Flipping the world upside down: Using eye tracking in virtual reality to study visual search in inverted scenes

Julia Beitner
Department of Psychology,
Goethe University Frankfurt, Germany
Corresponding author
beitner@psych.uni-frankfurt.de

Jason Helbing
Department of Psychology,
Goethe University Frankfurt,
Germany

Dejan Draschkow
Department of Experimental Psychology,
University of Oxford, UK
Oxford Centre for Human Brain Activity,
Wellcome Centre for Integrative
Neuroimaging, Department of Psychiatry,
University of Oxford, UK

Erwan J. David
Department of Psychology,
Goethe University Frankfurt,
Germany

Melissa L.-H. Võ
Department of Psychology,
Goethe University Frankfurt,
Germany

Image inversion is a powerful tool for investigating cognitive mechanisms of visual perception. However, studies have mainly used inversion in paradigms presented on two-dimensional computer screens. It remains open whether disruptive effects of inversion also hold true in more naturalistic scenarios. In our study, we used scene inversion in virtual reality in combination with eye tracking to investigate the mechanisms of repeated visual search through three-dimensional immersive indoor scenes. Scene inversion affected all gaze and head measures except fixation durations and saccade amplitudes. Our behavioral results, surprisingly, did not entirely follow as hypothesized: While search efficiency dropped significantly in inverted scenes, participants did not utilize more memory as measured by search time slopes. This indicates that despite the disruption, participants did not try to compensate the increased difficulty by using more memory. Our study highlights the importance of investigating classical experimental paradigms in more naturalistic scenarios to advance research on daily human behavior.

Keywords: Eye movements, eye tracking, virtual reality, scene perception, visual search, incidental memory, scene inversion

Received November 14, 2022; Published March 31, 2023.
Citation: Beitner, J., Helbing, J., Draschkow, D., David, E. J., & Võ, M. L.-H. (2023). Flipping the world upside down: Using eye tracking in virtual reality to study visual search in inverted scenes. *Journal of Eye Movement Research*, *15*(3):5.
Digital Object Identifier: 10.16910/jemr.15.3.5
ISSN: 1995-8692

Introduction

Research of visual perception often employs up-down inversion of visual stimuli as a powerful manipulation to investigate mechanisms of visual processing. The major advantage of image inversion is that it manipulates the whole image, while global image characteristics such as luminance, contrast, and color remain intact (Hayes &

Henderson, 2022; Kelley et al., 2003). Especially the field of face perception has utilized face inversion to probe mechanisms of face recognition (Taubert et al., 2011; Valentine, 1988; Yin, 1969).

In the study of scene perception, it has been shown that the inversion of scenes has disruptive effects on the top-down extraction of meaning and context (Brockmole & Henderson, 2006; Kelley et al., 2003; Rock, 1974; Shore & Klein, 2000), specifically on semantic guidance (Hayes & Henderson, 2022). Inversion further affects extraction of scene gist (Koehler & Eckstein, 2015; Lauer et al., 2020), scene categorization (Walther et al., 2009), scene memory (Beighley & Intraub, 2016; Meng & Potter, 2008), object recognition (Lauer et al., 2020), and eye movement behavior (Anderson et al., 2020; Foulsham et al., 2008).

Scene inversion effects have also been demonstrated for visual search (Koehler & Eckstein, 2015). In Koehler and Eckstein's study, participants had to decide whether a computer mouse was present or absent in two-dimensional images of office scenes. A trial was terminated either after a certain number of fixations (1, 2, or 3) or after 3 seconds, and scenes were either presented upright or inverted on a computer screen. Koehler and Eckstein found that performance dropped when scenes were inverted, that is, overall hit rate was significantly lower for inverted scenes. Eye movement analyses further revealed that the average distance of fixations to the target (if present) or its expected location (in case of absence) was larger for inverted than for upright scenes. Koehler and Eckstein concluded that scene inversion ultimately disrupts guidance in eye movement behavior.

