Turning the (virtual) world around: Patterns in saccade direction vary with picture orientation and shape in virtual reality

Nicola C. Anderson

Department of Psychology, University of British Columbia, Vancouver, BC, Canada ✉

Walter F. Bischof

Department of Psychology, University of British Columbia, Vancouver, BC, Canada ✉

Tom Foulsham

Department of Psychology, University of Essex, Colchester, UK ✉

Alan Kingstone

Department of Psychology, University of British Columbia, Vancouver, BC, Canada ✉

Research investigating gaze in natural scenes has identified a number of spatial biases in where people look, but it is unclear whether these are partly due to constrained testing environments (e.g., a participant with their head restrained and looking at a landscape image framed within a computer monitor). We examined the extent to which image shape (square vs. circle), image rotation, and image content (landscapes vs. fractal images) influence eye and head movements in virtual reality (VR). Both the eyes and head were tracked while observers looked at natural scenes in a virtual environment. In line with previous work, we found a bias for saccade directions parallel to the image horizon, regardless of image shape or content. We found that, when allowed to do so, observers move both their eyes and head to explore images. Head rotation, however, was idiosyncratic; some observers rotated a lot, whereas others did not. Interestingly, the head rotated in line with the rotation of landscape but not fractal images. That head rotation and gaze direction respond differently to image content suggests that they may be under different control systems. We discuss our findings in relation to current theories on head and eye movement control and how insights from VR might inform more traditional eye-tracking studies.

Introduction

A vast amount of eye movement research has been concerned with what people look at when shown images of the natural world; for example, when observers are shown pictures that include people, they tend to fixate the eyes and face (Birmingham, Bischof, & Kingstone, 2008; Birmingham, Bischof, & Kingstone, 2009). Similarly, there is a bias for people to look toward the center of objects in natural scenes (Anderson & Donk, 2017; Foulsham & Kingstone, 2013; Nuthmann & Henderson, 2010; 't Hart, Schmidt, Roth, & Einhäuser, 2013). We also know that what people look at is highly constrained by their task goals (e.g., Buswell, 1935; Einhäuser, Rutishauser, & Koch, 2008; Henderson, Brockmole, Castelhano, & Mack, 2007; Yarbus, 1967), the context and gist of the scene (e.g., Oliva & Torralba, 2006; Torralba, Oliva, Castelhano, & Henderson, 2006), and the underlying low-level visual properties of the image (e.g., Anderson, Ort, Kruijne, Meeter, & Donk, 2015; Itti & Koch, 2000; Parkhurst, Law, & Niebur, 2002). Much of this type of work has been geared toward predicting *what* people look at and modeling oculomotor control (e.g., Borji & Itti, 2013; Foulsham & Underwood, 2008; Itti & Koch, 2001; Parkhurst et al., 2002; Torralba et al., 2006). There is also a large body of work suggesting that *how* people look at images is also important. Most models of eye movements assume that people are equally likely to look in all directions when viewing images, but there is a growing literature showing that this is not the case. When people look at images, they tend to do so in specific, stereotypical ways. For example, they spend an unusually large amount of time simply looking at the middle of a scene (Tatler, 2007; Vincent, Baddeley, Correani, Troscianko, & Leonards, 2009), and they tend to make saccades

in the cardinal rather than oblique directions (Brandt, 1945; Gilchrist & Harvey, 2006; Tatler & Vincent, 2008; Torralba et al., 2006).

It has also been shown that observers have a particular bias to look along the horizon of an image. In a series of experiments, Foulsham, Kingstone, and Underwood (2008) asked participants to look at images of landscape and interior images that were rotated at various angles, but within a constant square frame (and also bounded by the rectangular monitor frame). They found that the dominant saccade direction followed the orientation of the scene, although this was less pronounced for interior scenes. Two major questions arose from this work. First, in light of the fact that saccade biases are sensitive to the content of the image, what, if any, bias will emerge when the contents of the images are complex but semantically meaningless and presumably isotropic, as with fractal images? Second, to what extent is the bias to follow the horizontal orientation of a rotated scene due to the fact that the viewing aperture contains straight borders that might accentuate a left–right horizontal eye-movement direction?

To answer these questions, Foulsham and Kingstone (2010) asked observers to look at rotated landscapes or fractal images through a circular aperture. They found that, for fractals, saccades were more likely to be made horizontally, irrespective of the orientation of the fractal, and thus within the participants' egocentric frame of reference. In addition, the predominant pattern of saccade directions in landscape images varied with image orientation, but, surprisingly and unlike Foulsham et al. (2008), the predominant saccade direction bias was orthogonal to, rather than parallel to, the image horizon. Why this occurred was unclear, although it is possible that some combination of the rectangular shape of the monitor and the loss of horizontal image content at the top and bottom of the circular images may have played a significant role. Taken together, the results of these studies suggest that image orientation, content, aperture shape, and even the monitor itself may impact saccade direction biases. Although the work of Foulsham and Kingstone (2010) was a step in the right direction (i.e., moving from a square to a circular aperture), there are still many constraints associated with lab-based eye tracking that are unavoidable, such as the requirement to maintain a steady head position and the fact that, whatever the aperture, the image is always situated within a rectangular monitor. Thus, it is not clear the extent to which the biases in Foulsham et al. (2008) and Foulsham and Kingstone (2010) are due to image content or image aperture or whether they are simply an artifact of a constrained testing environment.

One of the fundamental aims of the present study was to replicate and extend core elements of the studies by Foulsham et al. (2008), (2010) using a

viewing environment that eliminates visual cues beyond the image and the aperture itself. In short, will the different effects of a square versus circular aperture on saccades in landscape scenes be observed when potentially confounding visual features outside of the image—such as the monitor or the laboratory environment itself—are eliminated? To achieve this goal, we placed participants within a virtual reality (VR) environment and tracked their gaze while they viewed rotated images through an un-rotated square (Experiment 1), a circular aperture (Experiment 2), or a square that rotated with the image content (Experiment 3). Thus, in Experiments 1 and 2, only the content rotated, whereas in Experiment 3, both the content and the frame rotated (see Figure 1).

A second aim of our study was to examine the effect of rotated images and scene types on head rotation. Most of what we know about visual attention and eye movement control is derived from studies that require people to look at images presented on a computer monitor while they are sitting with their head restrained. There is growing recognition in the field that eye movement research collected in the lab when an observer's head movements are discouraged or not possible, are not representative of how people move their eyes in everyday life when their head is free to move (Backhaus, Engbert, Rothkegel, & Trukenbrod, 2020; Hooge, Hessels, Niehorster, Diaz, Duchowski, & Pelz, 2019; Kingstone, Smilek, & Eastwood, 2008; Risko, Richardson, & Kingstone, 2016; 't Hart, Vockeroth, Schumann, Bartl, Schneider, König, & Einhäuser, 2009). For example, Foulsham, Walker, and Kingstone (2011) asked participants to watch video clips of someone walking across campus to buy a coffee. Although there was some bias to look in the center of the video, their gaze was spread over the scene, looking at objects and the people that the walker encountered. When these same participants were asked to wear a mobile eye tracker and walk across campus to get a coffee, their eyes focused mainly on the path and remained relatively centered within their head. In short, when the head was free to move as people walked across campus, people tended to move their head in order to redirect their gaze to objects and people. In the lab, Solman, Foulsham, and Kingstone (2017) have shown that, when participants are required to look at a scene through an asymmetric window that is yoked to their eyes, eye movements target regions within the window. However, when the window is yoked to an observer's head movements, the head moves to reveal new information outside the window, presumably so that the eyes can then examine visual information within the new window. These and other studies (e.g., Land & Hayhoe, 2001; 't Hart et al., 2009) suggest that in the real world and in the lab, when the head is free to move, it does so in a manner that complements and operates in service of the eyes.
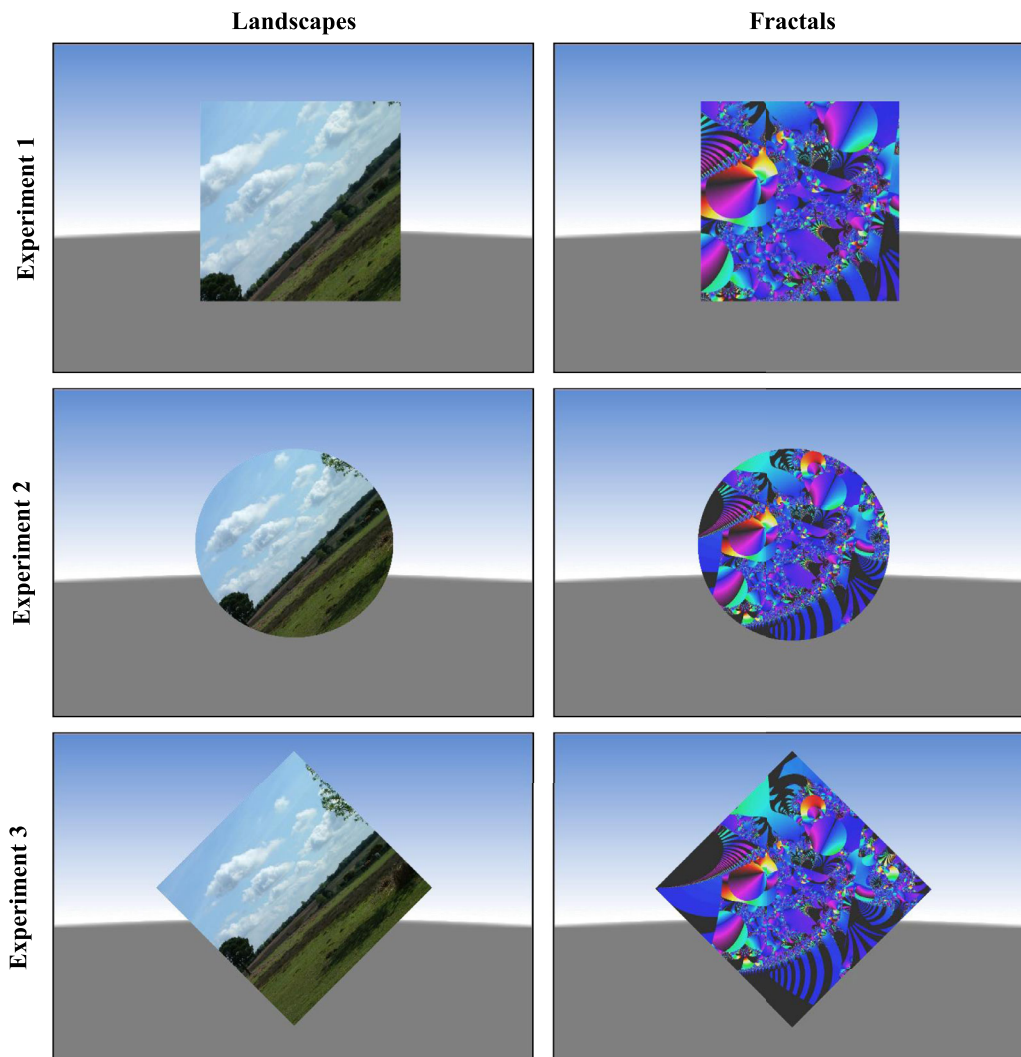
Figure 1. Example square (Experiment 1), circular (Experiment 2), and square frame (Experiment 3) landscape and fractal images rotated 45° counter-clockwise from the participant's perspective in the VR headset. Note that in Experiment 1 the rotated image is also zoomed relative to Experiments 2 and 3 due to the greater amount of cropping required (see Figure 2).

