Vision Research 51 (2011) 1920–1931

Contents lists available at ScienceDirect

Vision Research

journal homepage: www.elsevier.com/locate/visres

The where, what and when of gaze allocation in the lab and the natural environment

Tom Foulsham[a,*], Esther Walker[b], Alan Kingstone[c]

[a] *Department of Psychology, University of Essex, UK*
[b] *Department of Cognitive Science, University of California, San Diego, CA, USA*
[c] *Department of Psychology, University of British Columbia, Canada*

ARTICLE INFO

Article history:
Received 23 December 2010
Received in revised form 5 June 2011
Available online 23 July 2011

Keywords:
Eye movements
Scene perception
Active vision

ABSTRACT

How do people distribute their visual attention in the natural environment? We and our colleagues have usually addressed this question by showing pictures, photographs or videos of natural scenes under controlled conditions and recording participants' eye movements as they view them. In the present study, we investigated whether people distribute their gaze in the same way when they are immersed and moving in the world compared to when they view video clips taken from the perspective of a walker. Participants wore a mobile eye tracker while walking to buy a coffee, a trip that required a short walk outdoors through the university campus. They subsequently watched first-person videos of the walk in the lab. Our results focused on *where* people directed their eyes and their head, *what* objects were gazed at and *when* attention-grabbing items were selected. Eye movements were more centralised in the real world, and locations around the horizon were selected with head movements. Other pedestrians, the path, and objects in the distance were looked at often in both the lab and the real world. However, there were some subtle differences in how and when these items were selected. For example, pedestrians close to the walker were fixated more often when viewed on video than in the real world. These results provide a crucial test of the relationship between real behaviour and eye movements measured in the lab.

1. Introduction

Humans, along with most vertebrates, sample the visual world by making eye movements to direct a centralised region of visual acuity towards different parts of the environment (Land & Fernald, 1992). In natural vision, these saccadic eye movements are made around three or four times every second, and research in humans has sought to identify the spatiotemporal properties of the way that people move their eyes (e.g., Becker, 1991), as well as how eye movements reflect information processing in stereotyped tasks such as reading (Rayner, 1998). More recently, a large body of research has considered where people look in realistic scenes, and how this might be determined by properties of the image or the task being performed. However, the vast majority of these experiments measure the eye movements of observers presented with drawings (De Graef, Christiaens, & d'Ydewalle, 1990; Yarbus, 1967), photographs (Foulsham & Underwood, 2007, 2008; Henderson, 2003), or occasionally videos (Foulsham, Cheng, Tracy, Henrich, & Kingstone, 2010; Itti, 2005) of the real world. In the present study we compared the eye movements of people immersed in the real environment to fixations and saccades recorded in the laboratory while watching videos of the same environment. We asked whether people would distribute their gaze in the same way and look at the same things, at the same time, when freely acting in the world, compared to watching videos of the world in the laboratory.

There are several reasons why comparing eye movements between the real world and a video representation is important. First, there are many differences between the situations in which people make saccades in everyday experience and the constraints imposed by a typical laboratory set up. In the real world, eye movements are part of a coordinated gaze system involving the head and the body as we move around space and perform actions (Pelz, Hayhoe, & Loeber, 2001). In contrast, due to the set-up required by most eye-trackers, experiments in the laboratory are most often performed with participants who are seated, stationary and have their head and body restrained. The displays used to present pictures and videos of the world are also normally smaller than the full visual angle available to people in the real world, and obviously the participants know that they are watching a representation of the world, and not the world itself. Although much research suggests that motion has strong effects on attention and eye movements (Hillstrom & Yantis, 1994; Itti, 2005), most of what we know about how people look at natural scenes is based on the "free viewing" of static images presented one at a time for a few seconds. While we have undoubtedly uncovered much about how people respond to such stimuli, it is crucial to consider whether this research generalizes to the way that people behave when immersed in the real world.

* Corresponding author. Address: Department of Psychology, Wivenhoe Park, Colchester, Essex CO4 3SQ, UK.
E-mail address: foulsham@essex.ac.uk (T. Foulsham).

0042-6989/$ - see front matter © 2011 Elsevier Ltd. All rights reserved.
doi:10.1016/j.visres.2011.07.002

A second reason to compare real and video eye movements is that there is something of a disconnection between research investigating eye movements during real-world actions and that performed on natural scene viewing. There have now been several studies looking at how people coordinate their gaze while performing sequences of skilled action. For example, Hayhoe, Shrivastava, Mruczek, and Pelz (2003) recorded participants making sandwiches and pouring a drink. This study found that fixations were task specific and temporally coordinated: people looked at relevant items (such as the end of a knife) "just-in-time" for the completion of the next action (such as transferring jelly to bread). Whatever is guiding these eye movements, Hayhoe et al. argue, it is sensitive to highly specific and subtle details about the task (such as the fact that jelly needs more supervision as it is transferred than peanut butter, because peanut butter is more sticky; see also Ballard & Hayhoe, 2009). In other active tasks such as playing sports and driving, gaze is both highly specific to the task and planned in concert with a temporal sequence of motor actions such as hand, head or body movements (Land, 2009). In an everyday task that is more relevant to the study at hand, walking a path, Patla and Vickers (2003) reported that people fixate the points at which they will step approximately a second before stepping there, a pattern of gaze preceding action that is common to other real-world tasks.

In contrast to this research on *when* gaze is deployed during skilled actions, studies of eye movements during natural scene viewing have been largely concerned with accounting for patterns in *where* people fixate. For example, observers may be drawn to fixate objects that are out of place given their context (Loftus & Mackworth, 1978; Underwood & Foulsham, 2006) or items which convey social information, such as the faces of people in a scene (Birmingham, Bischof, & Kingstone, 2008; Yarbus, 1967). However, these findings have rarely been tested in the context of real world behaviour. One particular debate in natural scene viewing concerns the extent to which fixations are determined by the bottom-up visual saliency of image features (Itti & Koch, 2000; Parkhurst, Law, & Niebur, 2002), a hypothesis which has been refuted by many recent studies showing how participants can avoid fixating salient points if they are not task relevant (Einhauser, Rutishauser, & Koch, 2008; Foulsham & Underwood, 2007). Indeed, this is one area in which virtual reality studies have helped to bridge the gap between laboratory experiments and the real world. For example, Rothkopf, Ballard, and Hayhoe (2007) recorded eye and head movements while participants walked in a virtual reality environment collecting or avoiding blocks of a certain colour. Gaze was attuned to the demands of the task (i.e. to the task relevance of the different blocks), irrespective of their visual saliency. In general, however, the questions asked, and the findings uncovered, in studies of gaze during real-world actions and those of eye movements in scene viewing are not well matched (Kingstone, Smilek, & Eastwood, 2008). One of our aims in the present study is to address questions from scene perception, such as which items are frequently fixated, in the real world vs. a video-based setting.

A third and final reason why comparing real and laboratory eye movements is important is that, when the two have been explicitly compared, several differences have been found. In terms of the spatiotemporal dynamics of saccades and fixations, saccades tend to be somewhat larger in unconstrained action than in the lab (Bahill, Adler, & Stark, 1975). Stahl (1999) reports that these large gaze shifts are made with head movements in unconstrained tasks, something which has not been captured in natural scene viewing experiments. Hayhoe et al. (2003) report that fixations in their sandwich and drink making task have a wider range of durations, with frequent fixations that are shorter (100–200 ms) than is common in static scene viewing, as well as some very long fixations. In a recent study, 't Hart et al. (2010) reported what we believe is the only explicit comparison between eye movements in unconstrained, free exploration and free viewing on a monitor. In that study, head-centred videos from a mobile eye tracker (the "EyeSeeCam") were presented to participants in a laboratory situation. The real-world gaze data came from six individuals who were told to behave naturally in a range of different environments, and the results showed that some of the spatial characteristics of gaze in the laboratory matched those of the real world. Participants showed a central bias, previously documented in scenes (Foulsham & Underwood, 2008; Tatler, 2007), and this was most pronounced when isolated frames were presented randomly. Saliency was a weak but reliable predictor of where people looked, particularly in videos shown in laboratory conditions. Importantly, where people looked in video was a better predictor of gaze in the real world than eye movements in static scenes. On the other hand, even in this case, the eye movement distributions from the lab are only about 60% accurate at predicting gaze in the real world, which indicates that there are considerable differences between the two situations.

