Goal-Driven and Bottom-up Gaze in an Active Real-World Search Task

Tom Foulsham
University of Essex, UK
foulsham@essex.ac.uk

Alan Kingstone
University of British Columbia, Canada
alan.kingstone@ubc.ca

Abstract

Mobile eye tracking has become a useful tool in studies of vision and attention in real-world tasks. However, there remains a disconnection between such studies and the laboratory paradigms used by cognitive psychology. In particular, visual search has been studied intensively, but lab search often differs from search in the real world in many respects (e.g., in reality one must walk and move head and eyes to find the target, target and distractors are not equally visible, and objects are frequently occluded). Here, we took a broader view of search behaviour and analyzed the gaze of participants who were asked to walk around within a building, find a room, and then locate a target mailbox. Our aim was to describe the differences in behaviour according to principles of (lab-based) visual search, and we did this by testing the effects of top-down instructions (i.e. having more or less information about where to go) and target saliency (i.e. having a more or less distinctive target to look for). These factors made a difference in a real world context by changing the frequency with which signs and cues in the environment were fixated, and by affecting head and eye movements in the mail-room. Bottom-up saliency had little effect on search time, but our approach revealed how it influenced the coordination of gaze, while still allowing us to make contact with laboratory paradigms.

CR Categories: J.4 [Social and Behavioural Sciences]: Psychology

Keywords: mobile eye tracking, search, attention

1 Introduction

Think of the last time you located your keys on your desk. Or, on a much larger scale, finding a colleague's office in a building that you have not been to before. We might reasonably call both tasks "visual search", as they require matching a known target (a representation of keys or information about the office) to visual features in the environment. However, these tasks are also very different from the type of laboratory visual search traditionally studied by cognitive psychology. In laboratory search, displays normally consist of a target surrounded by several distractors, which are differentiated on the basis of one or two simple visual dimensions (for example, a 'Q' amongst 'O's, a 'T' amongst 'L's or a red horizontal line amongst green horizontal lines and red vertical lines [e.g., Wolfe 1998]. These items are typically arranged randomly on a blank background in a relatively small space (normally a monitor) that lies completely within the participant's visual field. Researchers have also explored search within pictures of real-world scenes, but these also tend to be confined to a monitor [Foulsham and Underwood, 2007].

During search in the real world, the target is not normally in the visual field at the onset of search. Occluding items or obstacles may have to be moved in order to see or access the target. Locating the target normally requires a whole sequence of complex actions in order to bring the target item into view so that it may be recognized and used. Eye, head and whole body movements must be made in order to locate the target, and in the example of finding an office, the searcher must change significantly their own location within the environment. Targets and background are often complex and defined in terms of a whole range of features. Finally, there are often other cues that can guide our search, and many ways of actively seeking such cues. Our knowledge of the places that items are likely to occur, as well as our memory for where we saw them before, will guide efficient real-world search. We place signs and maps around our environment in order to help people find locations and objects. The present investigation considers the performance and allocation of attention of participants engaged in a realistic, complex search task. Several previous studies have considered gaze allocation during mobile or active tasks [e.g., Hayhoe and Ballard 2005; Land et al. 1999]. However, these studies have normally been concerned with different topics (such as the selection of information over time) and used different measures of performance from laboratory visual search. The present study therefore sought to use mobile eye tracking to investigate search in a way that could be compared more easily to laboratory data. This also provides a test of how well principles of visual search scale up to real-world, active behavior. In particular, we focus on the key distinction between bottom-up and top-down guidance in attention. This distinction refers to the tendency for attention and fixation to be drawn to regions based either on stimulus features (bottom-up) or on task or environment knowledge possessed by the system (top-down). Many current theories and computational models focus on describing these processes, their interaction and their instantiation in the brain [e.g., Itti and Koch, 2001].

The current study recorded the gaze of participants as they completed a two-stage search task requiring them to first find a room in a building, and then find and retrieve an envelope from a mailbox in that room. Head-centred gaze was recorded using a mobile eye-tracker, and searchers were free to walk around, move their head and body and use whatever cues were available in the environment to complete their task. In addition, we introduced two manipulations in order to test the generalisability of principles from visual search in the lab. First, we varied the

ETRA 2012, Santa Barbara, CA, March 28 – 30, 2012.
© 2012 ACM 978-1-4503-1225-7/12/0003 $10.00

instructions for the room-finding task. Our hypothesis was that this top-down manipulation would affect attention and search time, so that more specific instructions (providing a room number) would lead to faster search and more fixations on these numbers in the environment. Second, we varied the bottom-up conspicuity of the target mailbox, making the hypothesis that a distinctive target would pop-out from the surroundings and be found more quickly.