Much of what is known about head movements comes from research that is more kinematic in nature. From this research, we have learned that the relative timing of eye and head movements may suggest whether attentional selection is reflexive or volitional (Doshi & Trivedi, 2012; Freedman, 2008; Zangemeister & Stark, 1982). For example, when the eyes move before the head, these are unplanned, reflexive movements (usually to a suddenly presented stimulus such as a flash of lightning) and involve shifts of less than 45° (Barnes, 1979). When the head leads the eyes, however, these are thought to be large, planned, purposeful movements, often to a known target location. These conclusions are mainly based on experiments where participants respond to simple light displays or targets on a screen but it has also been shown in more naturalistic settings, where, for example, it was found that the best predictor of a lane change in a driving simulator is not an

eye movement but a head movement that occurs 2 to 3 seconds beforehand (Doshi & Trivedi, 2012). In addition, very little research has been conducted regarding head rotation in conjunction with eye movements, although it is known that head rotation helps observers read rotated text (Risko, Medimorec, Chisholm, & Kingstone, 2014).

To date, however, with a few notable exceptions (Kothari, Yang, Kanan, Bailey, Pelz, & Diaz, 2020; Matthis, Yates, & Hayhoe, 2018), the relation between the head and the eyes has not been studied concurrently in complex settings. For example, in Solman et al. (2017) only the head or only the eyes were monitored. In mobile eye tracking, the eyes are tracked relative to the head (i.e., head-centered, not world-based coordinates; see Hessels, Niehorster, Nyström, Andersson, & Hooge, 2018), and head position is often only inferred from the camera position. There are limited scene perception

studies that provide information about both head and eye movements, and combining the two remains methodologically difficult (c.f., Backhaus et al., 2020). Fortunately, recent advances in VR technology have provided the opportunity to track the head and gaze in world-based coordinates simultaneously. In the present study, to the extent that the head operates in service of the eyes, one might, for example, expect that the head will rotate in a way that serves to facilitate saccades along the horizon of rotated landscapes scenes.

In sum, the present work seeks to examine the direction and relation of saccades and head rotation while viewers are presented with natural scenes and fractals in a VR environment. Our starting point for this new line of research begins with the work of Foulsham and colleagues who found that the effect of rotated landscape scenes viewed through a circular aperture (Foulsham & Kingstone, 2010) is paradoxically the opposite to the effect of a square aperture (Foulsham et al., 2008). Note, however, in the Foulsham studies and the many that are like it, head movements are prohibited and eye movements occur while images are presented within a rectangular computer monitor. In contrast, in the present study, observers were free to move the head and eyes while viewing rotated landscapes scenes and fractals through apertures displayed within a fully immersive VR environment; that is, the potentially confounding visual features beyond the images presented within the aperture—such as a monitor or the laboratory environment itself—were removed. In Experiment 1, the images were square and rotated within a fixed square frame (as in Foulsham et al., 2008). In Experiment 2, the images were circular and rotated within a circular frame (as in Foulsham & Kingstone, 2010). In Experiment 3, we add a novel condition in which both the square frame and the content within it rotated fully (Figure 1).

Methods

The methods for Experiments 1, 2, and 3 were the same, except where noted below.

Participants

In keeping with the Foulsham et al. (2008, 2010) studies, which had 12 (2008) to 20 (2010) participants, we targeted the upper end of this range in all of our experiments. A post hoc power analysis was conducted using G*Power (Erdfelder, Faul, & Buchner, 1996) to confirm whether our sample size was adequate for the within–between analysis of variance required for this work. The statistical power approached a power of 1 by using a moderate effect size ($\eta_p^2 = 0.25$) and overall sample size of 59 (across the three experiments) with three experimental groups and seven measurements

(corresponding to the two stimulus types and five levels of stimulus rotation) and the default correlation setting of 0.5.

In Experiment 1, 18 participants (ages 18–27 years, $M = 20$ years; 10 female) were recruited from the University of British Columbia and participated in this experiment for course credit. One participant was removed due to equipment failure. All participants reported normal or corrected-to-normal vision and were naive to the purpose of the experiment. Participants provided informed consent prior to participant, and the study was approved by the ethics board of the University of British Columbia (H10-00527).

Apparatus

Picture encoding

This portion of the study was conducted on a custom-built desktop computer running an VIVE virtual reality device (HTC Corporation, New Taipei City, Taiwan) equipped with an SMI eye tracker (iMotions, Copenhagen, Denmark). The VIVE headset has a resolution of 1080×1200 pixels, with a 110° field of view and a refresh rate of 90 Hz. The virtual experimental set-up was built using the Unity platform, a popular game development software (Unity Technologies, 2020). The SMI eye tracker has a sampling rate of 250 Hz and was controlled using the SMI-designed Unity plugin. Tracking accuracy was maintained by performing a calibration every 20 trials. Calibration consisted of following a moving white circle with a red dot in the middle to five different regions. The SMI Unity plugin reports only success or failure of calibration, not accuracy, so calibration was repeated upon failure. Head movements (movements of the VIVE headset) were tracked via two infrared base stations located on opposite corners of the room. Participants were provided with brief instructions on handling the VIVE headset and were then seated in a swivel chair in the center of the room (in clear view of both base stations). They were handed the keyboard and directed to the response key (keypad Enter).

The virtual testing environment was sparse, consisting only of the standard Unity skybox, which has a gray floor and neutral, blue horizon (Figure 1). Images were presented on a three-dimensional (3D) 4×4-m cube with a depth of 0.1 m, which was placed approximately 3 m away from the participant, subtending approximately 60×60 degrees visual angle. The cube was centered to the headset but was not yoked to head movement.

Stimuli

The images were landscape and fractal images extracted from the set used in Foulsham and Kingstone (2010). The landscapes were photos of outdoor
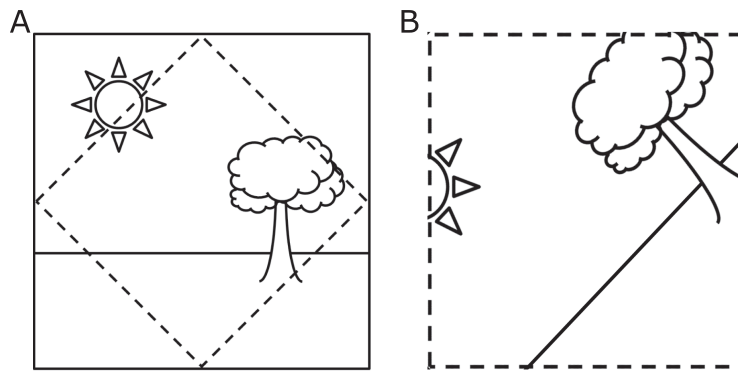
Figure 2. Example of the (A) crop and (B) zoom of an image rotated 45° in Experiment 1. The largest rotated square that fit obliquely into the original image was used to extract image content for each stimulus rotation. For cardinal image rotations (including un-rotated images), a square the same size as the largest rotated square was cut from the center of the image. This kept the content as close as possible to similar across image rotations in Experiment 1. In Experiments 2 and 3, the original, un-zoomed images were used.

environments, many of which contained a visible horizon. The fractal images were computer-generated images taken from the Spanky fractal database. Images were cropped to 768 × 768 to fit on a screen made of a 3D cube object in the virtual environment. Images were rotated counter-clockwise from the participant's perspective by 0°, 45°, 90°, 135°, 180°, 225°, 270°, and 315° but were made to fit into an upright, square frame (4 × 4 m). This required cropping a small amount of the image (Figure 1). Note that this meant that the images in Experiment 1 were zoomed relative to those in Experiments 2 and 3 (Figure 2). Horizon-aligned rotations were collapsed for the purposes of counterbalancing (i.e., 45°/225°, 90°/270°, 135°/315°) in order to reduce the length of the experiment, as the VR headset can get quite warm and cumbersome for participants to wear over long periods of time. Although fractal images do not have a horizon in the same way that landscape images do, for ease of communication we defined the fractal image horizon to be along the horizontal axis of the original, unrotated fractal, as if the upright fractal image had a horizontal horizon, like a landscape. This fractal "horizon" rotates along with the fractal rotations.

Procedure

Participants were instructed that there would be two phases to the study, an image-encoding phase and an image-recognition phase. The instruction that there would be a recognition component to the experiment was included to encourage participants to explore both the landscape and fractal images, but it was not of theoretical interest to the present study. Nevertheless, we conducted the test to ensure that anyone who might

hear about the study before taking part in it would know that the instruction regarding a recognition phase was truthful.

In the encoding phase, participants were familiarized with the VR equipment and given the instructions to remember each of the pictures they were about to see for a later memory test. Each trial began with an un-rotated, gray aperture (matching the aperture type used in the experiment) with a fixation cross in its center. Although this was not a drift correction, it allowed for the experimenter to monitor tracker accuracy. Participants pressed the Enter key on a number pad that they were holding to indicate that they were ready for the trial to begin. Each picture was presented for 10 seconds, a longer duration than in Foulsham and Kingstone (2010), to allow sufficient time to record any potential, slower, head movements. There were 160 trials (80 fractal images and 80 landscape images). The rotation condition was counterbalanced, such that across five participants each image was presented in each of the five rotation conditions. This part of the experiment took approximately 20 minutes.

In the recognition phase of the experiment, participants were asked to remove the VR headset and complete this phase on a computer monitor. Eighty landscape and fractal images were displayed in their canonical orientation (40 randomly chosen "old" images from the image set and 40 new) for 2.5 seconds. Participants were required to press the "Z" key if they had seen the image before, or the "/" key if they had not. If they indicated that they had seen the image before, they were asked to indicate which orientation that they remembered the image being in (0°, 45°, 90°, 135°, or 180°). We performed this test for each of the three experiments. Note that recognition performance was 62% correct across all three experiments, and there was

no significant difference in accuracy across experiments ($F < 1$). This aspect of the investigation is not discussed further.