Despite the limited research comparing real-world to laboratory tasks, it is not known if what people look at, and when they look there, is different in a laboratory set up vs. when people are immersed in the real world. In this paper we compare gaze recorded from participants using a mobile eye tracker while they undertook a naturalistic, unconstrained task (a walk around campus) with eye movements recorded from participants in the lab watching first-person videos of the same task. This procedure resulted in a large amount of naturalistic eye movement data, and our analysis and results focused on three main areas of interest, based on previous research from scenes and movies. First, we consider what spatial biases emerge from the general allocation of gaze to different points in the visual field, and whether these biases are tethered to landmarks in the world. Cristino and Baddeley (2009) found that fixations made by observers watching a video of a walk down a street were best explained by taking into account the structure of the world, and in particular the position of the horizon. We therefore investigate whether our observers showed consistent biases relative to egocentric head direction or allocentric horizon position, as well as whether these biases were different in video vs. the real world.

Second, we categorise what objects in the scene are fixated, and we consider whether the frequency with which some objects were fixated differs in the lab and the real world. For example, previous research suggests that people should look at the path when walking in laboratory conditions (Patla & Vickers, 2003) and that they often fixate other people in a static scene (Birmingham et al., 2008). In several studies, Jovancevic, Sullivan, and Hayhoe (2006) and Jovancevic-Misic and Hayhoe (2009) have shown that the gaze of participants walking in a virtual environment is very sensitive to virtual pedestrians, especially when they are likely to cause a collision. When a similar experiment was repeated in the real world, the results continued to show that participants looked at pedestrians according to the probability of them veering off course. In a related, virtual, study, Karacan and Hayhoe (2008) reported that participants spend most of the time looking at the ground and the environment, and relatively less time looking at pedestrians (although this increased with experience with the scene). A specific aim of the present research, therefore, is to quantify gaze on the path and people in both real, unconstrained behaviour and when watching a scene in the lab.

Finally, we evaluated the temporal fit between fixations on the different objects in the scene in the real world and the lab. Are objects looked at the same time in both cases? Jovancevic et al. (2006) report that most pedestrians in their virtual environment were fixated early and from a distance, within 1–2 s of them appearing in the scene. In studies with static stimuli, other people continue to be looked at over time (Birmingham et al., 2008). In the present study we are able to ask whether real people in the environment, or in a video, are looked at early and in the same temporal fashion.

Together, these analyses aim at describing the allocation of gaze in both natural behaviour (walking around campus) and "naturalistic" viewing (constrained in a laboratory).

2. Method

2.1. Participants

Our eye movement data set comes from fourteen participants (nine female) who were recruited from the student population at UBC and took part in exchange for course credit. All participants reported normal vision and none wore glasses. Participants were included on the proviso that they achieved a good calibration with the mobile eye tracker and that their real-world data had minimal data loss. The same participants subsequently took part in the laboratory part of the experiment.

2.2. Apparatus and stimuli

In the first part of the study, we recorded fully-mobile gaze direction using the ASL MobileEye (Applied Science Laboratories; Bedford, MA, USA; see Fig. 1, left). This eye tracker is mounted on a pair of lightweight, inconspicuous goggles, and features two cameras: a scene camera and a camera capturing the pupil image. The scene camera was aligned with the participant's line-of-sight and captured a video image of the visual field with a size of approximately 50° horizontally and 40° vertically. This camera included a microphone recording environmental sounds. The eye camera monitored the position of the pupil and the corneal reflection from the right eye. Video frames from each camera were written at 60 Hz to a digital video recorder that was carried by the participant in a backpack. Eye and scene camera frames were interleaved, resulting in a sampling rate of 30 Hz. The system has an instrumental spatial accuracy of 0.5° and in our own tests we found that gaze position was accurate to within 1°, although pupil data was sometimes lost due to changes in illumination when walking outdoors.

Videos recorded by the scene camera in the first part of the study were used as stimuli for the second part, where they were presented in the laboratory. Videos were presented at the recorded resolution (640 × 480 pixels, 30 fps), centred on a grey background on a 19-in. monitor. Participants used a chin-rest, constraining head movements and ensuring that there was a constant distance of 60 cm from the screen, meaning that the video was 25° by 19° of visual angle. It is important to note the video was not life-size and in fact took up only about half the visual angle of the same image in real life. This is one of the major simplifications typically imposed in scene perception research in the laboratory, with most studies making the assumption that absolute size will not make a difference to gaze behaviour.

Sound from the videos was also played through speakers that flanked the computer monitor. Eye movements in the laboratory were tracked using the EyeLink II system, which is also a head-mounted eye tracker. Eye position was recorded monocularly from the pupil image at 500 Hz. The EyeLink system used an on-line parser to extract fixations and saccades from the eye position samples, using velocity (30°/s) and acceleration (8000°/s^2) thresholds. A message was also written to the data file to mark the time of onset of each video frame with millisecond accuracy. This allowed eye position data to be synchronized with events in the movie with a high degree of precision.

2.3. Procedure

2.3.1. Session 1

In the first session, participants completed the mobile eye-tracking portion of the experiment. The eye tracker was fitted to the participant and secured with a strap, and a calibration was performed, using ASL's EyeVision software, by getting participants to look at each of nine points marked on the wall of the laboratory while standing at a distance of 2–3 m. This calibration was repeated at the end of session, allowing us to ensure that calibration had not slipped significantly during the session. In addition, several calibrations were conducted outside at the beginning of the walk, at the midway point, and at the end. Outdoor calibrations were recorded on tape and comprised the participant being instructed to look at a series of objects in the environment. The record of gaze location could later be compared to the recorded instructions, and the system could be recalibrated offline, ensuring that the system was correctly tracking gaze when outdoors.

After calibration, participants were given three Canadian dollars and instructed to walk to the students' union building, taking a route of their choice, purchase a coffee or a snack, and walk back. These instructions were chosen to be as unconstrained and natural as possible, whilst still ensuring that all participants would walk in a similar environment. There were several possible alternative routes, but all involved 5–10 min of walking in a pedestrian environment featuring sidewalks, buildings, cars and street furniture,

Fig. 1. Left: The ASL MobileEye tracker used to record gaze during walking. Right: The route walked by one participant, as drawn by them on a satellite map of the surrounding campus.

and other pedestrians. Fig. 1 (right) shows a map of the surrounding area, and examples from the scene camera are shown in Figs. 2 and 4. During the walk, an experimenter followed the participant at a distance. On their return, participants drew their route on a map and were debriefed.

2.3.2. Session 2

The same participants returned a week later and participated in the laboratory section of the experiment. In this session, participants were seated in front of a monitor and the EyeLink II system was calibrated using a 9-dot grid. In the first part of the session, participants watched a series of short clips taken from their own and other people's walks, with the task being to recognize whether the clip came from their own walk or somebody else's. The results from this part of the study, and from pilot studies, suggested that people were able to do this easily on the basis of the route that was taken, the weather conditions on the day, and possibly differences in gait. Because participants were re-viewing situations from their own experience, this part of the study risked introducing memory biases and anticipation into the eye tracking results. Therefore, in the remainder of the session we matched each participant with another from Session 1. This meant that we collected data on each route being viewed by one naïve observer in the real world and a different naïve observer in the lab. The routes were perfectly matched by having each participant in the lab view the walk of another person that had occurred in real life.

While watching the matched participant's walk, observers were asked to imagine that they were walking the route outside and behave as they normally would. In order to maintain accurate eye movement tracking, the videos were split into 3-min sections, allowing a pause between each section and a recalibration if required.

3. Analysis and results

The raw data from Session 1 consisted of a frame-by-frame record of eye position, within the head-centred video image that was specific to each walking subject. The ASL mobile eye also provided a digital video file with gaze cursor showing eye position overlaid onto the scene. In Session 2, the EyeLink system recorded eye position samples at 500 Hz, along with time-stamped messages indicating the exact time that a video frame was displayed, and allowing us to check for dropped frames (these turned out to be very rare). These data allowed a video showing gaze at each frame to be produced, using custom code written in MATLAB. Thus position coordinates for each frame were available for both sessions, as well as a video showing gaze location superimposed over the scene.

