Krupinski, Nodine, & Kundel, 1993; Kundel, Nodine, & Krupinski, 1990) In contrast, simply displaying fixations and saccades to observers yielded less positive results. Donovan, Manning, and Crawford (2008) found that only level-1 radiography students benefited from being presented with a (static) visualization of their own gaze on a chest radiograph, whereas naïve observers, level-2 radiography students, and experts did not benefit. Donovan, Manning, Phillips, Higham, and Crawford (2005) failed to find a benefit of gaze visualizations in a fracture detection task, and even found a negative effect for experts and second year radiography students.

Several researchers have investigated how gaze can be used to impact *learning*. Specifically, can the visualization of expert gaze be used to teach search strategies or visual tasks to learners? This method is called Eye Movement Modeling Examples (EMME; van Gog & Jarodzka, 2013). The results of EMMEs are mixed, but mostly positive when the learning task is a complex visual task (Jarodzka et al., 2012; Jarodzka, van Gog, Dorr, Scheiter, & Gerjets, 2013; Mason, Pluchino, & Tornatora, 2015, 2016; Mehta, Sadasivan, Greenstein, Gramopadhye, & Duchowski, 2005; Nalanagula, Greenstein, & Gramopadhye, 2006; Sadasivan, Greenstein, Gramopadhye, & Duchowski, 2005). In learning procedural problem-solving tasks that are less visually complex, EMMEs failed to improve performance (Van Gog, Jarodzka, Scheiter, Gerjets, & Paas, 2009; van Marlen, van Wermeskerken, Jarodzka, & van Gog, 2016), but observers learning complex visual tasks benefitted from EMME's.

## Gaze-contingent displays

In all of the previous studies the gaze visualizations were presented to the observer *after* the observer executed a task. Would it be useful to display the gaze *during* a visual search task, using a gaze-contingent display? Gaze-contingent displays use eye tracking data as input to update the display in real time. These displays have been mostly used to investigate the effects of peripheral vision on planning of saccades and processing during fixations by changing or taking away peripheral information. It has been widely applied in reading research (e.g., Rayner, 1975) and scene perception research (e.g., Nuthmann, 2014; Vö & Henderson, 2010).

We used a gaze-contingent display in order to show participants their gaze location online, during a search task. If this method is ever going to be useful in real-world tasks, the gaze-contingent display must not degrade performance on the primary task. At the same time, in order for a gaze-contingent display to be effective, it should be very obvious where the observer is looking. Classically, studies have shown gaze by displaying a moving circle or dot. The disadvantage of such displays is that the critical information at the point of fixation is blocked, at least to some extent by the eye tracking display. A better approach might be to take away information from the periphery rather than adding obscuring information at the point of fixation. Jarodzka et al. (2012) used a gaze display where the location fixated was sharp while information peripheral to fixation was blurred. This was more effective in an EMME study than a moving circle. However, since we tend to see sharply only at our point of fixation, blurring out information in the periphery might result

in the situation where the gaze-contingent change of display is practically invisible to the observer (e.g., Loschky, McConkie, Yang, & Miller, 2001). Since we want participants to be clearly aware of their gaze location, we developed a display in which the peripheral part of the image is darker, but still visible, while the gaze location is lighter—a quite literal “spotlight of attention.”

We investigated whether providing observers with online information about where they look during a search task would help them recall their fixation locations afterwards. To do so, we used an eye-tracking apparatus to implement a gaze-contingent spotlight that illuminates the currently fixated location and, thus, shows observers where they are looking. If this spotlight display is found to improve monitoring of visual behavior, a gaze-contingent spotlight could be implemented for learning or executing complex visual tasks.

## Methods

### Participants

Seventeen observers (seven male, 10 female) with a mean age of 28.8 years ( $SD = 7.13$ ) took part in the experiment. They all passed the Ishihara test for color blindness (Ishihara, 1987) and reported normal or corrected-to-normal vision. Ethical approval was granted by The Partners Health Care Corporation Institutional Review Board. All participants gave informed consent and were compensated for their participation. Due to technical difficulties, data of one participant had to be excluded, and search-performance data for the control condition for another participant was not recorded. Data for the manipulation check questionnaire is missing for three participants.

### Apparatus/material

Eye movements were recorded using an EyeLink 1000 desktop mount system (SR Research, Canada) at a sampling rate of 1000 Hz. The stimulus was displayed on a Mitsubishi Diamond Pro 91TXM CRT screen with a refresh rate of 75 Hz and a resolution of  $1024 \times 768$ . The screen was 19 in., and subtended  $33.4^\circ$  of visual angle horizontally and  $25.2^\circ$  of visual angle vertically at a viewing distance of 65 cm. Viewing was binocular, but only the position of the dominant eye was tracked. The Miles test was used to establish eye dominance (Miles, 1930), and only the dominant eye was tracked. A velocity-based, event-detection algo-



rithm was used with a velocity threshold of  $30^\circ/\text{s}$  and an acceleration threshold of  $8000^\circ/\text{s}^2$ .