Experiments 2 and 3

The methods for Experiments 2 and 3 were exactly the same as Experiment 1 with the following exceptions described below.

Participants

Experiment 2 had 23 participants (ages 18–28 years, $M = 21$ years; 17 female). These additional subjects were run with the goal of matching the sample size of Experiments 1 and 3, but more participants signed up than expected. We opted to keep these participants in rather than exclude them for the sake of matching sample sizes. Experiment 3 had 18 participants (ages 18–37 years, $M = 21$ years; 9 female). They were recruited in the same manner as in Experiment 1.

Stimuli

Stimuli were un-manipulated versions of those used in Experiment 1, with the exception that in Experiment 2 they were presented within a circular VR aperture that had the same maximum diameter as the square aperture used in Experiments 1 and 3. In Experiment 3, the square VR aperture rotated with the image content.

Results

Throughout the analyses, the experiment was treated as a between-subjects factor in order to compare performance across the different viewing conditions. We first examine general gaze behavior for landscape and fractal images in the virtual environment and then look specifically at saccade direction and amplitude biases as a function of the semantic content of the scenes and their rotations. We then turn to examining general head movement measures before examining head rotation.

Data handling

Using the SMI Unity plugin and treating gaze direction as a 3D vector, we recorded each participant's gaze direction in space at 250 Hz. Head position (position of the headset) was recorded at the fastest possible rate (which in Unity, is variable but close

to 90 Hz) and linearly interpolated to 250 Hz (our experiment scripts are available at https://osf.io/hc5k6/). Using the head position and gaze direction vector, gaze vectors were projected on the virtual screen to obtain two-dimensional gaze positions on the reference frame of the screen plane. Gaze positions outside the virtual screen were recorded as missing values. In this way, the 3D gaze data were converted into samples on the virtual screen which could then be examined in a way equivalent to traditional monitor-based studies. Head positions with respect to the virtual screen (for calculating head shifts and central tendency) was calculated in the same manner as gaze, using the 3D head direction vector.

It is worth clarifying here the reference frame we are dealing with (Hessels et al., 2018). In the present study, we examined gaze data and head data (specifically, rotation), separately. Most lab-based eye-tracking studies measure eye movements on a screen, where the frame of reference is world centered. In contrast, mobile eye tracking tends to track gaze position with respect to the world camera, where the frame of reference moves with the head (i.e., head centered or eye in head). In VR, we track gaze and head with respect to a common reference frame. That is, gaze and head position are referenced to the screen displayed, which is equivalent to world-based coordinates. Note that we could not differentiate neck movements from chair and torso movements, which likely contributed to the final head rotation measure used in this work.

As the SMI outputs only sample data, it was necessary to perform event detection on this raw data. Fixations were defined as stable gaze points and were extracted using the IDT algorithm (Salvucci & Goldberg, 2000; Blignaut, 2009; Komogortsev et al., 2010), using a minimum duration of 80 ms and a maximum dispersion threshold of 3°. Saccades were defined as differences between successive fixations, and saccades with amplitudes less than 1° and greater than 45 degrees visual angle were filtered out (Freedman, 2008).

All statistical analysis and follow-up tests were conducted using the R packages afex (Singmann, Bolker, Westfall, Aust, & Ben-Shachar, 2019) and emmeans (Lenth, Singmann, Love, Buerkner, & Herve, 2019); figures were produced using ggplot2 (Wickham, 2016). Fixations with durations longer than 800 ms were removed from the analysis (1.7% of data removed).

Gaze data

General gaze measures

We first investigated whether the type of image had an impact on general gaze measures (Table 1), following previous work (Foulsham & Kingstone, 2010). For each

Measure	Landscapes	Fractals
Experiment 1 (square fixed frame)		
Number of fixations per trial, *n*	27.62 (7.80)	27.22 (7.29)
Mean fixation duration, ms	200.85 (28.66)	203.43 (26.30)
Central tendency, degrees visual angle	15.49 (2.73)	15.62 (3.21)
Experiment 2 (circular frame)		
Number of fixations per trial, *n*	31.71 (6.39)	30.69 (6.21)
Mean fixation duration, ms	207.47 (24.66)	212.35 (24.63)
Central tendency, degrees visual angle	14.27 (1.85)	14.33 (1.88)
Experiment 3 (square rotating frame)		
Number of fixations per trial, *n*	30.42 (7.76)	29.00 (7.38)
Mean fixation duration, ms	192.70 (18.65)	199.97 (22.99)
Central tendency, degrees visual angle	17.05 (3.33)	17.35 (3.97)

Table 1. Means (standard deviations) of general gaze measures at encoding for landscape and fractal images.

experiment, we investigated whether the mean number of fixations, fixation duration (in ms), and central tendency differed across experiments and stimulus type. The mean central tendency was calculated from the absolute distance of each fixation, in degrees of visual angle, from the screen center. For each measure, we computed a three (experiment) by two (stimulus type) mixed analysis of variance, with experiment as a between-subjects factor.

For the mean number of fixations per trial, there was a main effect of stimulus type, $F(1, 56) = 17.63$, $p < 0.001$, $\eta_p^2 = 0.24$, such that there were more fixations in landscape images, $M = 30.02$ (7.35), compared to fractal images, $M = 29.08$ (6.95). There was no effect of experiment, $F(2, 56) = 1.51$, $p = 0.23$, $\eta_p^2 = 0.05$, and no interaction between experiment and stimulus type $F(2, 56) = 1.61$, $p = 0.21$, $\eta_p^2 = 0.05$.

For mean fixation durations, there was a main effect of stimulus type, $F(1, 56) = 28.85$, $p < 0.001$, $\eta_p^2 = 0.34$, such that fixation durations were longer for fractals, $M = 205.91$ (24.89) compared to landscapes, $M = 201.08$ (24.87). There was no main effect of experiment, $F(2, 56) = 1.56$, $p = 0.22$, $\eta_p^2 = 0.05$, and no interaction between experiment and stimulus type, $F(2, 56) = 2.04$, $p = 0.14$, $\eta_p^2 = 0.07$. Note that the analysis of fixation durations is based on subject-wise means of fixation durations, which have an approximate normal distribution even though the fixation durations are highly skewed.

For the mean central tendency, there was a main effect of experiment, $F(2, 56) = 5.27$, $p = 0.008$, $\eta_p^2 = 0.16$, such that fixations were farther from the center in Experiment 3, $M = 17.20$ (3.61) compared to Experiment 2, $M = 14.30$ (1.84), $t(56) = 3.25$, $p = 0.006$, $d = 0.43$. All other pairwise comparisons were non-significant (all $t < 1.8$, all $p > 0.25$). There was no main effect of stimulus type, $F(1, 56) = 1.40$, $p = 0.24$, $\eta_p^2 = 0.02$, nor an interaction between experiment and stimulus type ($F < 1$).

Taken together, there were more but shorter fixations in landscape images compared to fractals. This matches a similar (non-significant) trend in Foulsham and Kingstone (2010). Experiment, or image aperture, had no measurable effects on fixation numbers or durations; however, there was an effect on central tendency. Fixations were less centralized when both the image and frame were rotated in Experiment 3, compared to the circular aperture of Experiment 2, but Experiments 1 and 3 and Experiments 1 and 2 did not differ significantly.

Eye in head

Gaze positions on our virtual screen, and in VR more generally, are inherently a combination of eye position in the head and head direction. One advantage of eye and head tracking in VR is that it is possible to measure the contribution of the eye, over and above that of the head. To investigate this, we computed the divergence between eye and head positions by taking the angle subtended by the distance between the gaze and head positions on the screen. In order to assess whether the this measure varied as a function of experiment (image shape: square vs. circle), image type, and rotation, a three (experiment) by two (stimulus type: landscape vs. fractal) by eight (stimulus rotation) mixed analysis of variance was conducted on the mean divergence between the eye and head. Greenhouse–Geiser corrections were applied where sphericity violations occurred. Note that stimulus rotations were not collapsed across 45°/225°, 90°/270°, or 135°/315°, as we had no previous hypotheses about how image content or rotation might impact this measure.

There was no main effect of experiment, and experiment did not interact with any other factors (all $F < 1$); therefore, the data presented in Figure 3 are collapsed across experiment. There was also no
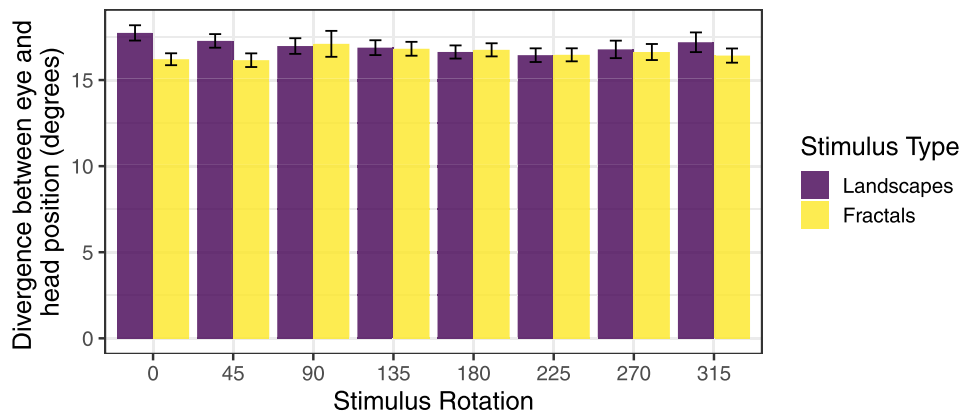
Figure 3. Angular deviation between eye and head positions as a function of stimulus type and stimulus rotation.

main effect of stimulus rotation ($F < 1$). There was a main effect of stimulus type, $F(1, 56) = 7.02$, $p = 0.010$, $\eta_p^2 = 0.11$, which was qualified by a marginal interaction between stimulus type and stimulus rotation, $F(3.58, 200.71) = 2.10$, $p = 0.089$, $\eta_p^2 = 0.04$. Bonferroni-corrected pairwise comparisons between stimulus type at each level of stimulus rotation revealed that this marginal interaction likely arose from significantly larger divergence between eye and head positions for unrotated landscapes and for landscapes rotated 45°, compared to their fractal counterparts (all $t > 2.42$, $p < 0.02$, $d > 0.11$). No other pairwise comparisons between landscapes and fractals were significant (all $t < 1.68$, $p > 0.09$) (see Figure 6). Taken together, these results suggest that the divergence between eye and head position is somewhat sensitive to image content and rotation but generally remained stable at around 16.5 degrees visual angle. This finding aligns with the central tendency data and the idea that these images were predominantly explored via eye movements while the head remained relatively stable in a central position.