2 Method

2.1 Participants

Thirty undergraduates (16 female) from the University of British Columbia took part in exchange for course credit. Participants gave their informed consent before beginning the experiment had normal vision and none were wearing glasses.

2.2 Apparatus and calibration

Head-centred gaze was recorded using the MobileEye system (Applied Science Laboratories; Bedford, MA), consisting of two small cameras mounted on a pair of lightweight glasses. The equipment recorded the position of the right eye and the scene in front of the observer. The scene camera was adjusted to have a field of view aligned with the participant's line of sight, and both cameras recorded to a digital videocassette recorder that the participant carried in a small backpack on their back. The MobileEye has an instrumental resolution of better than 1˚ and a tracking range of approximately 60˚ horizontally by 40˚ vertically. Video frames were recorded at 60Hz and scene and eye images were interleaved, giving an effective temporal resolution of 30Hz.

Calibrations were performed before and after the search task by recording gaze while participants fixated each of 9 points that were marked on the wall of a testing room with similar lighting conditions to the route that would be walked. Data from participants whose calibrations showed significant deterioration after they had completed the task (e.g. because the MobileEye glasses had slipped) were discarded.

2.3 Procedure

Following successful calibration, participants were given written instructions describing the search task and were led to the start point which was a door exiting the laboratory. Participants were instructed that they had to walk through the building, find the faculty mailroom, retrieve an envelope from a particular mailbox, and return it to the laboratory. This task took place in the Douglas T. Kenny building, which houses the Psychology Department at the University of British Columbia. We will discuss our results in terms of two stages within the task: Stage 1 (finding the correct room) and Stage 2 (finding the mailbox containing the envelope), although these components were not explicitly differentiated for the participants.

Stage 1 required a short walk in the four-story Kenny building. The most direct route from the start point (the laboratory) to the mailroom involved walking along two straight corridors, down a flight of stairs, through a small atrium and along another corridor, a walk that takes about 50-60 seconds. However, the majority of participants took a more circuitous route, and debriefing after the experiment confirmed that all were naïve to the location of the mailroom and had not visited it previously. The route featured, among many other landmarks such as windows, doors and posters, multiple signs, room numbers and floor plans, and there was the potential for seeing other individuals in the building. Figures 1 and 2 show example frames from the MobileEye scene camera, which are representative of the route.

We manipulated the specificity of the instructions given to participants, hypothesizing that this would influence the time taken to find the room and the cues that were selected along the way. Half the participants were given instructions which specified the room number but did not describe the room, while the other half were given less specific instructions telling them that the mailroom was "on the second floor of the building", that the door would be open, and that there was a photocopier near the door. Both sets of instructions specified that, once in the mailroom, participants should find the mailbox labeled "Kingstone Lab" and bring back the envelope inside.

Stage 2 required locating the correct mailbox on entering the mailroom, and retrieving the envelope (which was always the only item in the mailbox). The mailbox was contained in an array of approximately 120 highly similar boxes taking up the back wall of the mailroom, straight in front of the door through which the participants entered (see Figures 1 and 3). Thus finding the mailbox can be construed as a rather difficult visual search task. Mailboxes were 10 cm wide by 32 cm tall, and the target mailbox was 150 cm above the floor. The correct mailbox was inconspicuously labeled with the name of the laboratory (in letters approximately 1 cm high), and all participants were informed of this name. There were also multiple, irrelevant distractors in the room such as posters and magazines.

We manipulated the conspicuity of the target for half of the participants by adding a brightly-colored, pink paper frame which was affixed to the outside of the mailbox (see Figure 3), and which marked it out relative to the other, homogenously colored mailboxes. Participants were not informed of which type of target they would be searching for.

Figure 1 *Example frames from the MobileEye scene camera.*

3 Results

Gaze information from the eye camera was combined with the view from the head-mounted scene camera using software from ASL. This software generated a 30 frames-per-second video in which the point of regard at each point in time was superimposed over the scene with a red cursor. As well as looking at the time each participant took to complete Stages 1 and 2, the videos were hand coded using custom-written software to test several hypotheses about where people would look during the task. Coders recorded a fixation whenever eye position remained on an object for at least 2 consecutive frames (i.e. longer than approximately 66ms). Due to the difficulty of automatically parsing fixations, saccades and pursuit eye movements in a mobile situation (and with the available temporal frequency) we did not compute overall eye movement statistics for the task. This would, however, be interesting for future comparisons with search in the laboratory.