Stimulus presentation was controlled using MATLAB and the Psychophysics toolbox (Brainard, 1997; Pelli, 1997). The gaze-contingent display was adapted from the Gaze-contingent demo of the EyeLink toolbox of PsychToolbox (Cornelissen, Peters, & Palmer, 2002). For each stimulus, a darker version of the image was used in the periphery and a lighter version was used for foveal vision. These two images were blended into each other via a Gaussian weight mask (an aperture) with a 200 pixel diameter. This gaze-contingent window gave the impression of a roving  $7.5^\circ$  “spotlight” that illuminated everything fixated, while the rest of the display was still visible but darker, thus providing online information about where the observer was looking. The actual size of the spotlight is probably not critical since observers can see inside and outside of its beam. That said, the  $7.5^\circ$  spotlight was large enough to illuminate the  $5^\circ$ – $7^\circ$  useful visual field that is often assumed (e.g., Kundel, Nodine, Thickman, & Toto, 1987).

## Stimuli

We used 190 “Where’s Waldo” images as used in Vö et al. (2016) that were each  $1024 \times 768$  pixels and displayed full screen. Lighter and darker versions of each image were created using the bmp contrast toolbox from [http://www.mccausercontent.sc.edu/crnl/tools/bmp\\_contrast](http://www.mccausercontent.sc.edu/crnl/tools/bmp_contrast). We used the settings “no blur” and “linear transform.” For the darker images, the contrast was set to 0.5 and the brightness to 0.9; for the lighter images the contrast was set to 0.55 and the brightness to 0.4.

## Procedure

The experiment started with the Miles test (Miles, 1930) to establish eye dominance, followed by a nine-point calibration and a validation procedure for the eye tracker. Calibration was only accepted if the average error was under  $0.5^\circ$  for all points, and no larger than  $1.0^\circ$  on any point. If average error was too large after several attempts to calibrate (e.g., due to participants wearing glasses or contact lenses), participants were not included in the study. After every 20th trial, a drift check was conducted. Recalibration was conducted if the average error was over  $1.0^\circ$ . The experiment consisted of three blocks: normal vision (control condition), the gaze-contingent “spotlight” condition, and a 10-trial block in the end in which observers were asked to click on 12 locations where they thought *someone else* might look. The order of blocks one and

two were counterbalanced. Blocks 1 and 2 were 80 trials each, and had an identical structure (see Figure 1): Each trial started with a fixation-cross followed by a 1000 ms screen with a word that indicated what object the observers should search for—for example, “ball” or “bottle.” The target object was never “Waldo.” Then the stimulus was presented and the observers had 3 s to search for the target in either the spotlight or normal display conditions, depending on the block. After 3 s the display disappeared, and observers were asked to respond whether the target was absent or present by a key-press. On one third of the trials, instead of the target absent/present question, observers were asked to report on their fixation locations by clicking on 12 locations where they thought they had just looked in the scene. Only a third of the trials were queried to avoid having observers employ artificial eye movement patterns that would be easy to remember. For the same reason, observers were instructed that the search task was more important than remembering their eye movements. To make sure that observers understood that the gaze-contingent display represented their eye movements, they were not only instructed about this, but also asked in a short manipulation check questionnaire afterwards.

In the third block of trials, observers were asked to place 12 clicks on each of 10 images, based on where they thought another observer would fixate. We can obtain one baseline level of performance from the alignment between an observer’s report of his own fixations and his guesses about where any sensible individual would look. If there is one red poppy in the middle of a field of green grass, you do not need to remember that you looked at it. You know that any sensible person would look at it. In the Waldo pictures, the destination of fixations is less obvious. The data from block three measures the biases that exist, even with these stimuli.

## Analysis

We conducted the same analysis as in Vö et al. (2016). To measure the fidelity of the memory for fixations, we placed a virtual circle around each fixation and each click and measured the overlap. As the virtual circles increase in size, the overlap between fixations and click circles increases as well. Clicks and fixations are very unlikely to fall on exactly the same single pixel, so the overlap will be 0% with the smallest virtual circle size. If the virtual circles are large enough, overlap will approach 100%. We calculated the overlap for radii of  $0.5^\circ$ ,  $1^\circ$ ,  $2^\circ$ ,  $3^\circ$ ,  $4^\circ$ , and  $5^\circ$  of visual angle (see Figure 2). In order to do paired-sample *t* tests, we averaged over the six radii for each observer.