Here, our aim was to test whether these findings generalize to immersive 360-degree visual search and to demonstrate the feasibility of studying scene inversion effects on behavior, including eye movements, in virtual reality (VR). By embracing more unconstrained and naturalistic task settings, VR promises to increase the external validity of findings (Draschkow et al., 2021, 2022; Helbing et al., 2020, 2022; Parsons, 2015). In some cases, effects demonstrated in laboratory setups were weaker or absent when tested in more realistic settings which often engage behavior that is multimodal, immersive, and self-referential (David, Beitner, et al., 2020; Johnsdorf et al., 2023; Kisker et al., 2021; Li et al., 2016; Rubo et al., 2021; Schöne et al., 2019, 2021; Zhang & Pan, 2022). To bridge the gap between the need for high experimental control and naturalistic scenarios, VR offers a promising solution, combining both realism and control (Draschkow, 2022; Parsons, 2015; Tarr & Warren, 2002). Moreover, when comparing VR to experiments in the real world, VR further offers an easy implementation of setups that would physically not be feasible in the real world such as presenting a variety of environments in immediate succession, objects floating in mid-air, participants effortlessly moving objects that would be heavy in reality, or presenting whole environments flipped upside down. Last but not least, one great advantage of tracking eye movements in VR compared to real-world eye tracking is the ability to use the immediately available eye tracking data to directly change the visual input online. This enables interesting new experimental setups such as gaze contingent paradigms in VR (David, Beitner, et al., 2020; David et al., 2021, 2022).

Due to the aforementioned reasons, we decided to investigate the effects of scene inversion on visual search in VR. Previous studies using scene inversion typically found disruptive effects on many levels of visual processing (Beighley & Intraub, 2016; Brockmole & Henderson, 2006; Hayes & Henderson, 2022; Kelley et al., 2003; Koehler & Eckstein, 2015; Lauer et al., 2020; Meng & Potter, 2008; Shore & Klein, 2000). Visual search through scenes is a very efficient process that highly depends on semantic guidance while relying less on episodic memory (Boettcher et al., 2018; David et al., 2021; Draschkow & Võ, 2017; Helbing et al., 2022; Võ, 2021; Võ et al., 2019; Võ & Wolfe, 2012, 2013). Scene inversion, however, has been found to actively disrupt this semantic guidance (Hayes & Henderson, 2022; Koehler & Eckstein, 2015).

Since higher task difficulty as well as lack of access to semantic guidance often lead to more memory usage, as indicated by steeper response time slopes across search trials (Anderson & Lee, 2023; Draschkow et al., 2021; Helbing et al., 2022; Li et al., 2016; Võ & Wolfe, 2013), we expected to see worse performance as well as increased memory usage over time (i.e., steeper across-trial search time slopes) in searches through inverted scenes. Using VR in combination with eye tracking allowed a more granular investigation of the eye and head movement dynamics associated with scene inversion effects on visual search. Based on previous research (Foulsham et al., 2008; Hayes & Henderson, 2022), we expected to see more horizontal than vertical saccades and especially a bias for rightward saccades and more head rotations continuously made in the same direction (David et al., 2022) independently of the scene orientation. Given differences in gaze movement distributions, we should further find similar differences in the distributions of head movements as a function of scene orientation since head movements aid eye movements in naturalistic behavior (Einhäuser et al., 2007; Freedman, 2008; Stahl, 1999; von Noorden & Campos, 2002). With the vast amount of data generated by VR studies using eye tracking, our analysis of head and eye movements extends beyond our specific hypotheses and takes on an exploratory nature, providing possibly interesting insights for further research in this domain.

Methods

Participants

In total, 24 naïve German native speakers were recruited via the university's recruiting system. Four participants needed to be excluded, i.e., three participants failed to comply with the task instructions, and one participant aborted participation. The final sample consisted of 20 participants (16 women, four men) with a mean age of 22.3 years (*SD* = 3, range = 18–29 years). All participants had normal or corrected-to-normal vision (contact lenses, no glasses), no history of visual or neurological disorders, and were tested for visual acuity (at least 20/25) and normal color vision as assessed by the Ishihara test. All participants volunteered, gave informed consent, and were compensated with course credit or 8 €/h. The experimental procedure conformed to the Declaration of Helsinki and was approved by the local ethics committee of the Faculty of Psychology and Sport Sciences (2014-106R1) at Goethe University Frankfurt.

Apparatus

During the experiment, participants wore an HTC Vive head-mounted display (HMD) equipped with a Tobii eye tracker and held an HTC Vive controller in their dominant hand. The two 1080 × 1200 px OLED screens inside the HMD have a refresh rate of 90 Hz and a combined field of view of approximately 100° (horizontally) × 110° (vertically). The integrated Tobii eye tracker recorded eye movements binocularly with a refresh rate of 120 Hz and a spatial accuracy below 1.1° within a 20° window centered in the viewports. The experimental procedure was implemented in C# in the Unity 3D game engine (version 2017.3) using SteamVR (version 1.10.26) and Tobii eye-tracking software libraries (version 2.13.3) on a computer operated with Windows 10.