3.1 Stage 1: finding the mailroom

Participants took a mean of 182s to correctly find the mailroom, although this varied considerably between participants (SD=148s). One participant (from the vague instructions condition) failed to find the mailroom and gave up, so their data was excluded from further analysis. The time to find the mailroom was reliably affected by the specificity of the instructions, with participants taking an average of 114s (SD=51s) when given the room number and 254s (SD=183s) when given less specific directions (between group t test, $t(25)=2.7$, $p=.01$). Thus increasing the specificity of the search instructions led to faster search performance.

When walking the route, participants made fixations on walls and windows, and quite often they looked at the floor and at objects associated with upcoming actions (such as door handles and steps as going down the stairs). This is consistent with previous research showing the dependence of gaze on action in natural behaviour [Hayhoe and Ballard, 2005]. In order to investigate whether the change in instructions was also associated with different gaze patterns, we analysed the frequency with which participants looked at two types of cues in the environment: room numbers and other signs. Room numbers were present on all the doors in the building but were small and rather inconspicuous. Signs included signs indicating the exits and stairwells in the building, floor plans, and other miscellaneous postings such as directions to a particular laboratory. Of course, by virtue of taking longer to find the room, the group receiving vague instructions had more opportunity to fixate these objects of interest. We therefore equated the two conditions by analysing the frequency of fixations during only the first minute of the walk, because all participants took at least this long. Figure 2 plots the frequency of fixations during this minute on room numbers and signs in the two conditions.

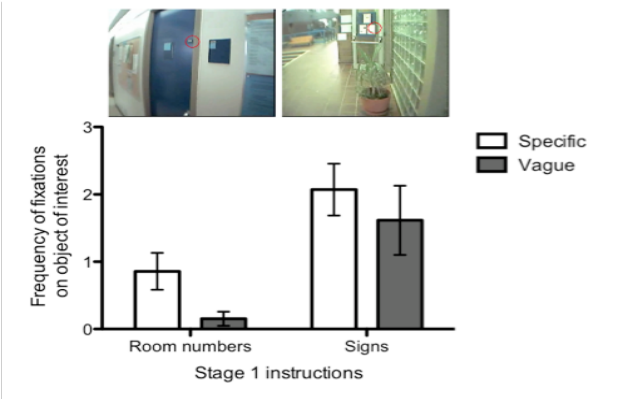

Figure 2 *Mean (+/- 1 SEM) frequency of fixations on room numbers and signs during the first minute of searching for the mailroom. Examples of fixation on each object are also shown.*

Participants who received specific instructions giving the target room number looked at room numbers reliably more often than those who received vague instructions ($t(25)=2.3$, $p<.05$). Most of the participants in the vague condition never fixated a single room number, despite walking past many of them. There was no reliable difference between conditions in the number of fixations on signs ($t(25)<1$).

3.2 Stage 2: finding the mailbox

We defined search time as the time between entering the mailroom and touching the envelope in the correct mailbox. One participant (from the homogenous mailbox condition) failed to find the correct mailbox and data from this participant, (along with data from the participant who failed to find the mailroom altogether!), are excluded from further analyses. The remaining participants took 32.1s on average to find the target (SD=19.3s). Surprisingly, given the extensive evidence from visual search in the laboratory that colour singletons should pop-out and be found more quickly, the pop-out mailbox was not found any quicker than the homogenously coloured mailbox (in fact it was found slightly less quickly, 33s vs. 31s; $t(24)<1$).

Although the pop-out mailbox did not affect search time, the mobile gaze recordings were inspected for differences in head and eye movements. In both conditions, participants spent the majority of the time moving their head and body around the room and the mailbox array, and for most of this time the target was not yet within their central visual field (as defined by the field of view of the scene camera). On average, 1.2 fixations were made on the target mailbox before the envelope was grasped and this did not differ between conditions ($t(24)<1$). In

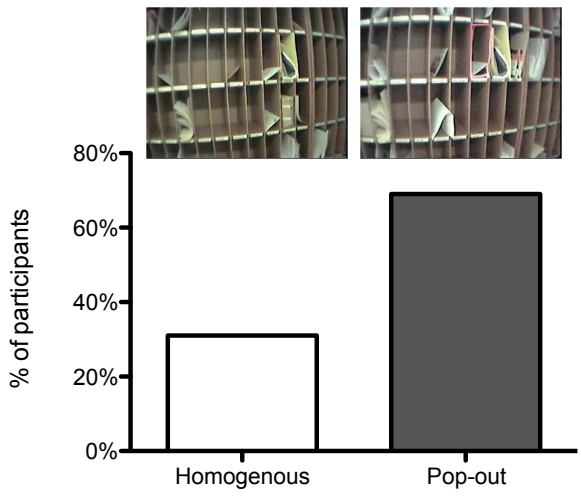

Figure 3 *The probability of the target being found on the first occasion that it entered the line of sight. An example of each type of mailbox is shown above.*

fact, the only significant difference between the two types of mailbox concerned the likelihood that the target was found rather than missed when it was brought into the camera's line of sight. Pop-out targets were more likely to be reached for the first time they were within the camera's field of view than homogenous targets (see Figure 3; χ^2=3.8, df=1, p<.05). Conversely, homogenous mailboxes were more likely to be passed over, requiring additional head and eye movements to orient back to the target. Thus a bottom-up factor (the brightness of the mailbox) had a subtle effect on search behaviour that was captured by our use of mobile gaze tracking.