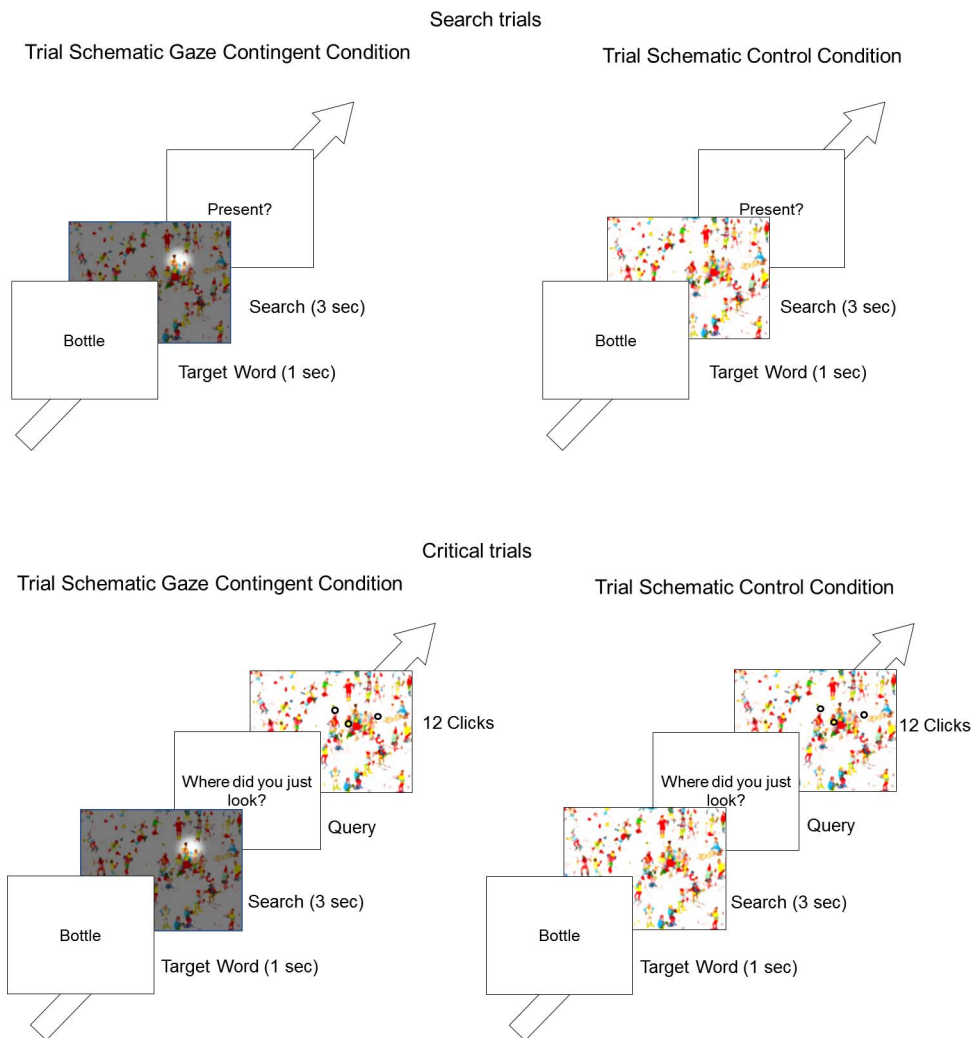


Figure 1. Trial schematic: On search trials, observers were presented with a target word (in the example “Bottle”) and then searched the image for the target word for 3 s after which they provided a target present/absent response. On critical trials (1/3 of all trials), observers skipped the present/absent response, and were instead asked to mark 12 locations fixated in the prior 3 s. The scene was now presented without the gaze-contingent spotlight.

We compared overlap between clicks (reported fixation locations) and actual fixation locations in both conditions with ceiling and chance performance. Ceiling performance was modeled by comparing clicks with a model that placed a “click” at the location of fixation with a  $1^\circ$  jitter around it (i.e., perfect fixation memory with modest noise in the placement of the “clicks”). Chance performance was modeled by comparing fixations with the clicks an observer made on a different image. Finally, we calculated the overlap between where an observer expected somebody else would look (on 10 images) and the actual fixations of the next observer on these trials. Cohen’s  $d$  was used as an effect size, with 0.2 representing small effects, 0.5 for medium effects, and 0.8 for large effects.

As our task requires participants to remember approximately 12 fixation locations, memory limita-

tions might impact monitoring performance in this task. Based on primacy and recency effects (Murdock, 1962), it might be expected that the first and last two fixations are remembered best. Thus, we additionally analyzed the percentage overlap between just the first and last two fixations with all clicks in the two conditions.