Stimuli

As in previous studies (Beitner et al., 2021; David, Beitner, et al., 2020; David et al., 2021; Helbing et al., 2022), we used a set of ten in-house developed three-dimensional virtual indoor scenes, two each from five different room categories: bathroom, kitchen, living room, bedroom, and office (see Figure 1). Every scene measured approximately 380 × 350 × 260 cm (length × width × height), which was fitted to the room size of the laboratory where the experiment took place so that participants could naturally move in virtual rooms without fear of colliding with walls. Each scene contained eight global objects which are large, usually static objects (e.g., sink, stove, couch, bed, desk; also known as "anchor" objects, see Boettcher et al., 2018; Draschkow & Võ, 2017; Helbing et al., 2022; Võ, 2021) and 20 local objects, which are smaller objects that are often interacted with (e.g., toothpaste, toilet paper, remote control, alarm clock, keyboard). An additional scene was used for practice trials,

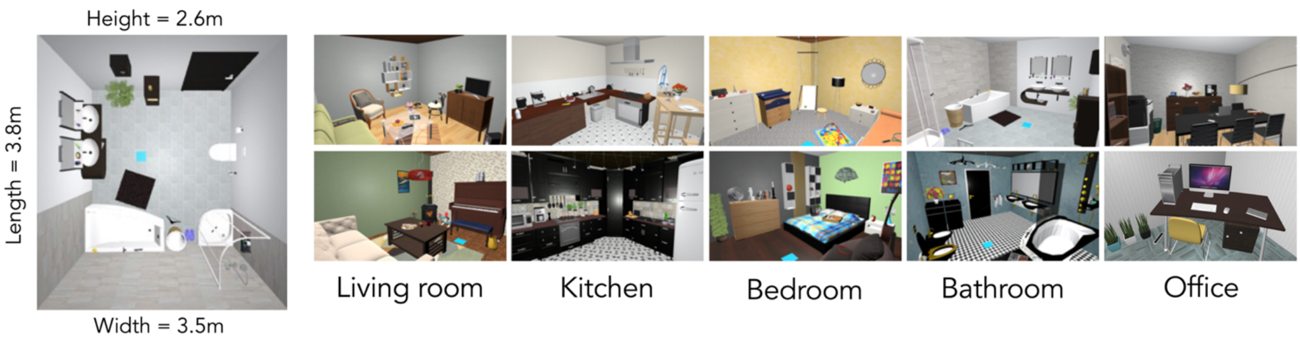

Figure 1. Bird's-eye view of one of the bathrooms and one sample view of all the scenes that were used in the experiment. Blue squares indicate the starting position of the participants and were not visible during searching.

which was a gray neutral room including 10 objects which were considered uncommon for typical living indoor spaces (e.g., hydrant, traffic light, diving helmet) to avoid any interference such as priming of any of the succeeding indoor scenes.

Design

We implemented a repeated visual search task followed by a surprise object recognition task. Critically, we manipulated the inversion of scenes (upright vs. inverted). Upright and inverted scenes were presented randomly interleaved to avoid carry-over effects and the acquisition of strategies over blocks. The condition in which a scene appeared was balanced across participants. Every condition contained five randomly chosen scenes, while each scene was from one of the five unique categories. For each participant individually, 10 out of the 20 local objects were randomly chosen as targets, defining the other 10 as distractors. The search order of objects was randomized. Each scene appeared only once and all 10 target objects were searched in immediate succession. Participants completed a total of 100 searches. The search task was preceded by 10 practice trials in the practice scene, of which five trials were upright and five were inverted. After finishing the visual search task, participants performed a surprise object recognition task to test incidental memory. Results from the object recognition task are not reported here, as eye movements were not recorded and the findings are beyond the scope of the current paper.