Saccade direction

Figure 4 shows the saccade direction density distributions across experiments, stimulus types, and stimulus rotations. Fixation positions were output by Unity in world-based coordinates that were translated on the fly into screen-based coordinates. From Figure 4, it is clear that saccade directions varied with image rotation across all three experiments. To mirror the work of Foulsham and Kingstone (2010), for statistical analysis of these data, saccade directions were grouped into four symmetrical bins of saccades in the horizontal (0°/180°), vertical (90°/270°), leftward oblique (45°/225°), and rightward oblique (135°/315°). In order to assess whether the frequency of saccade directions varied among experiments (and thus any potential effects of image framing), image type, and rotation, a three (experiment) by two (stimulus type:

landscape vs. fractal) by five (stimulus rotation) by four (saccade direction axis) mixed analysis of variance was conducted on the mean frequency of saccades in each direction, with experiment treated as a between-subjects factor. Greenhouse–Geiser corrections were applied where sphericity violations occurred.

There was no main effect of experiment and no interactions of experiment with any other factors (all $F < 1.4$), with one exception. There was a significant interaction among experiment, stimulus type, and saccade direction axis, $F(4.2, 117.7) = 2.62$, $p = 0.04$, $\eta_p^2 = 0.09$. This interaction appears to arise from the fact that, for landscape images across all stimulus rotations, horizontal saccades in Experiment 1 ($M_{freq} = 0.26$) were marginally less frequent than in Experiment 2 ($M_{freq} = 0.27$), $t(321) = 2.10$, $p = 0.09$, $d = 0.12$, and were significantly less frequent than in Experiment 3 ($M_{freq} = 0.28$), $t(321) = 2.51$, $p = 0.03$, $d = 0.14$. Consequently, for landscape images, there were significantly more vertical saccades in Experiment 1 ($M_{freq} = 0.29$) compared to Experiment 2 ($M_{freq} = 0.27$), $t(321) = 2.70$, $p = 0.02$, $d = 0.15$, and Experiment 3 ($M_{freq} = 0.27$), $t(321) = 2.93$, $p = 0.01$, $d = 0.16$. Given that there was no evidence that experiment (and thus, image framing) influenced saccade directions across different stimulus rotations and that experiment did not interact with any other factors, Figure 5 shows the binned saccade direction data collapsed across the factor experiment.

There was a main effect of saccade direction axis, $F(1.91, 10.75) = 64.05$, $p < 0.001$, $\eta_p^2 = 0.53$, an interaction between stimulus type and saccade direction axis, $F(2.10, 117.70) = 21.14$, $p < 0.001$, $\eta_p^2 = 0.27$, and between stimulus rotation and saccade direction axis, $F(2.13, 119.10) = 24.71$, $p < 0.001$, $\eta_p^2 = 0.31$. These interactions were qualified by a significant stimulus type by stimulus rotation by saccade direction axis interaction, $F(8.45, 472.96) = 2.92$, $p = 0.003$, $\eta_p^2 = 0.05$.[1]
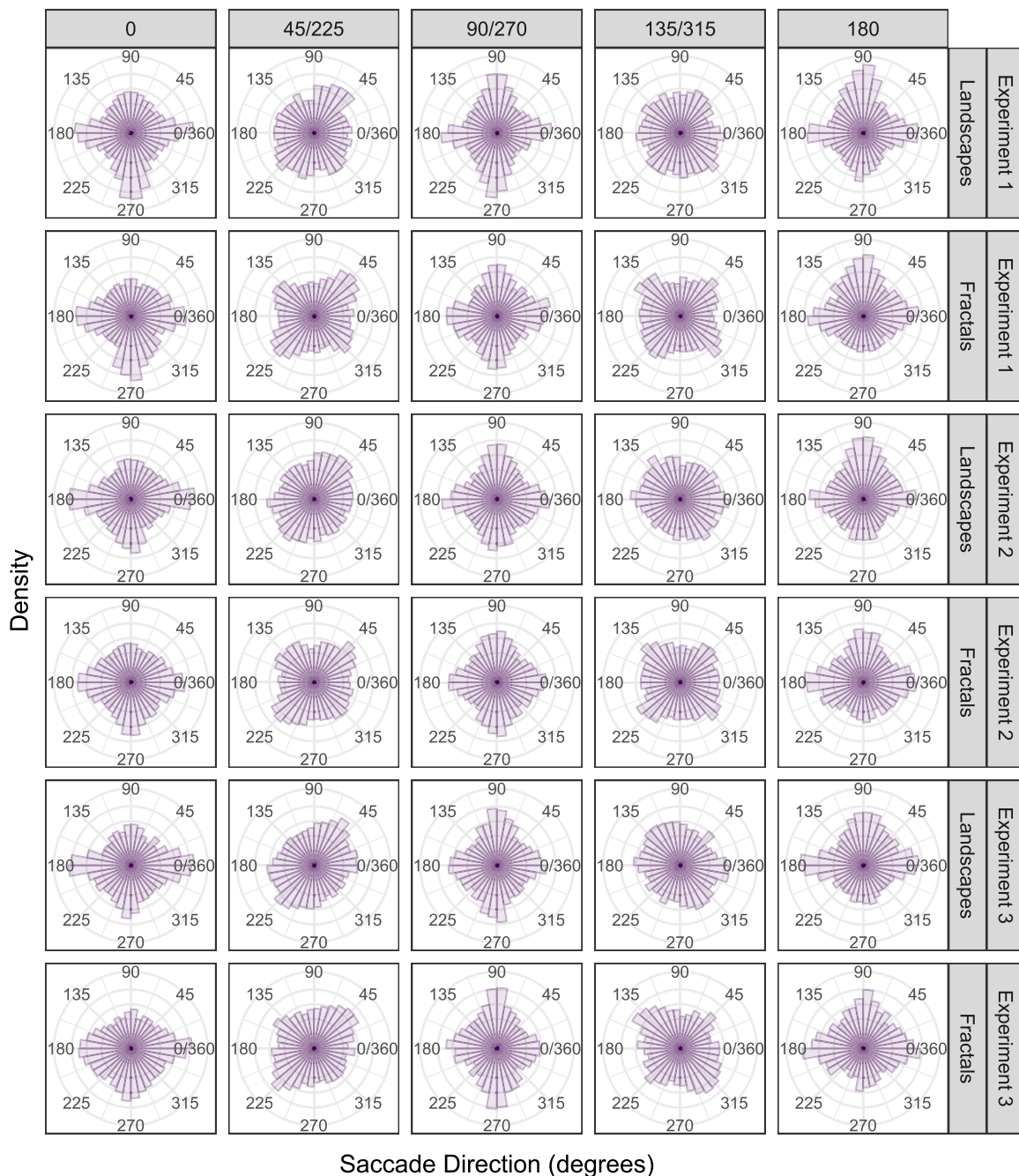
Figure 4. Saccade direction distributions (with respect to image coordinates) as a function of image orientation and image type. Each subplot shows the relative frequency of saccades in each of 36 bins.

In order to interpret this higher order interaction, separate analyses of variance were conducted for each stimulus rotation. That is, an analysis of variance was conducted across the data in the vertical panels of Figure 5. For example, a two (stimulus type) by four (saccade direction) analysis of variance was conducted on the relative frequency of saccades for stimuli rotated 45°/225°. In all cases, there was a significant stimulus type by saccade direction axis interaction (all $F >$ 4.95, $p < 0.005$), except for stimulus rotation 90°/270°

(accounting for the initial three-way interaction), $F(2.26, 126.46) = 2.14$, $p = 0.11$, $\eta_p^2 = 0.04$. In these cases, follow-up Bonferroni-corrected pairwise comparisons were made at each level of stimulus type. Although these varied statistically across stimulus type and rotation, it was generally the case that, in landscapes, the most frequent saccade direction was parallel to the image horizon in all cases except the 180° condition (where the most frequent direction was in the 90°/270° axis). The results were very similar in fractal
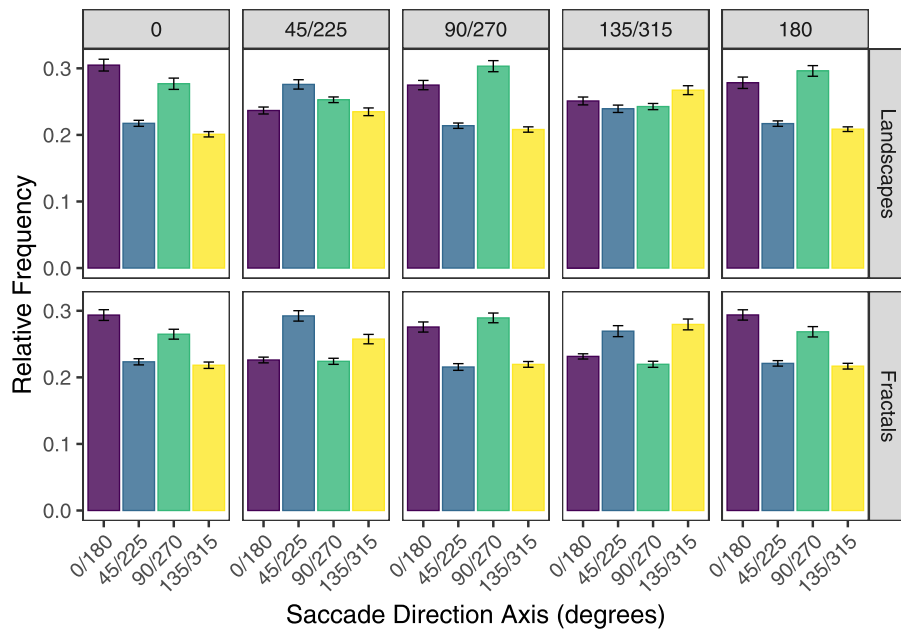
Figure 5. Saccade direction axes for each image type and rotation. Saccades were split into four symmetrical groups. Saccades were defined as screen-based such that, in an image rotated at 45°/225°, saccades along the image horizon are the saccade directions of 45°/225° in this figure.
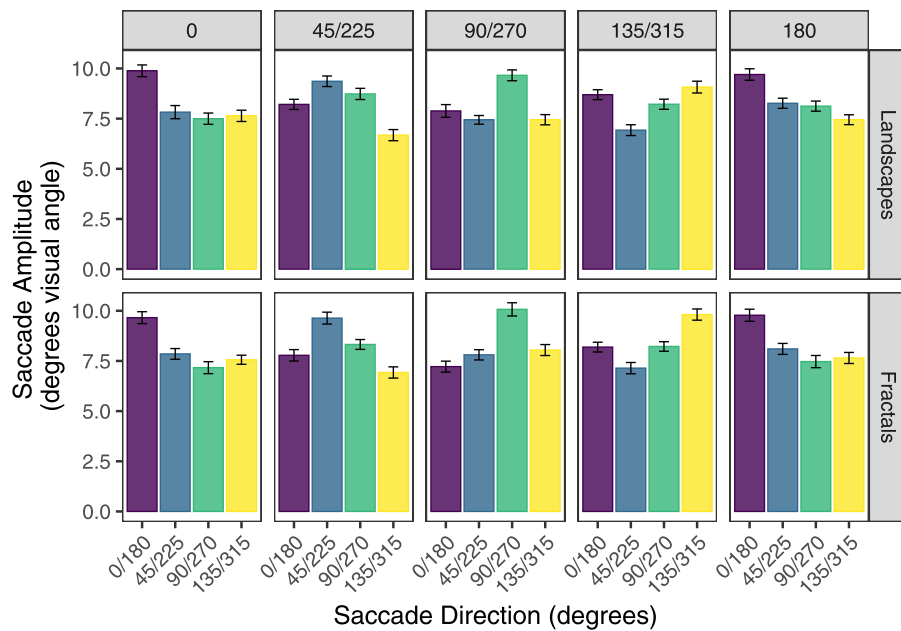