## Results

### Search performance

No significant differences between the control condition and the spotlight condition were found in search performance (number of true positives and true

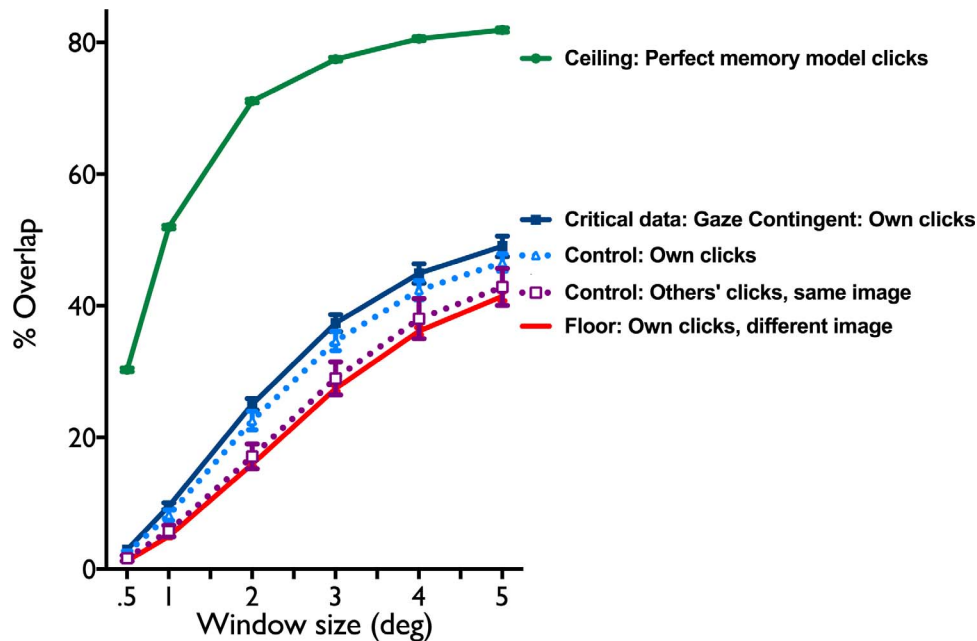


Figure 2. Data represent the percentage overlap between “windows” of different size radius (in degrees of visual angle) around each fixation and clicks on where observers thought they looked in the image for the spotlight (gaze-contingent) condition (dark blue) and the control condition (light blue, dotted). An observer’s fixations overlapped with where a different observer thought someone else might look at in the same image is shown in purple dotted. The perfect memory model represents ceiling performance (green), while the overlap with another image represents chance performance (red). Error bars represent *SE*.

negatives),  $M_{control} = 40.60$ ,  $SD = 4.70$ ,  $M_{spotlight} = 40.56$ ,  $SD = 4.18$ ,  $t(14) = 0.213$ ,  $p = 0.83$ . Recall that search trials are a filler task in the experiment, so the relatively poor performance on this task is not of much interest. It would often take longer than 3 s to actually find an item in these Where’s Waldo images.

### Manipulation check questionnaire

On the question “How noticeable was the spotlight?” with 0 representing “no different between the spotlight and the surrounding image” and 10 representing “very evident spotlight,” the average answer was 8.21 ( $SD = 1.89$ ) and the minimum was 4. On the question “How closely did the spotlight follow where you looked?” with 0 representing “not at all” and 10 representing “perfect match between where I looked and the spotlight,” the average was 7.93 ( $SD = 1.44$ ). Nine out of the 14 participants reported that the spotlight made remembering their eye-movements easier; one participant reported that it made matters worse while four participants reported that it made no difference. Nine participants reported that the spotlight made search harder, whereas three reported that it made search easier, and two reported that it made no difference.

### Overlap

Figure 2 shows the overlap between actual and reported fixations in the normal condition and the spotlight condition, ceiling performance, chance performance, and overlap with where an observer expected somebody else would look. Descriptive statistics reported below average over window sizes, yielding an average window size of  $2.6^\circ$ , and Figure 3 shows these averages for each condition. In the spotlight condition, the average percentage overlap was 28.1% ( $SD = 4.01$ ), which was slightly better than the control condition (26.1%,  $SD = 3.3$ ),  $t(15) = 2.48$ ,  $p = 0.03$ , Cohen’s  $d = 0.43$ .

Ceiling performance was estimated at 66.0% ( $SD = 1.28$ ). Both the spotlight and the control condition scored significantly below the ceiling performance: control condition,  $t(15) = 49.3$ ,  $p < 0.001$ ,  $d = 11.62$ ; and spotlight condition,  $t(15) = 39.62$ ,  $p < 0.001$ ,  $d = 9.2$ .

Chance performance was estimated at 21.2% ( $SD = 1.28$ ). Both conditions scored significantly higher than chance performance: control condition,  $t(15) = 6.28$ ,  $p < 0.001$ ,  $d = 1.44$ ; and spotlight condition,  $t(15) = 7.39$ ,  $p < 0.001$ ,  $d = 1.69$ .