In order to address the question of how similar or different gaze behaviour was in each session, we performed data analyses on both the pixel-based gaze coordinates ("where" the subject was looking), and on the objects at fixation ("what" was being looked at). Because each participant's walk, and therefore the scene in front of them, was different, knowing what the person was looking at required manual coding. This method will be discussed in more detail below. Analysis was restricted to three 30-s clips taken from each participant's walk. These clips were chosen on the basis of several criteria. First, they were representative of the environment and events found during the whole walk. Second, at least one clip was taken from the outbound trip, and at least one was taken from the return trip. Finally, accurate gaze data was available for a

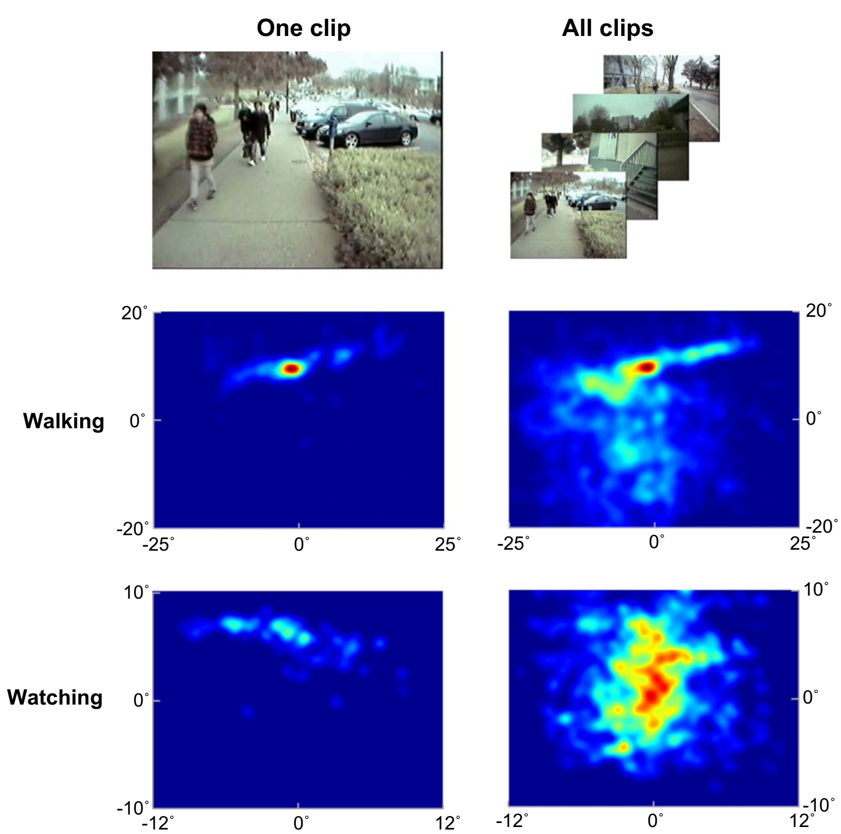

Fig. 2. Eye position distributions in one clip (left column) and across all clips (right column). Distributions are shown as heatmaps for the walking and watching conditions, with bright areas representing positions where the eye dwelt for longer, within the head-centred field of view. *x*- and *y*-coordinates indicate the visual angle, relative to the centre of the scene camera image (when walking) or the video display (when watching).

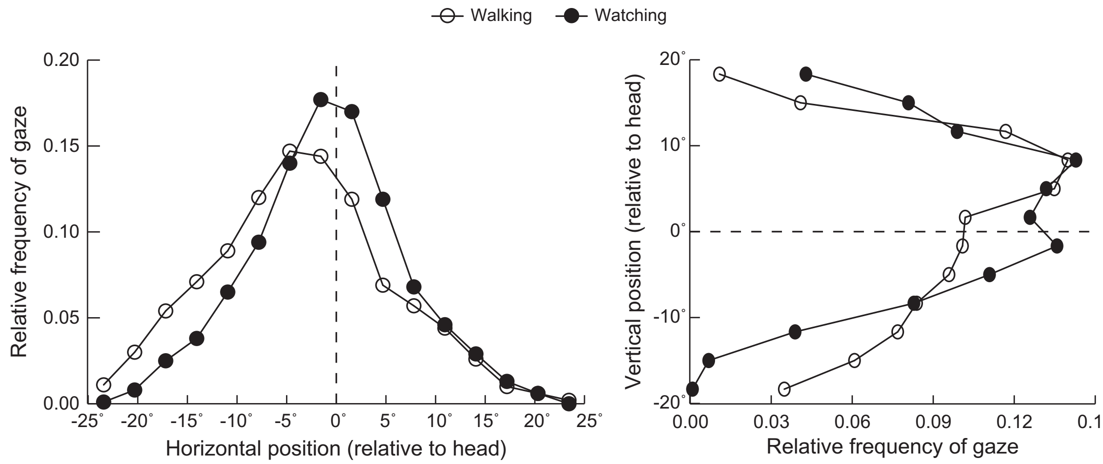

Fig. 3. Frequency distributions of gaze within the head camera view. Plots show the relative frequency (across all clips) of frames where the eyes were positioned at different points horizontally (left) and vertically (right). Position axes are scaled according to the visual angle in the real-world condition, with the centre indicated by the dashed line.

subject walking this section and a different subject viewing the walk in Session 2. We will begin by characterizing the spatial distribution of gaze within the scene.

3.1. Where do people look?

3.1.1. Spatial gaze distributions

How did people distribute their gaze within the head-centred field-of-view? In order to describe and compare the overall spatial distribution of gaze, we computed gaze density maps by overlaying a two-dimensional Gaussian at the gaze position for each frame, and progressively summing the resulting map. This process produced a gaze "landscape" for each clip, and the height of each map was normalized by the number of gaze samples available. The Gaussian had a sigma of 12 pixels, which was equivalent to approximately 1° in the real world. This value was chosen based on estimates of the size of the fovea and on the error in the position signal of both eye trackers.

Fig. 2 shows gaze distributions for one clip, and for the average of all clips taken from Session 1 (walking) and Session 2 (watching). Several observations can be made from this representation of the overall spatial distributions. First, the distributions from both conditions were highly centralised along the horizontal. A central bias has been previously seen and discussed in scene perception in the lab, and orbital reserve, or the position of the eyes within the head, is one possible cause of this bias (but not the only one: Tatler, 2007; Tseng, Carmi, Cameron, Munoz, & Itti, 2009). In our case, we found that in the real world the eyes tended to be positioned centrally, along the horizontal, relative to head direction. This was also the case in the lab, where head position was restricted.

Second, gaze in the watching condition was more distributed, despite the information in central vision being the same. When walking, participants spent most of the time fixating in one location above the centre of the head frame-of-reference; large eye movements to peripheral locations were rare. Fig. 3 plots the frequency distribution of gaze along the horizontal and vertical axes in the two conditions. These distributions are plotted relative to the field-of-view in the real-world condition. Non-parametric, two-sampled K–S tests indicated that walking and watching distributions were different, in both the x- and y-axis (both $ps < .001$). In the horizontal axis, gaze was approximately normally distributed around the centre, particularly in the laboratory condition. In the vertical axis, gaze was more frequently directed at regions in the