Figure 6. Mean saccade amplitudes at each binned saccade direction, stimulus type, and stimulus rotation.

images, where the modal direction was significantly different from other directions (all $t > 2.70$, $p < 0.05$, $d > 0.21$).

Taken together, it appears that saccades are made predominantly parallel to the image horizon of landscapes and fractal images, followed by saccades made orthogonal to the horizon. This pattern is strongest in landscape images in the cardinal rotations (see Figure 5) and, surprisingly, in fractals. The latter finding is an interesting and unexpected pattern that contrasts with the work of Foulsham and Kingstone (2010), where saccades were more likely to be made along the participants' egocentric horizon, rather than along the native horizontal orientation of the fractal.

We consider this finding in more detail in the discussion section.

Saccade amplitude

In previous work (Foulsham et al., 2008), saccade amplitudes varied as a function of image orientation, such that the longest amplitude saccades tended to be made along the image horizon. Here, we replicated those findings; yet, it appears that the most frequent saccades tended to be made cardinal with respect to the image horizon of both landscape and fractal images. Perhaps the saccade amplitudes might shed further light on these biases. In order to assess whether saccade amplitudes varied as a function of experiment (image shape: square vs. circle), image type, and rotation, a three (experiment) by two (stimulus type: landscape vs. fractal) by five (stimulus rotation) by four (saccade direction axis) mixed analysis of variance was conducted on the mean amplitude of saccades in each direction. Greenhouse–Geiser corrections were applied where sphericity violations occurred.

There was no main effect of experiment, $F(2, 56) = 2.33$, $p = 0.11$, $\eta_p^2 = 0.08$, and no two-way interactions of experiment with any other factors (all $F < 1$); however, there was a significant interaction among experiment, stimulus type, and saccade direction axis, $F(5.63, 157.68) = 2.23$, $p = 0.05$, $\eta_p^2 = 0.07$. Bonferonni-corrected comparisons between saccade amplitudes across experiment at each level of saccade direction axis and stimulus type revealed only a few marginally significant differences between amplitudes across experiments (all $t < 2.19$, $p > 0.08$). Because there were no discernable patterns that could explain this interaction and no interactions of experiment with any other factors (all $F < 1.18$), the data presented in Figure 6 are collapsed across experiments.

The pattern of results for the amplitude data is very similar to that of the saccade direction data. There was a main effect of saccade direction axis, $F(2.83, 158.42) = 44.61$, $p < 0.001$, $\eta_p^2 = 0.44$; an interaction between stimulus type and saccade direction axis, $F(2.82, 157.68) = 13.46$, $p < 0.001$, $\eta_p^2 = 0.19$; and an interaction between stimulus rotation and saccade direction axis, $F(2.63, 147.50) = 34.52$, $p < 0.001$, $\eta_p^2 = 0.38$. These interactions were qualified by an interaction among stimulus type, stimulus rotation, and saccade direction axis, $F(9.25, 518.08) = 2.72$, $p = 0.004$, $\eta_p^2 = 0.05$.

In order to investigate this latter interaction, separate analyses of variance were conducted for each stimulus rotation; that is, an analysis of variance was conducted across the data in the vertical panels of Figure 5. For example, a two (stimulus type) by four (saccade direction) analysis of variance was conducted on the mean saccade amplitude for stimuli rotated 45°/225°. In all cases, except for stimulus rotations 0 (accounting for the significant three-way interaction reported above)

($F < 1$), there was a significant stimulus type by saccade direction axis interaction (all $F > 3.62$, $p < 0.05$). Bonferroni-corrected pairwise comparisons among saccade amplitudes at each saccade direction axis across stimulus type and rotations revealed slightly different patterns of significance between saccade direction axes, but in most cases (including fractal images), the largest saccades were made in the direction associated with the image horizon. Notable exceptions included image rotation 135°/315°, where this pattern was observed for fractal images, (all $t > 5.77$, $p < 0.001$, $d > 0.41$), but not for landscapes ($t < 1$). Taken together, there is a general pattern where the longest saccades were made along the axis associated with the image horizon in both landscape images and in fractals.

Summary

A key question of the present work was whether the shape in which an image is shown to participants will affect their saccade direction biases. We aimed to replicate and extend the work of Foulsham et al. (2008) and Foulsham and Kingstone (2010), who found in one case (equivalent to our Experiment 1) biases for saccades along the horizon of an image and in another case (equivalent to our Experiment 2) biases for saccades against the horizon. Here, aside from a few statistical exceptions, we found that more saccades were made along the horizon of landscape and fractal images, and these saccades were generally the largest. Importantly, there was a similar (but smaller in magnitude) bias for saccades orthogonal to the horizon. This suggests that observers prefer to make saccades in cardinal directions *relative to the image content*. We extended these finding to the novel case in which the image and frame were both rotated. The only difference in this condition compared to Experiments 1 and 2 was a decrease in the central tendency of fixations, which is perhaps not surprising given that in Experiments 1 and 2 a small amount of the scenes had to be cropped to fit within the static frame (see Figures 1 and 2). In short, the aperture in which an image is framed had little effect on the bias to make cardinal saccades with respect to image content.

Head movement data

The second main question of this work was whether image aperture and rotation would affect head movements and, in particular, head rotation around the view axis (in the same plane as the image rotation). One of the key advantages to eye tracking in VR is that the head is free to move without affecting the ability to measure eye movements. Below we outline general head movement measures, then examine the effect of aperture and image rotation on head rotation.

Measure (degrees visual angle)	Landscapes	Fractals
Experiment 1 (square fixed frame)		
Mean head shift amplitude	1.00 (0.78)	1.03 (0.90)
Central tendency	15.30 (4.70)	15.87 (4.46)
Experiment 2 (circular frame)		
Mean head shift amplitude	1.00 (0.77)	1.01 (0.73)
Central tendency	9.37 (3.12)	9.45 (3.14)
Experiment 3 (square rotating frame)		
Mean head shift amplitude	1.98 (0.93)	1.94 (0.99)
Central tendency	11.96 (3.06)	12.19 (3.12)

Table 2. Means (standard deviations) of general head movement measures at encoding for landscape and fractal images.

General head movement measures

Table 2 presents head movement measures as a function of aperture (experiment) and stimulus type. In similar work where head movements were measured via a mobile eye tracker, they were smoother and slower than eye movements (Backhaus et al., 2020). For this reason, we opted to define them relative to the fixation information extracted via event detection (see Data handling). Mean head shift amplitude is defined as the difference in head position (pitch + yaw) between successive fixations. Head central tendency is the mean of the absolute distances, in degrees visual angle, between head positions during a given fixation and the center of the aperture. For each measure, we computed a three (experiment) by two (stimulus type) mixed analysis of variance, with experiment as a between-subjects factor.

For the mean head shift amplitude there was a main effect of experiment, $F(2, 56) = 7.71$, $p = 0.001$, $\eta_p^2 = 0.22$, such that mean head shift amplitude was larger in Experiment 3 compared to Experiment 1, $t(56) = 3.37$, $p < 0.005$, and Experiment 2, $t(56) = 3.54$, $p < 0.005$, $d = 0.95$ (see Table 2). There was no main effect of stimulus type nor an interaction between experiment and stimulus type (all $F < 1$).

For the mean central tendency of head movements there was a significant main effect of experiment, $F(2, 56) = 14.99$, $p < 0.001$, $\eta_p^2 = 0.35$; a main effect of stimulus type, $F(1, 56) = 10.95$, $p = 0.002$, $\eta_p^2 = 0.16$; and a marginal interaction between experiment and stimulus type, $F(2, 56) = 2.86$, $p = 0.07$, $\eta_p^2 = 0.09$. Follow-up pairwise comparisons between stimulus types at each level of experiment revealed that this interaction arose from the fact that the head remains closer to the center of landscape compared to fractal images in Experiment 1, $t(56) = 3.68$, $p < 0.001$, $d = 0.49$. There were no significant differences between central tendency across landscapes and fractals in the other experiments (all $t < 1.44$, $p > 0.15$).

Taken together, the general head movement measures tended to align with the gaze measures reported above. Head shift amplitudes were larger in Experiment 3 compared to Experiments 1 and 2, consistent with the idea that there was slightly more image content in Experiment 3 due to the fact that image cropping was not required. Stimulus type had little impact on general head movement measures, except for a reduced central tendency (or greater image exploration) of the fractal images, compared to the landscape images of Experiment 1. Interestingly, the head moved away from the center of the image more in Experiment 1 compared to Experiments 2 and 3. This contrasts with the gaze data, where central tendency was largest in Experiment 3. Overall, by looking at the central tendency values in Tables 1 and 2, one finds that the eye and head central tendencies corresponded the most in Experiment 1 and were most divergent in Experiments 2, where the central tendency was greater for the eyes compared to the head.