Data on “where someone else might look” was missing for three participants. The average overlap between where someone else might look and actual fixations was 22.4% ( $SD = 4.78$ ). This was slightly but

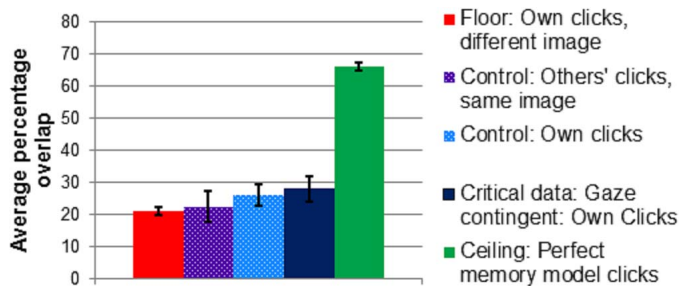


Figure 3. Average percentage overlap between “windows” of different size radius (average overlap  $2.6^\circ$  of visual angle) around each fixation and clicks on where observers thought they looked in the image for the spotlight (gaze-contingent) condition (black and yellow) and the control condition (dark blue). An observer’s fixations overlapped with where a different observer thought someone else might look at in the same image is shown in purple. The perfect memory model represents ceiling performance (green), while the overlap with another image represents chance performance (red). Error bars represent standard deviations.

significantly lower than the spotlight condition,  $t(13) = 4.2$ ,  $p = 0.001$ ,  $d = 0.8$  and the control condition,  $t(13) = 5.22$ ,  $p < 0.001$ ,  $d = 1.1$ .

There was no significant correlation between the percentage overlap between clicks and fixations and behavioral performance on the search task for the control condition ( $r = -0.18$ ,  $p = 0.53$ ) or the spotlight condition ( $r = -0.15$ ,  $p = 0.57$ ), indicating that people with a better memory for their fixation locations did not do significantly worse on the search task.

### Overlap of clicks with first two and last two fixations

When only the first two and the last two fixations were included, the average percentage overlap in the spotlight condition was 28.9% ( $SD = 4.70$ ), which was slightly better than the control condition’s 26.2% ( $SD = 2.2$ ),  $t(15) = 2.41$ ,  $p = 0.03$ , Cohen’s  $d = 0.50$ . The average percentage overlap for both conditions with only the first two and last two fixations included is very similar to the average percentage overlap when all fixations are included (spotlight condition, 28.9% vs. 28.1%; control condition, 26.2% vs. 26.1%). Indeed, no significant difference is found between the mean values in the analysis with only those four fixations included versus the analysis with all fixations included for the spotlight condition,  $t(15) = 1.11$ ,  $p = 0.28$ , and the control condition,  $t(15) = 0.08$ ,  $p = 0.93$ .

## Discussion

We investigated whether using a gaze-contingent “spotlight” to provide observers with online information about where they looked during a search task would help them recall their own fixations afterwards. In both the spotlight condition and the control condition, observers showed suboptimal introspection about their own fixation locations: The overlap between their reported fixation locations and actual fixation locations is significantly better than chance performance, but far from perfect monitoring performance. The overlap between reported and actual fixation locations was significantly larger in the spotlight condition than in the control condition. Since the search performance did not decrease, we can conclude that the gaze-contingent spotlight did improve recall of fixations without decreasing search performance. However, while the difference between the conditions was statistically significant, it was not very large (only 2%). The real message from these results is that explicit information about eye position does not produce a clear memory for those eye positions when observers are asked about them a few seconds later.

Interestingly, observers reported that they *thought* that the spotlight improved their memory for the fixations. We do not have enough power to make a statistical comparison, but anecdotal evidence suggests that participants’ introspection about the degree to which the spotlight helps might be, if anything, inversely correlated with actual performance. The nine participants who reported that the spotlight made it easier to remember eye movements improved on average by 1.29% in the spotlight condition, whereas the four participants who reported that they experienced no difference, actually improved by 5.5%.

Gaze visualizations have previously been used to improve performance or learning (e.g., Kundel, Nodine, & Carmody, 1978; van Gog & Jarodzka, 2013), but these have been visualizations presented to the observer after the eye movements have been completed. Our investigation of the use of online gaze visualization as a way to improve monitoring in visual tasks is an approach that, to our knowledge, has not been previously reported. Our gaze spotlight of visualizing gaze by darkening the periphery proved to be clearly visible without disturbing search performance, and could be implemented in EMME’s or other gaze visualizations.

Whereas self-monitoring may be critical in learning complex tasks, our results suggest that self-monitoring of eye position in day to day life may not be critical as it is not available. Performance is not very good in the control condition and, whereas our “spotlight” display improved performance, the improvement was small



and did not bring performance to anything like the level of full memory for even a few seconds' worth of fixations. Of course, a substantial limit on performance in this task may lie in the general limits on short-term or working memory. We are asking for 12 clicks, and it could be that observers only recall 3–5 (Cowan, 2001). However, the data suggest that this is not the constraint on performance. If it was, one would expect observers to do significantly better on the first and last two fixations, and we did not find such a difference. The similarity of the two measures suggests that only very little specific information about fixation is surviving in memory, not even 3–5 items.