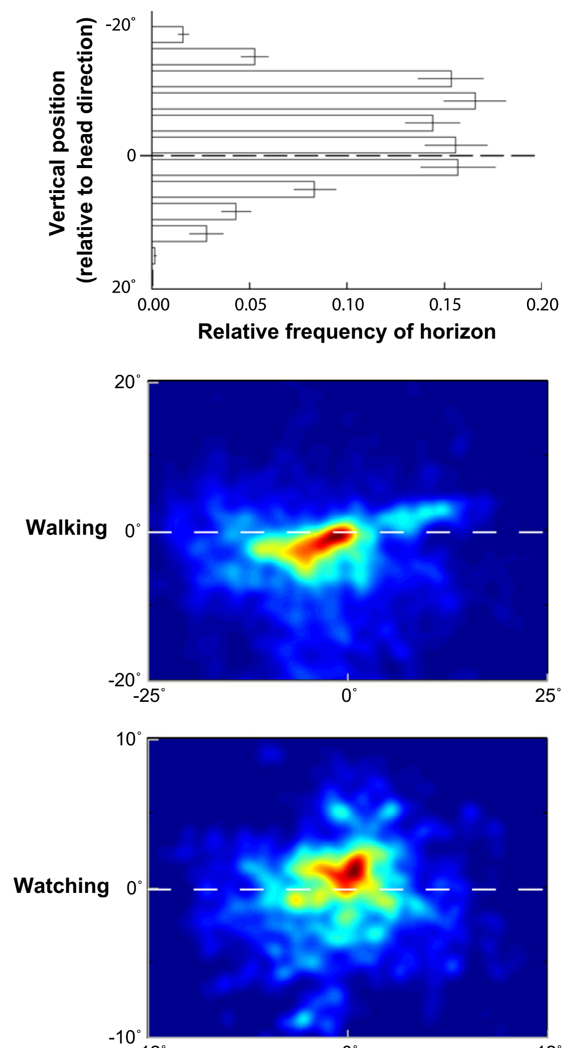

Fig. 4. Tethering gaze distributions to the position of the horizon. The top panel depicts the distribution of the horizon along the vertical axis within the real-world field of view, with bars representing the mean (with standard error bars) across clips. The centre of the head frame is shown for reference (dashed line). Bottom panels show world-centred eye position distributions, plotted in the same way as Fig. 2 but relative to the position of the horizon at each point in time (white dashed lines).

top half of the field-of-view, peaking at about 10° above the centre of the head direction while walking. In the watching condition, the point of gaze was more evenly distributed, and somewhat more centralised along the vertical. To test this difference, we compared the standard deviation of the *x*- and *y*-coordinates of eye position between walking and watching. A higher standard deviation in one condition would indicate that eye movements in that condition were more spread out. Testing across clips, there was no difference between conditions in the standard deviation in the horizontal direction (average SD walking = 7.6°, SD watching = 7.4°, $t(13) < 1$), but there was a significant difference in the standard deviation in the *y*-coordinate ($t(13) = 2.5$, $p = .03$). Eye position was more spread out along the vertical axis in the watching condition (6.8°) than when participants were walking in the real world (5.3°). It is clear from the gaze distributions that the modal eye position along the vertical axis is also different in the two experiments. Specifically, fixation in the walking condition was highly likely to be directed toward a "heading point" above the centre of the head frame-of-reference, while gaze in the lab was more likely to be positioned around the horizontal midline.

3.1.2. "World salience": head direction, eye direction and the horizon

The previous section considered eye position regardless of the environment presented to the participant at any point in time. In order to tether the head-centred gaze record to the scene, we analysed the position of the horizon at each point in the clips. In scene perception in the lab, the horizon may be an important factor in determining where people initially orient their attention (Foulsham & Kingstone, 2010). In dynamic scenes, Cristino and Baddeley (2009) found that the position of the horizon was a strong predictor of fixation density. This indicated that locations were selected in a world- or scene-centred frame of reference and buttressed their claim that "world salience"—the meaning of items in their real world context—was a more important factor in guiding the eyes than image salience. Importantly, that study used realistic, first-person video clips, but these were viewed by participants constrained in a laboratory eye tracker. In the present study we were able to test the generalisability of the world-centred reference frame, in both a real walker and somebody watching a video. Thus in this section we consider eye position relative to the layout of the world.

To approximate the layout of the world in front of the participant, a research assistant coded the vertical location of the horizon in each frame. This coding was performed using custom programmed software that allowed the observer to place a horizontal line at the point where the horizon was located, and to move the cursor outside the frame when the horizon was missing from the frame (e.g. if the participant looked at their own feet). Horizon position coding was performed for all clips from the scene camera, without reference to the eye position data.

The horizon analysis provided two insights. First, because the image frame is head-centred, the position of the horizon within this frame gives us a proxy measure of head direction in the vertical axis. For example, if the head were always pointed at the horizon, the horizon will always be positioned in the centre of the video frame. Second, we can re-evaluate the eye position distributions from walking and watching in terms of their relationship to the scene frame of reference. As we have seen, the eyes tended to dwell in the centre of the field of view, but were these fixations planned egocentrically, or might they be explained by the structure of the scene? Fig. 4 shows a histogram of the vertical position of the horizon (top panel) alongside plots of eye position relative to the horizon (bottom panels). The latter plots were produced by shifting the *y*-coordinate of the eye position for each frame according to where the horizon was located in that frame. Moments when the horizon was missing were excluded from this analysis.

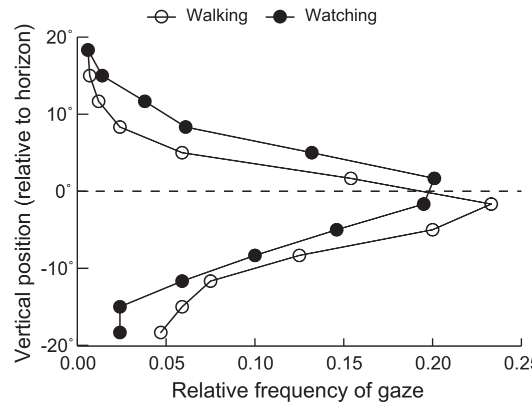

Fig. 5. Frequency distribution of gaze in the vertical axis, plotted in the same way as Fig. 3 but relative to vertical position of the horizon.

Rather than being always positioned exactly in the centre of the frame, the horizon tended to be in the upper half of the visual field. When the horizon was visible, it was above the centre of the head frame in 69% of all frames. To express this in a different way, participants walking in our clips tended to point their heads below the horizon. Furthermore, despite the fact that this sampling of the scene by head direction was also imposed on participants watching the clips in the lab, the plots of eye position relative to the horizon (Fig. 4, bottom) reveal that differences in eye position between the two conditions persisted. Fig. 5 shows the horizon-shifted frequency distributions along the vertical axis across all clips (note that the horizontal distribution remained the same as that in Fig. 3). While walking, participants tended to fixate slightly below the horizon, with the peak of the gaze distribution being approximately 0.5° below the horizon. However, when watching video in the lab, participants tended to look above the horizon more often (gaze distribution peak at 1° above the horizon). Again the K–S test confirmed that the two distributions were reliably different ($p < .001$). The difference in mean vertical eye position approached significance ($t(13) = 2.1$, $p = .06$). Across clips the mean vertical eye position relative to the horizon was lower in walking (M = 1.9° below the horizon) than in watching (M = 0.6° below the horizon).

It is notable that both the central tendency and the spread of the horizon-shifted distributions are different from the average head-centred ones in Fig. 3. Specifically, while gaze tended to be above the centre of the head direction, it was systematically *below* the horizon. The head- and horizon-centred vertical distributions were different, both in walking and when video watching (K–S tests, both $ps < .001$). The distributions were also more normal and slightly less spread out when aligned with the horizon, although this was not reliable across the limited number of clips (shifted SD walking = 5.2°, watching = 6.4°; both $ts < 1$). While this merits further investigation, the differences support the idea that the position of the horizon (and probably of the world in general) accounted for much of the variance in where people were looking.

3.2. What do people look at?

3.2.1. Coding gaze to objects of interest

Our analyses thus far have been confined to the head- and scene-centred location of fixations rather than the objects at which these fixations are targeted. In order to determine what people were looking at we manually coded the object at the point of gaze for each frame in each of the clips. Coding was accomplished using purpose built MATLAB software written by the first author (and available at http://barlab.psych.ubc.ca/people/tom/programming). This software displayed each video one frame at a time and