4 Discussion

Visual search has been well studied in cognitive psychology, where it is a crucial paradigm for the investigation of attention. However, most studies have been confined to investigating the deployment of attention in simple displays within the central visual field of stationary participants (i.e. searching on a screen in the lab). Eye movements during real-world activities have been studied previously, [Hayhoe and Ballard 2005; Land et al. 1999] and here we used these methods to study search.

Attentional selection is often described according to top-down and bottom-up control, principles derived from experiments in the lab. We addressed the question of whether these principles would have observable effects on a realistic search task, and if so, how these effects would be manifested in active gaze.

Our top-down manipulation—the instructions given to participants looking for the mailroom—had an effect on the strategy adopted by participants. Those that were given specific instructions including a room number fixated on these numbers significantly more than those given vague instructions. In contrast, both groups used building signs equivalently, indicating that the use of room numbers was not an artifact of an overall between-group difference in head position or participants' willingness to select and attend to building information. It is likely that other differences could be tested in these data (for example the vague group may have looked more at large objects in the hall because they were told the mailroom contained a photocopier). The data show clearly that the volitional top-down search strategy adopted by participants in the specific instruction condition was distinct. More broadly, this is a simple example of top-down gaze control in a natural setting: rather than being determined only by the visual information in the environment, participants allocated their gaze differently depending on their knowledge about the task.

Bottom-up attention was manipulated by making the targeted mailbox in the mailroom either visually distinct by surrounding it with brightly colored paper or equivalent to the other mailboxes by removing this border. This manipulation had an effect on the way that participants oriented their head towards the target because additional head movements were made to bring non-salient targets back into the scene camera's field of view. In this sense, a uniquely colored mailbox captured attention, even when participants did not have a top-down set for that singleton, as predicted by lab-based data [e.g., Bacon and Egeth 1994]. However, it did not have an effect on the overall search time. This is likely because, unlike lab search, most of the task was spent moving around the mailroom, during which time the target was not within the visual field and accordingly its visual saliency could not have an impact.

Very little is known about how people coordinate their body, head, eyes and covert attention during search in a natural and complex environment, even though it is one we experience every day. The present experiment demonstrates that principles of visual search from the lab can be studied using mobile eye-tracking, and most importantly, both top-down instructions, and to bottom-up stimulus saliency, have clear and demonstrable effects on human eye movement behaviour. These results echo findings from the lab, which is an interesting confirmation of previous research. There were, however, surprises in how these principles manifested themselves in behaviour. For instance, instruction did not impact the use of any available information in the environment but rather it was selective to room numbers; and stimulus salience did not affect search time in the mailroom or the number of fixations on the target before it was grasped, but rather the probability that the target would be detected and fixated once it fell within the line of sight.

Each of the many thousands of previous lab-based visual search studies were mute on if, and how, top-down and bottom-up processes would impact gaze behaviour within a complex natural environment. We have discovered that these processes do indeed impact performance, but they do so in a remarkably selective manner. Determining the principles and boundary conditions that guide these selective influences will be an exciting and worthwhile enterprise for experimental and applied researchers alike. The present study therefore highlights these selective influences on active behavior as valuable empirical and theoretical issues for further investigation.

References

BACON, W. F., AND EGETH, H. E. 1994. Overriding stimulus-driven attentional capture. *Perception & Psychophysics, 55,* 485–496.

FOULSHAM, T., AND UNDERWOOD, G. 2007. How does the purpose of inspection influence the potency of visual saliency in scene perception? *Perception, 36*, 1123-1138.

HAYHOE, M. M., AND BALLARD, D. 2005. Eye movements in natural behavior. *Trends In Cognitive Sciences*, 9(4), 188-194.

ITTI, L., AND KOCH, C. 2001. Computational modelling of visual attention. *Nature Reviews Neuroscience*, *2*, 194-203.

LAND, M. F., MENNIE, N., AND RUSTED, J. 1999. The roles of vision and eye movements in the control of activities of daily living. *Perception, 28*(11), 1311-1328.

WOLFE, J. M. 1998. What can 1 million trials tell us about visual search? *Psychological Science, 9*(1), 33-39.