In this experiment, we informed participants in the control and spotlight condition that we would ask them about their fixation locations. In Vö et al. (2016), participants were not informed about this in the equivalent of our control condition. They found that participants were only marginally better in locating their own (average overlap is 27%) as compared to someone else's (average overlap is 24%) fixations. We found a significantly higher performance in locating their own (average overlap 26%) as compared to somebody else's fixations (average overlap is 22%). Whereas the difference between locating your own and somebody else's fixations has increased (and thus reached statistical significance), the mean overlap between clicks and fixations between the earlier experiment and the present experiment did not change much. Thus, this suggests that our instruction did not produce better recognition memory than the earlier experiment. Van Wermeskerken et al. (2017) found similar results, showing no differences between participants who were informed that they would be asked to recognize their fixation locations and participants who were not informed about it.

In the context of complex visual tasks such as radiology, one might ask whether experts in these tasks would be better able to remember their own viewing behavior. It has been found that experts in complex visual tasks remember visual stimuli in their expertise domain differently from nonexperts (Evans et al., 2011; Myles-Worsley, Johnston, & Simons, 1988). Future directions include repeating this experiment with experts viewing stimuli in their domain of expertise. However, given the results from experiments testing memory for fixations, it seems likely that we are simply not equipped to record our own eye position, even when it might be useful.

Differences between Waldo images and domain-specific stimuli such as luggage scans and radiographs might limit the generalizability of our results. Thus, it would be useful to replicate these findings with expert observers, working in their domain of expertise. It is possible that learned and systematic behaviors like “drilling” through a stack of CT images, or always

using the same order of checking anatomical regions in a chest radiograph (Drew et al., 2013; Kok et al., 2016) could compensate for the lack of memory for fixations revealed in our data. For example, we are all reading “experts,” and we know how much of a page we have read, not by monitoring our fixations, but by knowing the current point of fixation and that one begins reading in the upper left corner of the page. Like the structure of a page, the anatomy in radiographs may further help monitoring. Indeed, Vö et al. (2016) found more overlap between actual fixations and reported fixations in semantically structured photographs as compared to less semantically structured Waldo images. This might result in slightly more overlap between reported and actual fixations by experts in a visual task. Furthermore, experts might target certain relevant areas of the stimulus and purposefully ignore others as a function of their expertise. This approach might also impact the generalization of our results.

Another limitation of this study is that we used a short task. Our observers often failed to find the target in the 3 s of exposure. It might be interesting to give observers unlimited time for a search. We would not expect observers to recall dozens of precise fixations; however, it is possible that the moving spotlight would speed search or increase accuracy in a manner that the present experiment was not designed to detect.

## Conclusion

In conclusion, monitoring viewing behavior could be useful in learning and/or conducting complex visual tasks in radiology, air traffic control, meteorology, etc. However, observers have poor introspection about their own eye movements. A spotlight display that provides online information about gaze location did not help monitoring very much. It seems that a different approach would be needed if we wanted to improve self-monitoring of fixations.

*Keywords:* eye movements, fixation memory, gaze-contingent display, introspection

## Acknowledgments

This work was funded by DFG grant VO 1683/2-1 to MLV and NIH grants EY017001 and CA207490 to JMW. Special thanks to Abla Alaoui-Soce and Chia-Chen Wu for helping us collect data, and to Krista Ehinger for help with programming the experiment.

Commercial relationships: none.

Corresponding author: Ellen M. Kok.

E-mail: e.m.kok@uu.nl.

Address: School of Health Profession Education,  
Department of Educational Research and  
Development, Maastricht University, Maastricht, The  
Netherlands.