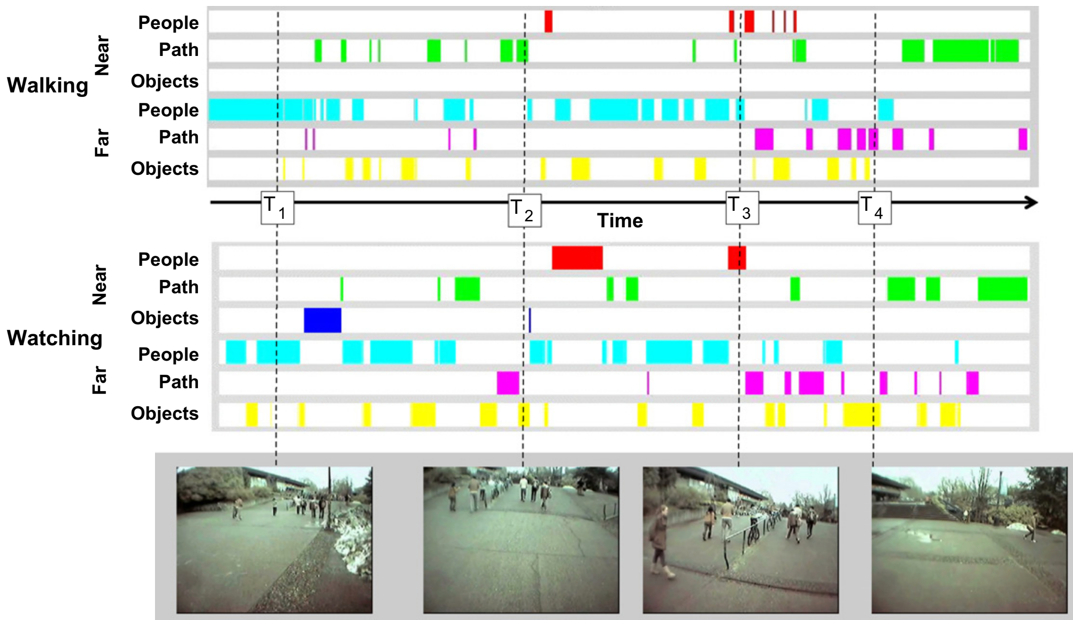

Fig. 6. Gaze allocation in a participant walking (top panel) and somebody watching video of the same walk in the lab (middle panel), coded over time (horizontal axis) for one 30 s clip. Timing matrices classify gaze as being on one of six different categories (labelled on the left axes). Bottom panel shows the scene in front of the participant at four example time points (T_1–T_4, dotted lines). Categories were mutually exclusive and the missing and miscellaneous categories are not shown.

allowed the user to categorise the object being looked at by pressing one of several keys on the keyboard. Clips were generated in exactly the same way in both walking and watching conditions, with point of gaze illustrated with a circular cursor of diameter 25 pixels, equivalent to about 2° of visual angle in the real world. All clips were coded by at least two independent research assistants, one of whom was naïve as to whether any individual clip came from the real-world or lab condition. This enabled us to exclude possible experimenter biases and check on the between-coder reliability (which was high in all cases: all correlations between codings >.90).

We defined six objects of interest in the clips: people, the path and other objects, all of which were defined as either near or far from the person walking. These categories were chosen based both on prior research and on our impressions of what objects were often inspected. A significant body of research has indicated that observers are biased to attend to and fixate other people when they appear in search displays or natural scenes, and this is believed to reflect a hard-wired human tendency for social attention (Emery, 2000; Kingstone, 2009). However, few if any studies have examined social attention using eye tracking in the real world. Gaze was counted as being on another person if the gaze cursor intersected any part of the body of one of the pedestrians in the clip (meaning that eye position was within 1° of the person). A fixation on the path was logged when gaze intersected the ground ahead of the walker, which consisted of grey pavement for the majority of the route. Prior inspection of the clips suggested that people often fixated the path, and Cristino and Baddeley (2009) anecdotally observed that their lab-based participants looked at the ground frequently, despite it being of low contrast and brightness. Several studies have looked at eye movements during walking in the lab and found that looks to the ground are important, and here we sought to investigate this in a naturalistic setting (Patla & Vickers, 2003). The "other objects" category was defined as gaze on discernible objects (other than people or the path). For example in our clips these objects included lampposts, cars and trees.

Each of these categories was subdivided into whether they were "near" or "far" in relation to the walker. Coders were instructed to make their own judgements about what was near and far. No accurate physical measurements were available in practice but coders reported that distance could be inferred from the position of the object in the frame, with in most cases a near object being in the lower half of the video frame. A final category designated when gaze was on a different part of the scene (such as the sky) or when eye position data was missing due to blinks or tracking loss.

Fig. 6 illustrates the resulting coding matrix for one clip, showing the location of gaze at each point in time, both for the participant walking the route in real life, and for a different participant watching the clip in the lab. Gaze regularly moved between objects in the field of view. In the example in Fig. 6, the participant walking in the scene spends the first few seconds looking at people in the distance (e.g. at T_1) and she also makes multiple short glances toward the near path (e.g. at T_2). When a pedestrian crosses her path (e.g. T_3), she makes some brief fixations on this near person, before moving back to people and the path. Towards the end of the clip she often looks at the path and the steps in the distance (e.g. T_4), which she will have to climb in a few minutes time. The coding of gaze over time provided a very detailed description of where people were looking, and our subsequent analysis looks at the similarity between this description for walkers and watchers.

3.2.2. Comparing gaze across conditions

In order to evaluate whether different objects were looked at in the real world and in the lab we summed the time spent on each of our six categories within each clip, and then expressed it as a proportion of the total clip duration. It is important to note that here we were interested only in comparing between conditions, and not the difference in frequency of gazes to different regions of interest, because the latter comparison would depend on the incidence of each type of object in the videos (and therefore it's availability for being fixated). We treated each clip as an independent data point and compared gaze on each type of object between someone walking the route and someone watching exactly the same clip, using planned, related t-tests.

Fig. 7 shows the results of this analysis. In many cases the frequency of gazes on each type of object was similar in both

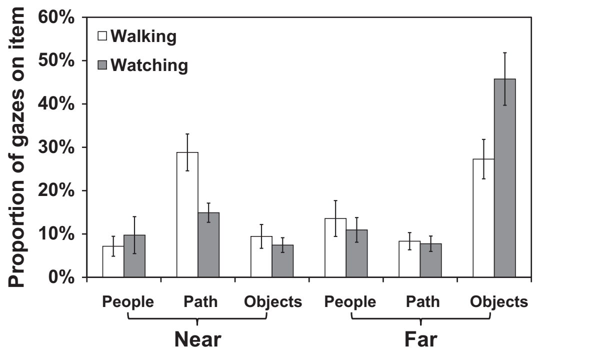

Fig. 7. The proportion of frames participants walking and watching spent gazing at different items in their environment. Graph plots the mean (with standard error bars) across clips.

conditions. However, people walking in the real world spent significantly more time looking at the near path than participants watching videos in the lab ($t(13) = 2.9$, $p = .01$). Conversely, participants in the lab spent more time looking at objects that were far away than people in real life ($t(13) = 2.7$, $p = .02$). The frequency of gazes on each type of object did not differ significantly in any of the other categories (all $ts < 1$). To summarise this analysis then, while people walking spent more time gazing at the path close to them, people watching the clips in the lab tended to spend this time looking instead at objects further away.

3.3. When do people look

3.3.1. Comparing gaze over time

The example record of gaze over time in Fig. 6 gives a very detailed picture of the target of gaze at different points in time. A further question, therefore, concerns whether there was a close temporal fit between gazes in the real world and in the lab. Informal inspection of Fig. 6 suggests that in that clip there was some correspondence between gaze in the two conditions: sometimes both the person walking and the participant watching in the lab looked at the same type of object at the same time. For example, both participants start the clip looking at a pedestrian in the distance (see T_1, Fig. 4). To quantify this we calculated the proportion of frames where the two gaze records converged, and we can compare this to a control comparison between two gaze records from different clips. This control comparison will allow for the fact that, given certain categories of object were gazed at more often than others, we would expect some degree of overlap by chance.

Surprisingly, gaze in the two conditions converged on only 22% of frames (mean across clips, SEM = 3%). Although this is greater than we would expect if gaze were randomly assigned to one of the six categories (where chance overlap would be 1/6 or 16.7%), it is only marginally higher than the control comparison between gaze records for different clips (which showed a mean overlap of 20% of all frames). Indeed there was no statistically significant difference between the observed overlap and the control comparisons ($t < 1$). In other words, although both walkers and watchers looked at some types of objects for a similar amount of time across all clips (see previous section), these objects were not fixated at the same moment in time.