Head rotation

We hypothesized that observers might rotate their heads in line with the stimulus rotation in order to view the image horizon at a more canonical orientation (i.e., upright) in their visual field. Here, head rotation is defined as the rotation around the viewing axis (in our particular experimental setup, this was rotation around the *z*-axis, or roll). Because the head, unlike the eyes, takes more time to move, we examined head rotation over the course of the trial. Head rotations in angular degrees (around the viewing axis) at each fixation were subtracted from the head rotation position at the start of the trial (i.e., at fixation 1). Figure 6 shows these rotations over the course of 15 fixations for each stimulus rotation. To statistically examine the rate of head rotation, the first 15 head rotations were analyzed using mixed-effects quadratic growth-curve analyses separately for each experiment, stimulus type, and stimulus rotation. These analyses generally yielded regression slopes significantly different from zero. For rotated landscape scenes (45°, 90°, 135°, 215°, 270°, and 315°), slope magnitude exceeded 0.1°/fixation. For the other landscape scene rotations and for the fractal scenes, the slopes were either non-significant or smaller than 0.1°/fixation.

It is clear from Figure 6 that many (but not all) participants rotated their head and that when they did so their head rotated quite substantially in line with the stimulus rotation for landscape images (e.g., maximum 35°), but there was relatively little rotation for the fractal images. To examine differences in the amount of head rotation among image apertures, stimulus type, and stimulus rotation, a three (experiment) by two (stimulus type) by five (stimulus rotation) mixed analysis of variance was conducted on the maximum
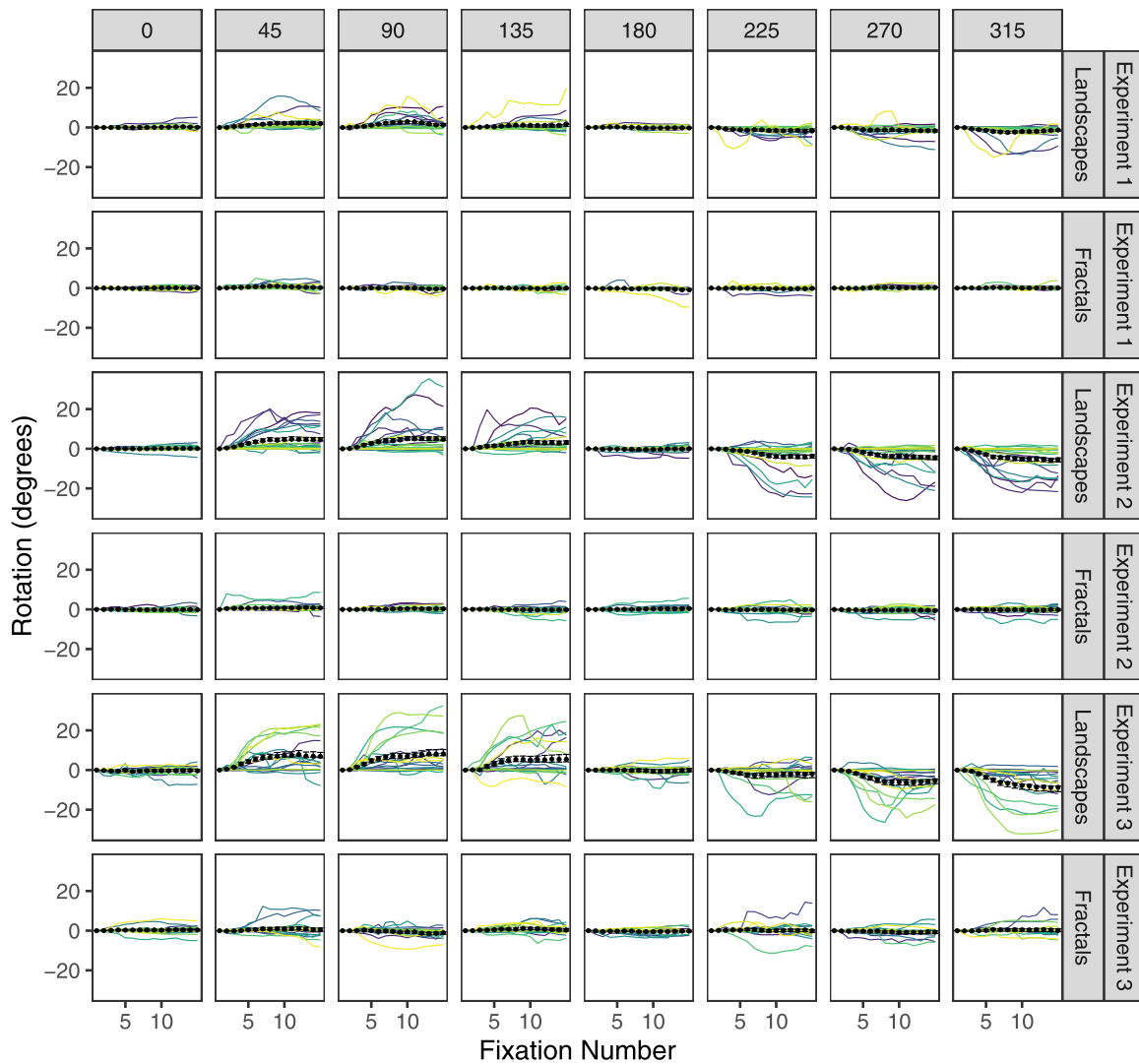
Figure 7. Cumulative head rotation across fixation index for each image rotation. Individual subjects are plotted as separate lines, and overall subject means are plotted as black lines with error bars representing the standard error of the mean.

of the absolute value of head rotations across the 15 fixations presented in Figure 7. There was a main effect of experiment, $F(2, 56) = 5.98$, $p < 0.005$, $\eta_p^2 = 0.18$. Follow-up Bonferroni-corrected comparisons of maximum head rotation across experiments revealed significantly more head rotation in Experiment 3 compared to Experiment 1, $t(56) = 3.43$, $p < 0.005$, $d = 0.92$, and marginally more head rotation in Experiment 3 compared to Experiment 2, $t(56) = 2.27$, $p = 0.07$, $d = 0.61$. There was no significant difference in head rotation between Experiments 1 and 2, $t(56) = 1.36$, $p = 0.37$, $d = 0.36$, and there were no significant interactions among experiment and any other factors (all $F < 2.27$, $p > 0.05$).

There was a main effect of stimulus type, $F(1, 56) = 29.60$, $p < 0.001$, $\eta_p^2 = 0.35$, and stimulus rotation, $F(3.7, 207) = 16.89$, $p < 0.001$, $\eta_p^2 = 0.23$,

qualified by a stimulus type by stimulus rotation interaction, $F(4.03, 225.59) = 11.32$, $p < 0.001$, $\eta_p^2 = 0.17$. Bonferroni-corrected pairwise comparisons of maximum head rotation between stimulus type at each stimulus rotation revealed significantly greater head rotation for landscapes compared to fractals for all stimulus rotations (all $t > 3.53$, $p < 0.001$, $d > 0.21$), except for stimulus rotations 0° and 180° (all $t < 1$).

Figure 6 also reveals that some participants were rotators, but others were not. For each experiment, stimulus type, and stimulus rotation (i.e., each cell in Figure 7), we counted the number of participants with mean head rotation greater than 2.5 standard deviations outside the mean for that cell. Based on this criterion, 39% of participants in Experiment 1, 35% of participants in Experiment 2, and 33% of participants in Experiment 3 were identified as rotators.

Summary

A second key question of the present work was whether and to what extent observers might rotate their heads when allowed to do so. We found clear evidence that, indeed, people do rotate their heads, but that they do so significantly more in line with the horizon of landscape images, with relatively little head rotation occurring for fractals. Observers rotated their heads more in Experiment 3 compared to Experiments 1 and 2. This latter finding is interesting because rotating both the image and frame is something that is not practically possible in current lab-based eye tracking experiments. One striking observation is that some observers were head rotators, but others were not (see Figure 7). This strong individual difference is a fruitful avenue for further research. For example, the extent to which people recruit their head for exploration may depend on several factors, including musculature constraints (Solman & Kingstone, 2014), or even curiosity (Risko, Anderson, Lanthier, & Kingstone, 2012), which may have functional consequences (Goldin-Meadow, 1999; Risko et al., 2014).

Discussion

In the present work, we presented participants with natural landscape and fractal scenes in a fully immersive virtual environment. The images were rotated and participants' gaze and head positions on the scenes, as well as head rotations (roll), were tracked while they looked at the scenes in preparation for a later memory test. One aim was to replicate and extend the work of Foulsham and colleagues (2008, 2010) in an environment that eliminated any bias in eye movements that could be due to extraneous environmental factors, such as presenting different aperture shapes and rotated image content within a fixed rectangular laboratory monitor. In Experiment 1, images were presented on a square frame such that the frame always remained upright and only the image content was rotated, whereas in Experiment 2 images were presented in a circular aperture to eliminate directional cues associated with a square or rectangular frame. Experiment 3 extended this work to a situation where both a square image frame and content rotated, akin to rotating the entire computer monitor along with the scene content.

We found that observers explored these scenes by moving their eyes and head (albeit the head moved less than the eyes) and that image aperture had minimal effects on general gaze measures, with no significant difference in fixation numbers and fixation durations across experiments. There was, however, less central tendency in Experiment 3 compared to Experiments 1 and 2. This is likely due to the fact that, although the different apertures subtended approximately the same visual angle, there was more image content visible in the corners of the square aperture in Experiment 3, as no image cropping was required to keep rotated content in an un-rotated square (see Figure 2).

We found that the most frequent saccade direction was along the horizon of landscape images, replicating the work of Foulsham and colleagues (2008) and conforming to the predictions (but not the findings) of Foulsham and Kingstone (2010). Both square (Experiment 1) and circular (Experiment 2) apertures produced a predominant bias for observers to make saccades parallel to the landscape horizon. The next most frequent saccade direction was orthogonal to the horizon. Foulsham and Kingstone (2010) suggested that the task ("remember the pictures for a later memory test"), the interleaved presentation of the images, or the circular aperture may have made vertical saccades more likely. Our data point to the likelihood that other, extraneous factors unique to their experimental setup may have nudged the predominance of vertical saccades in that case, but not in earlier work (Foulsham et al., 2008). When these were eliminated by virtue of the VR setup, we found that the majority of saccades were made in the cardinal directions of the landscape images, with a preferential bias for the horizon. This was true for both square and circular apertures and whether or not the square frame rotates with the image.