## References

- Brainard, D. H. (1997). The psychophysics toolbox. *Spatial Vision*, *10*, 433–436.
- Clarke, A. D. F., Mahon, A., Irvine, A., & Hunt, A. R. (2016). People are unable to recognize or report on their own eye movements. *The Quarterly Journal of Experimental Psychology*, *70*(11), 2251–2270, doi:10.1080/17470218.2016.1231208.
- Cornelissen, F. W., Peters, E. M., & Palmer, J. (2002). The EyeLink Toolbox: Eye tracking with MATLAB and the Psychophysics Toolbox. *Behavior Research Methods, Instruments, & Computers*, *34*(4), 613–617, doi:10.3758/BF03195489.
- Cowan, N. (2001). The magical number 4 in short-term memory: A reconsideration of mental storage capacity. *Behavioral and Brain Sciences*, *24*(1), 87–114; discussion, 114–185.
- Donovan, T., Manning, D. J., & Crawford, T. (2008). *Performance changes in lung nodule detection following perceptual feedback of eye movements*. Paper presented at the Proceedings of SPIE—The International Society for Optical Engineering, San Diego, CA.
- Donovan, T., Manning, D. J., Phillips, P. W., Higham, S., & Crawford, T. (2005). *The effect of feedback on performance in a fracture detection task*. Paper presented at the Proceedings of SPIE—The International Society for Optical Engineering, San Diego, CA.
- Drew, T., Vo, 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*(10):3, 1–13, doi:10.1167/13.10.3. [PubMed] [Article]
- Dunlosky, J., & Rawson, K. A. (2012). Overconfidence produces underachievement: Inaccurate self evaluations undermine students' learning and retention. *Learning and Instruction*, *22*(4), 271–280, doi:10.1016/j.learninstruc.2011.08.003.
- Evans, K. K., Cohen, M. A., Tambouret, R., Horowitz, T., Kreindel, E., & Wolfe, J. M. (2011). Does visual expertise improve visual recognition memory? *Attention, Perception, & Psychophysics*, *73*(1), 30–35, doi:10.3758/s13414-010-0022-5.
- Foulsham, T., & Kingstone, A. (2013). Where have eye been? Observers can recognise their own fixations. *Perception*, *42*(10), 1085–1089, doi:10.1068/p7562.
- Greene, J. A., & Azevedo, R. (2010). The measurement of learners' self-regulated cognitive and metacognitive processes while using computer-based learning environments. *Educational Psychologist*, *45*(4), 203–209, doi:10.1080/00461520.2010.515935.
- Holmqvist, K., Nyström, M., Andersson, R., Dewhurst, R., Jarodzka, H., & van de Weijer, J. (2011). *Eye tracking: A comprehensive guide to methods and measures*. Oxford, UK: Oxford University Press.
- Ishihara, I. (1987). *Ishihara's tests for color-blindness (Concise ed.)*. Tokyo, Japan: Kanehara & Co.
- Jarodzka, H., Balslev, T., Holmqvist, K., Nyström, M., Scheiter, K., Gerjets, P., & Eika, B. (2012). Conveying clinical reasoning based on visual observation via eye-movement modelling examples. *Instructional Science*, *40*(5), 813–827, doi:10.1007/s11251-012-9218-9225.
- Jarodzka, H., van Gog, T., Dorr, M., Scheiter, K., & Gerjets, P. (2013). Learning to see: Guiding students' attention via a Model's eye movements fosters learning. *Learning and Instruction*, *25*(0), 62–70, doi:10.1016/j.learninstruc.2012.11.004.
- Kok, E. M., & Jarodzka, H. (2017). Before your very eyes: The value and limitations of eye tracking in medical education. *Medical Education*, *51*(1), 114–122, doi:10.1111/medu.13066.
- Kok, E. M., Jarodzka, H., de Bruin, A. B. H., BinAmir, H. A. N., Robben, S. G. F., & van Merriënboer, J. J. G. (2016). Systematic viewing in radiology: Seeing more, missing less? *Advances in Health Sciences Education*, *21*(1), 189–205, doi:10.1007/s10459-015-9624-y.
- Kostons, D., van Gog, T., & Paas, F. (2009). How do I do? Investigating effects of expertise and performance-process records on self-assessment. *Applied Cognitive Psychology*, *23*(9), 1256–1265, doi:10.1002/acp.1528.
- Krupinski, E. A., Nodine, C. F., & Kundel, H. L. (1993). Perceptual enhancement of tumor targets in chest-x-ray images. *Perception & Psychophysics*, *53*(5), 519–526.
- Kundel, H. L., Nodine, C. F., & Carmody, D. (1978). Visual scanning, pattern recognition and decision-making in pulmonary nodule detection. *Investigative Radiology*, *13*(3), 175–181.
- Kundel, H. L., Nodine, C. F., & Krupinski, E. A. (1990). Computer-displayed eye position as a visual