3.3.2. When do people look at people?

Previous research suggested one case where the timing of gazes might differ between the lab and the real world: when observers are looking at other people. We were particularly interested in how the other pedestrians in the clips were looked at, because of the large number of laboratory studies documenting an attentional bias towards people, their faces and their eyes (Birmingham et al., 2008; Foulsham et al., 2010; Yarbus, 1967), and because observational studies of interpersonal behaviour have found that people avoid looking at others in some social settings (e.g., Goffman, 1963). Jovancevic-Misic and Hayhoe (2009) found that participants looked at real pedestrians when walking around a set route, although in this experiment the pedestrians were also explicitly important for the task (participants had to follow a leader and avoid collisions). We have experimental evidence that people are less likely to look at people who are physically present than at those who are on a video screen (Laidlaw, Foulsham, Kuhn, & Kingstone, 2011). However, no other experiments, to our knowledge, have used eye trackers to measure gaze towards people in real situations where the observers are free to move around.

In order to measure how likely pedestrians in our experiment were to be fixated, we supplemented our coding of the clips in the following way. First, we counted how many unique pedestrians there were in the scene and when they entered and exited the field of view. Then, we recorded the time of each discrete gaze made towards these pedestrians.

There were 133 different pedestrians across all the clips and the walker in the scene fixated 83% of these people at least once. In the lab, 72% of the people were looked at. Fig. 8 (left panel) shows the cumulative probability of a pedestrian being looked at, and together these data confirm that the other people in the scene were potent attractors of attention. We also observed a subtle difference between the conditions arising in gazes toward pedestrians that

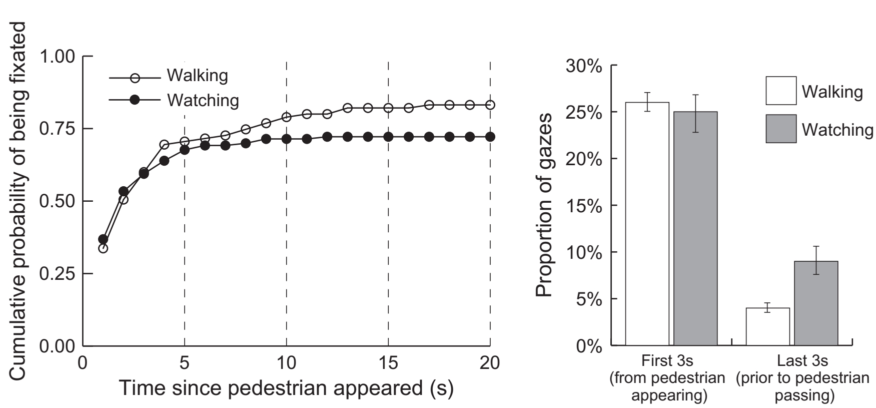

Fig. 8. The time course of looks toward other pedestrians. Left panel shows the cumulative probability of a pedestrian being looked at least once, as a function of the time since that pedestrian first entered the field of view. Right panel shows the proportion of discrete gazes that were directed towards pedestrians walking towards the observer, in the first or last 3 s in which they appeared in the scene. Bars show the mean plus standard errors across clips.

were close to the participant. In Fig. 7 one can see that, while people far away from the observer were more likely to be fixated when that observer was in the real world than in the lab, the opposite pattern was found in pedestrians close to the walker. We confirmed this by counting gazes directed at pedestrians who were walking towards the walker and crossed their path. While in general pedestrians were slightly more likely to be looked at in the real-world condition, this was not the case when looking at the last few seconds before a pedestrian passed the observer (Fig. 8, right panel). Most gazes towards these pedestrians occurred when they first appeared in the field of view, typically several metres away from the observer. During the first 3 s in which they were in the scene there was no difference between the proportion of pedestrian gazes in walking and watching ($t(13) < 1$). However, during the last 3 s before passing the approaching pedestrian, there were significantly more pedestrian gazes in participants watching a video ($t(13) = 2.6$, $p = .02$). Looks at the pedestrian when he or she was close to the observer were relatively rare, particularly when the observer was actually in the real world. Therefore one of the temporal differences between gaze in the real world and when watching video was that participants in the real world were less likely to look at pedestrians when they were close to them than people watching video of exactly the same event.

3.4. The role of head-induced motion

Our results have described some differences between the real-world walking condition and the case of watching the scene unfold in the laboratory. One of the differences between this study and previous laboratory research is that the clips viewed in our video condition came directly from the head-mounted scene camera in the walking condition and thus featured motion from head movements. This allowed a clearer comparison between conditions because the visual stimulus was very similar and permitted us to yoke each "watcher" to a "walker". t' Hart et al. (2009) also used head-centred videos for the laboratory component of their study. In contrast, previous studies looking at eye movements in video have used professionally shot footage or video taken from a fixed camera where head movements were minimized or excluded (Foulsham et al., 2010; Itti, 2005) Cristino and Baddeley (2009) used a customized head-mounted camera, but deliberately minimized head movements. Jerky camera movements can make video viewing uncomfortable, although the participants included in the present study reported feeling comfortable and immersed in the scene. Our own phenomenological experience converged with those reports. The fact that spatial gaze distributions and the things looked at in each condition were broadly similar suggests that camera motion elicited by head movements is unlikely to have had a large impact on the results. Nevertheless, an interesting question is whether the unpredictable motion from head movements may have caused some of the differences we observed. To address this question we performed two additional analyses.

First, we compared general eye movement statistics from participants in our watching condition to those from observers viewing fixed-camera videos in the same eye-tracking set-up. Participants watching the walking videos in the present study made an average of 61.6 fixations during the 30 s clips analysed here (i.e. around 2 fixations per second), with a mean fixation duration of 441 ms. This is within the range reported by other studies. On the other hand, fewer (and longer) fixations were made compared to the 25 participants in Foulsham et al. (2010) who watched fixed camera videos of people talking and made 2.5 fixations/s with a mean duration of 377 ms. Observers watching a professionally shot Hollywood movie (with the same eye tracker) also made more fixations which were shorter than those recorded here (Birmingham et al., 2006; mean fixation rate = 2.5/s; mean duration = 346 ms). Between-group *t*-tests confirmed these differences (all $ps < .05$). Thus the walking videos seemed to elicit fewer, longer fixations, and this may have partly been due to the head-movements. However, it certainly does not show that gaze allocation in the head-mounted videos is more random (as it might be if participants were uncomfortable). Instead, and because we found that participants walking in the real world also tended to stay fixated for longer periods, we suggest that the eye movement statistics reflect the fact that items within the field-of-view had already been selected by head movements from the walking group and thus watchers did not need to move their eyes as much as in other video presentations.

Second, we recoded the video data according to the presence of head movements. If movements from the head camera have a large effect on eye position during watching, we would expect the difference between walking and watching to be largest around the time of a large head movement (for example if video-watching participants were not expecting the head movement and took some time to re-orient themselves). We coded the onset time of any large head movements in each clip and calculated the Euclidian distance between eye positions of walkers and watchers. Eye positions were compared in each frame and the distance averaged across a 500 ms bin either side of the head movement. To generate a baseline distance, eye positions were compared at random times during each clip. Eye position distance at the time of a head movement was not significantly different from the baseline (in fact gazes were slightly closer during a head movement: mean distance in real-world degrees around time of head movement = 15.8°; random baseline = 17.8°; $t(13) = 1.3$, $p = .2$). Thus, although further research would be necessary to fully establish the effects of head translations and scene changes around the time of a head movement, this preliminary analysis suggests that head movements did not have a large impact on the difference between walking and watching.

4. Discussion

In everyday life, humans shift their gaze as part of a much larger behavioural repertoire. This repertoire often consists of moving around within a continuous three-dimensional environment featuring many different objects and inhabited by other people. In the present study we asked whether the allocation of gaze in this type of realistic situation is the same as that observed in laboratory experiments. This investigation is important because it permits an evaluation of the degree to which findings from past and present research into visual attention in scenes and movies are reflective of how humans behave in the real world. It might be that some (or all) examples of visual attention occur in the same way in both complex natural behaviour (e.g., when people are walking down a street) and in the controlled version of this behaviour that researchers often investigate in a lab (e.g. viewing a static image of the same street). We believe that comparisons between these levels must be made before such a conclusion can be reached (Kingstone et al., 2008).

We measured where people looked, when they looked there, and what people looked at for multiple participants walking across campus. We then compared the results to matched participants who watched the exact same walks while seated in the laboratory. Our discussion focuses on what is preserved between these situations and what is different.