Saccade amplitudes mirrored the saccade direction biases observed in Experiments 1, 2, and 3, where the longest saccade amplitude was usually made along the horizon of landscape images. Unlike the saccade direction biases, saccade amplitudes were not significantly larger in the axis orthogonal to the image horizons. It has been shown that the shape of the information visible at fixation affects the amplitudes of horizontal and vertical saccades (e.g., Foulsham, Teszka, & Kingstone, 2011). For example, on a wide (e.g., 4:3 or 16:9) monitor, the amplitudes of horizontal saccades tend to be larger than those of the vertical saccades. The images and frames used here had a 1:1 aspect ratio; thus, the larger horizontal saccades cannot be explained by the aspect ratio.

Interestingly, for fractal images that presumably are isotropic and have no discernible upright, we found that saccade biases varied in line with fractal rotations. Saccades were more likely to be made along the cardinal directions relative to the fractal horizons, and the longest amplitude saccades were made along the horizon of fractal images. These directional effects observed for fractal images are unusual. In Foulsham and Kingstone (2010), observers made saccades horizontally within an egocentric frame of reference. This was important, because it was strong evidence for an image-independent bias. Why did we not observe the same in our work? It is possible that observers in our

study were picking up on some directionality inherent in the fractals that was only apparent due to their relatively large size: 60 × 60 degrees visual angle compared to the 34° × 27° monitor size in Foulsham et al. (2008) and 40° × 31° in Foulsham and Kingstone (2010). Observers may be less likely to use directional information from the clear virtual horizon (the Unity skybox) behind the images due to the paucity of other gravitational cues. Indeed, if the virtual horizon influenced saccade biases, this would be the case across all conditions. Instead, it appeared that observers relied more on directional cues sourced from the images themselves (Rai, Gutiérrez, & Le Callet, 2017; Sitzmann, Serrano, Pavel, Agrawala, Gutierrez, Masia, & Wetzstein, 2018). Regardless, it is clear that the oculomotor system was picking up on some intrinsic, directional structure in the fractal images and tracking the fractal rotations in a manner similar to that for the landscape rotations. Interestingly, as noted below, this was not the case for head rotations.

The bias for saccades to be made along image horizons is well established in fully immersive, 360° panoramic scenes (Bischof, Anderson, Doswell, & Kingstone, in press; Rai et al., 2017; Sitzmann et al., 2018) and videos (Corbillon, De Simone, & Simon, 2017; David, Gutiérrez, Coutrot, Da Silva, & Callet, 2018; Wu, Tan, Wang, & Yang, 2017; Xu, Li, Liu, Deng, & Lu, 2017). Bischof and colleagues (in press) demonstrated that this bias persists when landscape panoramas are rotated in a manner similar to the present work. The general consensus, then, has been that saccades biases are allocentric and tied to image content. In contrast, for fractal 360° scenes, fixation distributions have been found to be isotropic and saccade directions biased in the egocentric cardinal directions. As mentioned earlier, image-independent, intrinsic eye movement biases have been proposed (Foulsham & Kingstone, 2010) but were not demonstrated in the present work. This discrepancy in our findings warrants further investigation. Saccade biases could be tied directly to the directional characteristics of the fractal images either by extracting their low-level features or by asking participants to indicate where they think the horizon might be positioned. In any case, determining under what conditions observers may shift their saccade biases in line with internal or external reference frames is a fruitful avenue for future research.

A second main aim of this work was to investigate the effect of aperture shape, image type, and rotation on head movement. We found that head shift amplitudes were larger in Experiment 3 compared to Experiments 1 and 2, consistent with the idea that no cropping was required to fit rotated images into square or circular frames in Experiment 3. Central tendency was lower in Experiment 1 compared to Experiments 2 and 3, suggesting that observers explored with

the head more when viewing an unrotated square frame. Most importantly, we confirmed our main hypothesis that the head would rotate in line with the horizon in order to bring landscape images into a more cardinal viewing position for the eyes. We found that observers rotated their heads in line with the horizon in landscape images. This is consistent with previous work demonstrating the use of our bodies in cognitive offloading (Risko & Gilbert, 2016), in a process termed "external normalization" (Risko et al., 2014). In addition, we found clear evidence for individual differences in the magnitude of head rotation. Some participants were head rotators and others were not.

The idea of individual differences in the propensity to move the head (not just rotate it) is not new (Delreux, Abeele, Lefevre, & Roucoux, 1991; Fuller, 1992a; Fuller, 1992b; Goldring, Dorris, Corneil, Ballantyne, & Munoz, 1996; Pozzo, Berthoz, & Lefort, 1992). In these works, which was largely concerned with gaze and head shifts to targets via audio cues or simple light displays, researchers noticed that, when left to their own devices, some participants would move their heads and others would not. Fuller (1992b) dubbed his participants either "movers" or "non-movers" but noted that the differences varied along a continuous spectrum. He attributed the differences in head recruitment to an innate behavioral trait that reflects an individual's method of "constructing central nervous coordinate systems" (p. 163). Movers are more likely to choose world-based, allocentric reference frames, whereas non-movers are more likely to rely on intrinsic, or egocentric frames because keeping the head still avoids having to recalibrate internal and external coordinate systems. Whether this is indeed the case has rarely been studied, but later work by Stahl (2001) has demonstrated that it is not likely a kinematic effect but instead may be related to how visual space is mapped at cerebral levels. In the present work, many participants were non-head movers, yet if these non-movers were relying on internal, egocentric reference frames, we might expect their saccades to follow suit, which was not the case. This suggests that Fuller's (1992b) hypothesis is not correct, or that the head and eye movement systems are relying on different coordinate reference frames. This is a fruitful avenue for further research.

Interestingly, we did not observe head rotation in line with the horizon in fractal images. Clearly, the head was not responsive to whatever directionality that the eyes had picked up on in the fractal images. This is strong evidence suggesting that the eyes and head respond to different cues and could potentially be under different control strategies. This contrasts with early work in simple displays, where the head and eyes have generally been tightly coupled (Freedman, 2008; Zangemeister & Stark, 1982).

In the present work, the finding that the head and eyes responded to different cues has several implications. First, it implies that the head may not simply act in service of the eyes, contrary to our predictions and those of earlier work (Land & Hayhoe, 2001; Solman et al., 2017). In particular, saccades in line with a rotated image horizon do not always trigger a head rotation, and the head does not automatically follow gaze behavior. Second, it implies that the head is not necessarily responding to low-level (or mid-level) visual cues that the eyes may be picking up from the fractal images. Rather, the head may be acting in a more deliberate manner, responding to the overtly meaningful component of landscape images (e.g., the canonical up and down). This dovetails with work in more natural tasks, where the head is shown to be under more deliberate, cognitive control. Lee (1999) demonstrated that, in reading, the head (but not the eye) was sensitive to text familiarity; however, in driving it has been shown that the best predictor of a lane change is not an eye movement but a head movement that occurs 2 to 3 seconds beforehand (Doshi & Trivedi, 2012).

Taken together, our findings confirm that head movements are an important aspect of human attention. When participants were allowed to freely move their heads, they did so (despite wearing a relatively heavy VR headset) but in a manner that was sensitive to the content of the visual environment. Our findings reveal that saccade biases observed in standard head-restrained eye tracking extend to situations where observers are allowed to move their heads. In free movement, observers preferentially explore scenes by making cardinal saccades relative to the coordinate system inherent to the image itself. However, observers rotated their heads in order to more closely align the allocentric image coordinate system to the egocentric one, albeit only for landscape scenes and not fractals. Understanding the recruitment of, and relationship between, allocentric and egocentric coordinate systems, as well as the control strategies that govern eye and head movement control, represents a fruitful avenue for future VR research.

Conclusions

In summary, we demonstrated that saccades are generally made parallel to the horizon of both landscape and fractal images presented in a virtual environment when they are rotated within a square or circular aperture and when the entire image frame and content are rotated (equivalent to rotating an entire computer monitor). This was true for both landscape and fractal images, suggesting that the eyes were sensitive to

directional content in both. Interestingly, this was not the case for head rotations; the head rotated in line with the landscape but not the fractal horizons. This suggests that the head need not always operate in service of the eyes and as such may be sensitive to a different control system.

Keywords: saccade biases, head movements, head rotation, virtual reality

Corresponding author: Nicola C. Anderson.
Email: nicola.anderson@ubc.ca.
Address: Department of Psychology, University of British Columbia, Vancouver, BC, Canada.

Footnote

[1]Post hoc power analysis for this interaction, based on an observed η_p^2 of 0.05, with three groups and 11 measurements and using default settings in G*Power (Erdfelder et al., 1996), revealed a power of 0.13.

References

Anderson, N. C., & Donk, M. (2017). Salient object changes influence overt attentional prioritization and object-based targeting in natural scenes. *PLoS One, 12*(2), e0172132.

Anderson, N. C., Ort, E., Kruijne, W., Meeter, M., & Donk, M. (2015). It depends on *when* you look at it: Salience influences eye movements in natural scene viewing and search early in time. *Journal of Vision, 15*(5):9, 1–22, https://doi.org/10.1167/15.5.9.

Backhaus, D., Engbert, R., Rothkegel, L. O. M., & Trukenbrod, H. A. (2020). Task-dependence in scene perception: Head unrestrained viewing using mobile eye-tracking. *Journal of Vision, 20*(5):3, 1–21, https://doi.org/10.1167/jov.20.5.3.

Barnes, G. R. (1979). Vestibulo-ocular function during co-ordinated head and eye movements to acquire

visual targets. *The Journal of Physiology, 287*(1), 127–147.

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

Birmingham, E., Bischof, W. F., & Kingstone, A. (2009). Saliency does not account for fixations to eyes within social scenes. *Vision Research, 49*(24), 2992–3000.

Bischof, W. F., Anderson, N. C., Doswell, M. T., & Kingstone, A. (in press). Visual exploration of omni-directional panoramic scenes. *Journal of Vision*.

Blignaut, P. (2009). Fixation identification: The optimum threshold for a dispersion algorithm. *Attention, Perception, & Psychophysics, 71*(4), 881–895, https://doi.org/10.3758/APP.71.4.881.

Borji, A., & Itti, L. (2013). State-of-the-art in visual attention modeling. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 35*(1), 185–207.

Brandt, H. F. (1945). *The psychology of seeing*. New York: Philosophical Library.

Buswell, G. T. (1935). *How people look at pictures*. Chicago, IL: University of Chicago Press.

Corbillon, X., De Simone, F., & Simon, G. (2017). 360-degree video head movement dataset. In: *Proceedings of the 8th ACM on Multimedia Systems Conference* (pp. 199–204). New York: Association for Computing Machinery.