- aid to pulmonary nodule interpretation. *Investigative Radiology*, 25(8), 890–896.
- Kundel, H. L., Nodine, C., Thickman, D., & Toto, L. C. (1987). Searching for lung nodules: A comparison of human performance with random and systematic scanning models. *Investigative Radiology*, 22, 417–422.
- Loschky, L. C., McConkie, G. W., Yang, J., & Miller, M. E. (2001). *Perceptual effects of a gaze-contingent multi-resolution display based on a model of visual sensitivity*. Paper presented at the the ARL Federated Laboratory 5th Annual Symposium—ADID Consortium Proceedings, College Park, MD.
- Marti, S., Bayet, L., & Dehaene, S. (2015). Subjective report of eye fixations during serial search. *Consciousness and Cognition*, 33, 1–15, doi:10.1016/j.concog.2014.11.007.
- Mason, L., Pluchino, P., & Tornatora, M. C. (2015). Eye-movement modeling of integrative reading of an illustrated text: Effects on processing and learning. *Contemporary Educational Psychology*, 41, 172–187, doi:10.1016/j.cedpsych.2015.01.004.
- Mason, L., Pluchino, P., & Tornatora, M. C. (2016). Using eye-tracking technology as an indirect instruction tool to improve text and picture processing and learning. *British Journal of Educational Technology*, 47(6), 1083–1095, doi:10.1111/bjet.12271.
- Mehta, P., Sadasivan, S., Greenstein, J. S., Gramopadhye, A. K., & Duchowski, A. T. (2005). Evaluating Different Display Techniques for Communicating Search Strategy Training in a Collaborative Virtual Aircraft Inspection Environment. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting, Orlando, FL*, 49(26), 2244–2248, doi:10.1177/154193120504902606.
- Miles, W. R. (1930). Ocular Dominance in Human Adults. *The Journal of General Psychology*, 3(3), 412–430, doi:10.1080/00221309.1930.9918218.
- Murdock, B. B. (1962). The serial position effect of free recall. *Journal of Experimental Psychology*, 64(5), 482.
- Myles-Worsley, M., Johnston, W. A., & Simons, M. A. (1988). The influence of expertise on X-ray image processing. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 14(3), 553.
- Nalanagula, D., Greenstein, J. S., & Gramopadhye, A. K. (2006). Evaluation of the effect of feedforward training displays of search strategy on visual search performance. *International Journal of Industrial Ergonomics*, 36(4), 289–300, doi:10.1016/j.ergon.2005.11.008.
- Nelson, T. O., & Nahrens, L. (1990). Metamemory: A theoretical framework and new findings. In G. H. Bower (Ed.), *Psychology of learning and motivation* (Vol. 26, pp. 125–173). New York: Academic Press.
- Nuthmann, A. (2014). How do the regions of the visual field contribute to object search in real-world scenes? Evidence from eye movements. *Journal of Experimental Psychology: Human Perception and Performance*, 40(1), 342.
- Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: Transforming numbers into movies. *Spatial Vision*, 10(4), 437–442.
- Rayner, K. (1975). The perceptual span and peripheral cues in reading. *Cognitive Psychology*, 7(1), 65–81, doi:10.1016/0010-0285(75)90005-5.
- Rubin, G. D., Roos, J. E., Tall, M., Harrawood, B., Bag, S., Ly, D. L., . . . Choudhury, K. R. (2014). Characterizing search, recognition, and decision in the detection of lung nodules on CT scans: Elucidation with eye tracking. *Radiology*, 274(1), 276–286.
- Sadasivan, S., Greenstein, J. S., Gramopadhye, A. K., & Duchowski, A. T. (2005). *Use of eye movements as feedforward training for a synthetic aircraft inspection task*. Paper presented at the Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, Portland, Oregon, USA.
- van Gog, T., & Jarodzka, H. (2013). Eye tracking as a tool to study and enhance cognitive and metacognitive processes in computer-based learning environments. In R. Azevedo & V. Aleven (Eds.), *International handbook of metacognition and learning technologies* (pp. 143–156). New York: Springer Science and Business media.
- Van Gog, T., Jarodzka, H., Scheiter, K., Gerjets, P., & Paas, F. (2009). Attention guidance during example study via the model's eye movements. *Computers in Human Behavior*, 25(3), 785–791, doi:10.1016/j.chb.2009.02.007.
- van Marlen, T., van Wermeskerken, M., Jarodzka, H., & van Gog, T. (2016). Showing a model's eye movements in examples does not improve learning of problem-solving tasks. *Computers in Human Behavior*, 65, 448–459, doi:10.1016/j.chb.2016.08.041.
- Van Wermeskerken, M., Litchfield, D., & Van Gog, T. (2017). What am I Looking at? Interpreting dynamic and static gaze displays. *Cognitive Science*, E-pub ahead of print, doi:10.1111/cogs.12484.
- Vö, M. L. H., Aizenman, A. M., & Wolfe, J. M. (2016).

You think you know where you looked? You better look again. *Journal of Experimental Psychology: Human Perception and Performance*, 42(10), 1477–1481, doi:10.1037/xhp0000264.

Vö, 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):14, 1–13, doi:10.1167/10.3.14. [PubMed] [Article]

Winne, P. H., & Hadwin, A. F. (1998). Studying as self-regulated learning. *Metacognition in Educational Theory and Practice*, 93, 27–30.

Zimmerman, B. J., & Schunk, D. H. (2001). *Self-regulated learning and academic achievement: Theoretical perspectives*. Mahwah, NJ: Lawrence Erlbaum Associates.