4.1. Walking vs. watching: what is the same?

By presenting video of the same events to people in our lab condition we kept the contents of central vision as similar as possible:

the same objects were available to be fixated and the situation unfolded in the same way. We found some similarities between gaze in the real world and gaze in the lab.

First, both participants walking in the real world and those watching video on a screen spent much of the time looking at the centre of the visual field. A pronounced central bias has been reported previously in viewing of static scenes and videos on a monitor (Foulsham & Underwood, 2008; Itti, 2005; Tatler, Baddeley, & Gichrist, 2005). We replicated this result, and observed a similar bias in the distribution of eye position, within the head frame, shown by people free to move around the real world. That there was a central bias in the real world is probably a reflection of both orbital reserve—the tendency for the eyes to be positioned centrally in the orbits (Tseng et al., 2009)—and the fact that people oriented their head towards important features in the environment. These features (e.g. objects) would then have been more likely to be in the centre of the visual field, something that carried over into the head-centred video we presented to watching subjects in Session 2. One critical feature that we were able to track was the horizon, and we found that it was normally positioned slightly above the centre of the visual field, suggesting that people oriented their head to points *below* the horizon. By recording videos from multiple angles or framing these videos differently in the lab it would be possible to tease apart the influence of orbital reserve and the position of features in the frame. Cristino and Baddeley (2009) have also demonstrated that a change in perception of the gravitational vertical, a laboratory manipulation they induced with a contoured "tent", can cause vertical eye position to deviate. It would be interesting, therefore, to explore this in people walking outdoors on different terrain and at different inclines, as well as indoors where the horizon is not visible.

In a preliminary investigation of eye movements in real-world "free-exploration" and laboratory-based "free-viewing", t' Hart et al. (2009) also found a central bias in both conditions, particularly along the horizontal axis. In the vertical axis, gaze tended to be around 3° above the centre of the head frame-of-reference, particularly in the real-world condition. This is consistent with the present data, which also showed a real-world gaze bias to the upper half of the visual field. Some of the other findings from that study were that a saliency model was a weak predictor of gaze in both settings (which is consistent with the similarity between the objects that were looked at in the present experiment) and that matching the temporal aspects of real world stimuli (as we did by using a video rather than static images) is important when moving into the lab.

The types of objects that were inspected were also quite similar in both real walking and in video watching. Because both sets of participants saw exactly the same stimuli, it may be that some of these objects were selected due to their bottom-up salience, e.g., some objects were looked at in both cases because they were the biggest or brightest things in the visual field, or because they were the only things moving in the environment. However, these cases were relatively rare and many of the things that were fixated often did not contrast heavily with the rest of the scene. For example, people often looked at the path, even though it was a usually a constant grey colour with no high contrast edges or markings.

Other pedestrians in the scene were also fixated often, and this was something that did not differ between participants walking in the environment and those watching a video. In both conditions, most (>70%) pedestrians were fixated just 4 s after appearing in the scene. Given that pedestrians were small, and there were many other items to look at, some of which were also moving (e.g., cars and trucks), this confirms a wealth of research that other agents are prioritized for selection by covert attention and eye movements (Birmingham et al., 2008; Emery, 2000; Foulsham et al., 2010). Critically, it also demonstrates that attentional priority is given to other people both during active behaviour in a real environment as well as during the watching of complex video clips in the lab. That is, the social attention mechanisms observed in the lab are also at work in natural behaviour.

4.2. Walking vs. watching: what is different?

Despite the noted similarities in the general distribution of gaze, and in some of the things that were looked at, we also demonstrated some differences between walking and watching. Describing these differences is an important enterprise as it will allow ourselves and other researchers to begin to investigate why these situations differ and how we can study them in a controlled manner.

First, the measurement of gaze distributions showed some interesting differences between people in the real world and those watching video, and because the stimulus in central vision was the same in both cases these differences are likely due to different scanning strategies in the different environments. When people were walking, we observed a surprisingly constrained gaze distribution: walkers did not spend long fixating away from where their head was pointing. To express this in another way, participants walking in the real world selected objects with head movements, and their eyes tended to stay fixated on a "heading point" slightly above the centre of the head frame-of-reference. When participants watched the same events on a monitor, with their head constrained in a laboratory eye tracker, they shifted their gaze more often to the edge of the visual field. This is even more striking when one considers that the items in the video frame had been pre-selected by virtue of the head-mounted scene camera. Most research recording shifts of gaze in scenes and other naturalistic stimuli does so by measuring eye position while fixing head position. Our results imply that this may not reflect the dynamics of gaze selection in the real world, which involves head movements rather than large scanning eye movements. It also implies that researchers can use head direction as a good first approximation of gaze direction, because in many cases, at least when moving through their environment, people are fixating in the centre of the head-centred field of view. One possible factor in the vertical distribution of gaze during walking, and the difference from the watching condition, is the location of the head-mounted camera, which was positioned slightly above the eyes. Although this merits investigation with different technology, it is unlikely to have caused the observed above-centre bias in the real world. If the eyes were in fact positioned centrally in the head (i.e. according to orbital reserve) then the position of the camera above the eyes would have lead to a bias to look below the centre of the camera's field-of-view, which is the opposite to that found here. This means that the above-centre bias observed in the walking condition, if anything, underestimates that present in the true head-frame (rather than the approximation given by the head-mounted camera). In addition, 't Hart et al. (2010) also reported an above-centre bias, using a different set-up and methodology, which supports this interpretation.

Second, when aligned with a prominent feature in the world—the horizon—the gaze distributions from walking and watching continued to be different. Walking participants spent most of the time fixating slightly below the horizon, whereas participants in the lab were more likely to look above the horizon. This may have been because, when walking, participants were more concerned with items in the lower visual field, which could have interrupted walking. This is consistent with the categorical coding of the objects at fixation. However, the difference between conditions relative to the horizon was relatively small and based on a limited number of frames. Further investigations are needed to confirm this, and to rule out any influence of the particular testing apparatus (e.g. artifacts from the position or angle of the head-mounted camera).

Of course, a major difference between walking and watching conditions was that the participants watching video were not in control of the head and thus the field-of-view afforded by the scene camera. In the real world, gaze is probably influenced by a prediction of how the scene will change given a certain head movement. The absence of this information in most laboratory studies, and in the watching condition here, provides an important caveat for research aiming to make sense of behaviour in the real world. It is possible that our use of a head-mounted camera, and thus the presence of unpredictable camera movements, influenced the results here. For example, participants watching the video may have been planning their eye movements with reference to the horizon in just the same way as when walking, but, being unable to predict the change of head direction, were more erratic in doing so. This could explain the slight difference in horizon-centred eye position. However, analysis of convergence in eye position at the time of a head movement did not reveal a sudden change in the relationship between gaze in walking and watching. Further research is needed to tease apart the effects of real-world immersion and head movements, perhaps by combining scene cameras or tracking head movements more precisely in a virtual environment.

Third, within the categories of objects on which we measured fixation, we found some differences in what was looked at in the lab, as compared to the real world. The main difference was that participants who were immersed in an active task in the real world spent more time looking at the path than those watching the video. Moreover, this difference was found only in gazes on the path near to the observer. We can attribute the decrease in gazes on the path when participants watched a video of the walk to one of the main differences between the conditions, namely the requirement to engage in the active walking task. In Session 1 this task required paying attention to the route, planning where to step next and avoiding obstacles or uneven terrain. In Session 2, although participants watching the video were also asked to pay attention to the route, they were not actually walking, and had no need to monitor their footsteps and no capability of planning or influencing the route. Our categorical coding of gaze suggested that observers watching the video spent time looking at far objects, rather than the near path. These objects tended to include lampposts, trees and distant buildings, which may have been distinctive or salient but were normally not necessary for the task of walking across campus. More generally, our results demonstrate a strong top-down influence of task and action on gaze in natural scenes, something which has been identified and modelled by other researchers (see Ballard and Hayhoe (2009) and Land (2009) for recent reviews). Despite being exposed to the same visual scene, participants selected different objects according to their relevance for current actions. Many laboratory studies of eye movements in scenes ask people to look at a static image with little or no task (i.e. "free viewing"). People may sometimes perform a similar task in the real world (e.g. when looking out of a window or admiring a view) but one suspects that these occasions are rather rare. Our results demonstrate that the task matters, and thus the objects looked at frequently in free viewing may not be looked at as much, or in the same way, when engaged in a real task.