David, E. J., Gutiérrez, J., Coutrot, A., Da Silva, M. P, & Callet, P. L. (2018). A dataset of head and eye movements for 360 videos. In: *Proceedings of the 9th ACM Multimedia Systems Conference* (pp. 432–437). New York: Association for Computing Machinery.

Delreux, V., Abeele, S. V., Lefevre, P., & Roucoux, A. (1991). Eye-head coordination: Influence of eye position on the control of head movement amplitude. In: J. Paillard (Ed.). *Brain and Space* (pp. 38–48). London: Oxford University Press.

Doshi, A., & Trivedi, M. M. (2012). Head and eye gaze dynamics during visual attention shifts in complex environments. *Journal of Vision, 12*(2):9, 1–16, https://doi.org/10.1167/12.2.9.

Einhäuser, 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, 1–19, https://doi.org/10.1167/8.2.2.

Erdfelder, E., Faul, F., & Buchner, A. (1996). GPOWER: A general power analysis program. *Behavior Research Methods, Instruments, & Computers, 28*(1), 1–11.

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., & Kingstone, A. (2013). Optimal and preferred eye landing positions in objects and scenes. *Quarterly Journal of Experimental Psychology, 66*(9), 1707–1728.

Foulsham, T., Kingstone, A., & Underwood, G. (2008). Turning the world around: Patterns in saccade direction vary with picture orientation. *Vision Research, 48*(17), 1777–1790.

Foulsham, T., Teszka, R., & Kingstone, A. (2011). Saccade control in natural images is shaped by the information visible at fixation: Evidence from asymmetric gaze-contingent windows. *Attention, Perception, & Psychophysics, 73*(1), 266–283.

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*(2):6, 1–17, https://doi.org/10.1167/8.2.6.

Foulsham, T., Walker, E., & Kingstone, A. (2011). The where, what and when of gaze allocation in the lab and the natural environment. *Vision Research, 51*(17), 1920–1931.

Freedman, E. G. (2008). Coordination of the eyes and head during visual orienting. *Experimental Brain Research, 190*(4), 369.

Fuller, J. (1992a). Comparison of head movement strategies among mammals. In: A. Berthoz, W. Graf, & P. P. Vidal (Eds.). *The head-neck sensory motor system* (pp. 101–112). New York: Oxford University Press.

Fuller, J. (1992b). Head movement propensity. *Experimental Brain Research, 92*(1), 152–164.

Gilchrist, I. D., & Harvey, M. (2006). Evidence for a systematic component within scan paths in visual search. *Visual Cognition, 14*(4–8), 704–715.

Goldin-Meadow, S. (1999). The role of gesture in communication and thinking. *Trends in Cognitive Sciences, 3*(11), 419–429.

Goldring, J. E., Dorris, M. C., Corneil, B. D., Ballantyne, P. A., & Munoz, D. R. (1996). Combined eye-head gaze shifts to visual and auditory targets in humans. *Experimental Brain Research, 111*(1), 68–78.

Henderson, J. M., Brockmole, J. R., Castelhano, M. S., & Mack, M. (2007). Visual saliency does not account for eye movements during visual search in real-world scenes. In: R. P. G. Van Gompel, M. H. Fischer, W. S. Murray, & R. L. Hill (Eds.).

Eye movements: A window on mind and brain (pp. 537–562). Amsterdam: Elsevier.

Hessels, R. S., Niehorster, D. C., Nyström, M., Andersson, R., & Hooge, I. T. (2018). Is the eye-movement field confused about fixations and saccades? A survey among 124 researchers. *Royal Society Open Science, 5*(8), 180502.

Hooge, I. T., Hessels, R. S., Niehorster, D. C., Diaz, G. J., Duchowski, A. T., & Pelz, J. B. (2019). From lab-based studies to eye-tracking in virtual and real worlds: Conceptual and methodological problems and solutions. *Journal of Eye Movement Research, 12*(7), https://doi.org/10.16910/jemr.12.7.8.

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.

Itti, L., & Koch, C. (2001). Computational modelling of visual attention. *Nature Reviews Neuroscience, 2*(3), 194–203.

Kingstone, A., Smilek, D., & Eastwood, J. D. (2008). Cognitive ethology: A new approach for studying human cognition. *British Journal of Psychology, 99*(3), 317–340.

Komogortsev, O. V., Gobert, D. V., Jayarathna, S., Koh, D., & Gowda, S. (2010). Standardization of automated analyses of oculomotor fixation and saccadic behaviors. *IEEE Transactions on Biomedical Engineering, 57*(11), 2635–2645, https://doi.org/10.1109/TBME.2010.2057429.

Kothari, R., Yang, Z., Kanan, C., Bailey, R., Pelz, J. B., & Diaz, G. J. (2020). Gaze-in-wild: A dataset for studying eye and head coordination in everyday activities. *Scientific Reports, 10*, 2539.

Land, M. F., & Hayhoe, M. (2001). In what ways do eye movements contribute to everyday activities? *Vision Research, 41*(25), 3559–3565.

Lee, C. (1999). Eye and head coordination in reading: Roles of head movement and cognitive control. *Vision Research, 39*(22), 3761–3768.

Lenth, R., Length, R., Singmann, H., Love, J., Buerkner, P., & Herve, M. (2019). emmeans: Estimated marginal means, aka least-squares means. Retrieved from https://CRAN.R-project.org/package=emmeans.

Matthis, J. S., Yates, J. L., & Hayhoe, M. M. (2018). Gaze and the control of foot placement when walking in natural terrain. *Current Biology, 28*(8), 1224–1233.

Nuthmann, A., & Henderson, J. M. (2010). Object-based attentional selection in scene viewing. *Journal of Vision, 10*(8):20, 1–19, https://doi.org/10.1167/10.8.20.

Oliva, A., & Torralba, A. (2006). Building the gist of a scene: The role of global image features in recognition. *Progress in Brain Research, 155*, 23–36.

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.

Pozzo, T., Berthoz, A., & Lefort, L. (1992). Head kinematics during complex movements. In: A. Berthoz, W. Graf, & P. P. Vidal (Eds.). *The head-neck sensory motor system* (pp. 587–590). New York: Oxford University Press.

Rai, Y., Gutiérrez, J., & Le Callet, P. (2017). A dataset of head and eye movements for 360 degree images. In: *Proceedings of the 8th ACM on Multimedia Systems Conference* (pp. 205–210). New York: Association for Computing Machinery.

Risko, E. F., Anderson, N. C., Lanthier, S., & Kingstone, A. (2012). Curious eyes: Individual differences in personality predict eye movement behavior in scene-viewing. *Cognition, 122*(1), 86–90.

Risko, E. F., & Gilbert, S. J. (2016). Cognitive offloading. *Trends in Cognitive Sciences, 20*(9), 676–688.

Risko, E. F., Medimorec, S., Chisholm, J., & Kingstone, A. (2014). Rotating with rotated text: A natural behavior approach to investigating cognitive offloading. *Cognitive Science, 38*(3), 537–564.

Risko, E. F., Richardson, D. C., & Kingstone, A. (2016). Breaking the fourth wall of cognitive science: Real-world social attention and the dual function of gaze. *Current Directions in Psychological Science, 25*(1), 70–74.

Salvucci, D. D., & Goldberg, J. H. (2000). Identifying fixations and saccades in eye-tracking protocols. In *Proceedings of the EyeTracking Research and Applications Symposium* (pp. 71–78). New York: ACM Press.

Singmann, H., Bolker, B., Westfall, J., Aust, F., & Ben-Shachar, M. S. (2019). afex: Analysis of factorial experiments. Retrieved from https://CRAN.R-project.org/package=afex.

Sitzmann, V., Serrano, A., Pavel, A., Agrawala, M., Gutierrez, D., Masia, B., . . . Wetzstein, G. (2018). How do people explore virtual environments? *IEEE Transactions on Visualization and Computer Graphics, 24*(4), 1633–1642.

Solman, G. J., Foulsham, T., & Kingstone, A. (2017). Eye and head movements are complementary in visual selection. *Royal Society Open Science, 4*(1), 160569.

Solman, G. J., & Kingstone, A. (2014). Balancing energetic and cognitive resources: Memory use during search depends on the orienting effector. *Cognition, 132*(3), 443–454.

Stahl, J. S. (2001). Eye-head coordination and the variation of eye-movement accuracy with orbital eccentricity. *Experimental Brain Research, 136*(2), 200–210.

't Hart, B. M., Schmidt, H. C. E. F., Roth, C., & Einhäuser, W. (2013). Fixations on objects in natural scenes: Dissociating importance from salience. *Frontiers in Psychology, 4*, 455.

't Hart, B. M., Vockeroth, B., Schumann, J., Bartl, F., Schneider, K., König, E. P., . . . Einhäuser, W. (2009). Gaze allocation in natural stimuli: Comparing free exploration to head-fixed viewing conditions. *Visual Cognition, 17*(6–7), 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, 1–17, https://doi.org/10.1167/7.14.4.

Tatler, B. W., & Vincent, B. T. (2008). Systematic tendencies in scene viewing. *Journal of Eye Movement Research, 2*(2), https://doi.org/10.16910/jemr.2.2.5.

Torralba, A., Oliva, A., Castelhano, M. S., & Henderson, J. M. (2006). Contextual guidance of eye movements and attention in real-world scenes: The role of global features in object search. *Psychological Review, 113*(4), 766–786.

Unity Technologies. (2020). Unity for all. Retrieved from https://unity3d.com.

Vincent, B. T., Baddeley, R., Correani, A., Troscianko, T., & Leonards, U. (2009). Do we look at lights? Using mixture modelling to distinguish between low- and high-level factors in natural image viewing. *Visual Cognition, 17*(6–7), 856–879.

Wickham, H. (2016). *ggplot2: Elegant graphics for data analysis*. New York: Springer-Verlag.

Wu, C., Tan, Z., Wang, Z., & Yang, S. (2017). A dataset for exploring user behaviors in VR spherical video streaming. In: *Proceedings of the 8th ACM on Multimedia Systems Conference* (pp. 193–198). New York: Association for Computing Machinery.

Xu, M., Li, C., Liu, Y., Deng, X., & Lu, J. (2017). A subjective visual quality assessment method of panoramic videos. In: *2017 IEEE International Conference on Multimedia and Expo (ICME)* (pp. 517–522). Piscataway, NJ: Institute of Electrical and Electronics Engineers.

Yarbus, A. L. (1967). *Eye movements during perception of complex objects*. Boston, MA: Springer.

Zangemeister, W. H., & Stark, L. (1982). Types of gaze movement: Variable interactions of eye and head movements. *Experimental Neurology, 77*(3), 563–577.