Finally, we highlighted one further case where immersion in a real scene may make a difference—in the allocation of gaze towards other people. Although pedestrians were frequently fixated both by people walking and by people watching video, there was a subtle difference in *when* these pedestrians were looked at. Pedestrians who were far away from the observer were equally likely to be looked at by walkers in the scene and those watching video. However, when these pedestrians were close to the observer, and were passing by, they were more likely to be gazed at by observers watching the event on video. Jovancevic et al. (2006) provide one possible explanation for this. In their study, pedestrians were fixated early, especially when they were likely to collide with the walker. This suggests that the early fixation of people is important for planning the path being walked, something which obviously was not necessary when just watching the video. When the pedestrians are close to the walker, in the real world, the planning process may have moved onto a later part of the route so they no longer need to be fixated. We propose an additional explanation: that the timing difference is due to the authentic social context afforded by a real person, who can look back at the observer, as opposed to a video of a person. This proposal is supported by work documenting gaze avoidance in social situations (Goffman, 1963). It also dovetails with a recent study where we found that participants were less likely to look at a confederate who was seated nearby than at a monitor showing an image of the same confederate (Laidlaw et al., 2011). The present finding confirms that, when it comes to looking at people, there may be important differences between real life and video.

4.3. Conclusions

Laboratory investigations of attention in natural scenes, by design, represent a restricted simulation of the real world. The present study sought to compare gaze behaviour in an immersive, real world task—including where, what and when people look at items in the environment—to the same behaviours in people watching videos of the real world in a lab. The results suggested both similarities and differences, with the differences being attributable to the freedom of immersed participants to move their head and their body, to their engagement in an active task, and to the social constraints of having other real people nearby. Describing these differences, and their causes, should be of the upmost importance to those looking to investigate attention in natural stimuli and situations. Doing so provides new avenues for research, as well as validating past and present work investigating eye movements and attention.

References

Bahill, A. T., Adler, D., & Stark, L. (1975). Most naturally occurring human saccades have magnitudes of 15 degrees or less. *Investigative Ophthalmology, 14*(6), 468–469.

Ballard, D. H., & Hayhoe, M. M. (2009). Modelling the role of task in the control of gaze. *Visual Cognition, 17*(6–7), 1185–1204.

Becker, W. (1991). Saccades. In R. H. S. Carpenter (Ed.), *Eye movements* (pp. 95–137). London: Macmillan Press.

Birmingham, E., Bischof, W. F., & Kingstone, A. (2008). Gaze selection in complex social scenes. *Visual Cognition, 16*(2–3), 341–355.

Birmingham, E., Yanko, M., Barresi, J., Bischof, W. F., Kingstone, A., & Smith, T. J. (2006). The role of attention in complex event perception. Poster presented at the Canadian society for brain, behaviour and cognitive science 16th annual meeting, Saskatoon, Canada, July 22–24.

Cristino, F., & Baddeley, R. (2009). The nature of the visual representations involved in eye movements when walking down the street. *Visual Cognition, 17*(6–7), 880–903.

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

Einhauser, W., Rutishauser, U., & Koch, C. (2008). Task-demands can immediately reverse the effects of sensory-driven saliency in complex visual stimuli. *Journal of Vision, 8*(2), 2. doi:10.1167/8.2.2.

Emery, N. J. (2000). The eyes have it: The neuroethology, function and evolution of social gaze. *Neuroscience and Biobehavioral Reviews, 24*(6), 581–604.

Foulsham, T., Cheng, J. T., Tracy, J. L., Henrich, J., & Kingstone, A. (2010). Gaze allocation in a dynamic situation: Effects of social status and speaking. *Cognition, 117*, 319–331.

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

Foulsham, T., & Underwood, G. (2007). How does the purpose of inspection influence the potency of visual saliency in scene perception? *Perception, 36*, 1123–1138.

Foulsham, T., & Underwood, G. (2008). What can saliency models predict about eye movements? Spatial and sequential aspects of fixations during encoding and recognition. *Journal of Vision, 8*(6), 1–17.

Goffman, E. (1963). *Behavior in public places*. New York: Free Press.

Hayhoe, M. M., Shrivastava, A., Mruczek, R., & Pelz, J. B. (2003). Visual memory and motor planning in a natural task. *Journal of Vision, 3*(1), 49–63.

Henderson, J. M. (2003). Human gaze control during real-world scene perception. *Trends in Cognitive Sciences, 7*(11), 498–504.

Hillstrom, A. P., & Yantis, S. (1994). Visual-motion and attentional capture. *Perception and Psychophysics, 55*(4), 399–411.

Itti, L. (2005). Quantifying the contribution of low-level saliency to human eye movements in dynamic scenes. *Visual Cognition, 12*(6), 1093–1123.

Itti, L., & Koch, C. (2000). A saliency-based search mechanism for overt and covert shifts of visual attention. *Vision Research, 40*(10–12), 1489–1506.

Jovancevic, J., Sullivan, B., & Hayhoe, M. (2006). Control of attention and gaze in complex environments. *Journal of Vision, 6*, 1431–1450.

Jovancevic-Misic, J., & Hayhoe, M. (2009). Adaptive gaze control in natural environments. *Journal of Neuroscience, 29*(19), 6234–6238.

Karacan, H., & Hayhoe, M. (2008). Is attention drawn to changes in familiar scenes? *Visual Cognition, 16*, 346–374.

Kingstone, A. (2009). Taking a real look at social attention. *Current Opinion in Neurobiology, 19*(1), 52–56.

Kingstone, A., Smilek, D., & Eastwood, J. D. (2008). Cognitive ethology: A new approach for studying human cognition. *British Journal of Psychology, 99*, 317–340.

Laidlaw, K. E. W., Foulsham, T., Kuhn, G., & Kingstone, A. (2011). Social attention to a live person is critically different than looking at a videotaped person. *Proceedings of the National Academy of Sciences, 108*(14), 5548–5553.

Land, M. F. (2009). Vision, eye movements, and natural behavior. *Visual Neuroscience, 26*(1), 51–62.

Land, M. F., & Fernald, R. D. (1992). The evolution of eyes. *Annual Review of Neuroscience, 15*, 1–29.

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

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

Patla, A. E., & Vickers, J. N. (2003). How far ahead do we look when required to step on specific locations in the travel path during locomotion? *Experimental Brain Research, 148*(1), 133–138.

Pelz, J., Hayhoe, M., & Loeber, R. (2001). The coordination of eye, head, and hand movements in a natural task. *Experimental Brain Research, 139*(3), 266–277.

Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin, 124*(3), 372–422.

Rothkopf, C. A., Ballard, D. H., & Hayhoe, M. M. (2007). Task and context determine where you look. *Journal of Vision, 7*(14), 16. doi:10.1167/7.14.16.

Stahl, J. S. (1999). Amplitude of human head movements associated with horizontal saccades. *Experimental Brain Research, 126*(1), 41–54.

't Hart, B. M., Vockeroth, J., Schumann, F., Bartl, K., Schneider, E., König, P., et al. (2009). Gaze allocation in natural stimuli: Comparing free exploration to head-fixed viewing conditions. *Visual Cognition, 17*(6), 1132–1158.

Tatler, B. W. (2007). The central fixation bias in scene viewing: Selecting an optimal viewing position independently of motor biases and image feature distributions. *Journal of Vision, 7*(14), 4. doi:10.1167/7.14.4.

Tatler, B. W., Baddeley, R. J., & Gichrist, I. D. (2005). Visual correlates of eye movements: Effects of scale and time. *Vision Research, 45*(5), 643–659.

Tseng, P. H., Carmi, R., Cameron, I. G. M., Munoz, D. P., & Itti, L. (2009). Quantifying center bias of observers in free viewing of dynamic natural scenes. *Journal of Vision, 9*(7), 4. doi:10.1167/9.7.4.

Underwood, G., & Foulsham, T. (2006). Visual saliency and semantic incongruency influence eye movements when inspecting pictures. *Quarterly Journal of Experimental Psychology, 59*(11), 1931–1949.

Yarbus, A. L. (1967). *Eye movements and vision*. New York: Plenum.