Procedure

Upon entering the lab, participants gave informed consent, performed both vision tests, and were familiarized with the HMD and how to use the controller. Next, the eye tracker was calibrated with a nine-point calibration grid. Participants were then instructed on the visual search task and how to navigate within the environment. Participants were told to search as fast and precisely as possible, and that they could move and walk freely within the virtual room during searching. No information regarding strategies was given. Before entering a scene, participants were presented with an empty gray room with instructions written on a wall. Participants then had to position themselves on a blue square on the floor, which was the starting position for the scene and from where they could see most of the objects without obstructions. When the participants were ready, they pulled the trigger button on the controller in their hand to start the trials. Depending on the condition the scene belonged to, the scene was either upright or inverted. When search trials in a scene started, a fixation cross appeared on a large black square in the center of the participants' visual field of view for 1 s, followed by a verbal target cue for 1.5 s that informed participants which object to search for (e.g., "Zahnbürste", toothbrush in German). When the cue disappeared, participants had 15 s to find the target object. Participants completed the trial by pointing a laser beam emerging from their controller at the target and pulling the trigger button with their index finger. In case the selected object was not the target or the timeout was reached, participants heard an error sound. Upon pulling the trigger or after the timeout, the fixation cross cueing the next search appeared. After searching for all 10 target objects successively, participants again entered a gray room with instructions on the wall. The eye tracker was recalibrated after half of the search scenes (i.e., after 50 searches). Participants were allowed to take a break before the eye tracker re-calibration and before the object recognition task if they wanted to. An example video of the task is available at https://osf.io/2ntpj/. Finally, participants completed the object recognition task, which is not reported here. After successful completion of the experiment, participants were debriefed, and exploratively asked how they experienced the different conditions. The whole experiment lasted approximately 1 h. The visual search task reported here lasted approximately 15 minutes.

Analysis

Only correct trials were included in the analysis of behavior and eye movements measured during the visual search task. We counted cases where a wrong object was selected but gaze was on the target object during selection as correct (5.7%). We chose to perform this correction because participants sometimes had difficulties with aiming the laser beam precisely at the target while pulling the trigger button, resulting in a trial initially being logged as incorrect despite the participant actually having found the target object. All incorrect trials, either where the selected object was not the target or where the timeout was reached, were discarded from further analyses (except for the analysis of search accuracy), which led to the removal of 7.4% of the trials. We further excluded trials in which no saccades were detected (2.15%) or in which the first fixation already landed on the target object (3.55%). This happened when participants already positioned themselves towards the target object while only the cue was visible. Fixations and saccades were determined based on the toolbox of the Salient360! Benchmark (David, Lebranchu, et al., 2020; Gutiérrez, David, Coutrot, et al., 2018; Gutiérrez, David, Rai, et al., 2018) using Python (version 3.7.1). To analyze the eye tracking data, we accounted for both head and eye movements and refer to the combination of both as gaze (eye-in-space). By calculating the orthodromic distance and dividing it by the time difference, we obtained the velocity between gaze samples (°/s) which was smoothed with a Savitzky–Golay filter (Nyström & Holmqvist, 2010). Fixations were identified as filtered samples with a velocity of less than 120 °/s. To further investigate saccade directions, we calculated relative and absolute saccade direction angles based on the analyses described in David et al. (2022). Absolute angles are in reference to the longitudinal axis, while relative angles are in reference to the previous saccade angle.

With the preprocessed data, we could further obtain those dependent measures related to visual search (Hollingworth & Bahle, 2019; Malcolm & Henderson, 2009), that is, initiation time (time until the first saccade was made after cue offset), scanning time (time to first target fixation), and verification time (also known as decision time; after fixating the target, time until button press), as well as gaze durations for each object before it became the target. We excluded trials in which initiation time or verification time exceeded a +3 *SD* cut-off, (4.43%). We analyzed our data using the R statistical programming language (version 4.2.2, R Core Team, 2020) with RStudio (version 2022.7.1.554, RStudio Team, 2018).

To investigate search and eye movement behavior, we calculated generalized linear mixed-effects models (GLMMs) and linear mixed-effects models (LMMs) on all variables of interest using the lme4 package (version 1.1-23, Bates et al., 2015). Performing a mixed-models approach allowed us to estimate both between-subject and between-stimulus variance simultaneously which is advantageous compared to traditional F1/F2 analyses of variance (Baayen et al., 2008; Kliegl et al., 2011). We analyzed search accuracy, response time, initiation time, scanning time, verification time, fixation duration (i.e., average duration of all fixations per trial), fixation count, fixated objects count, target refixations (i.e., measure of how many times a participant returned their gaze to the target after a non-target fixation), gaze latitude, saccade amplitudes for gaze and head, as well as gaze and head movement directions.

All variables of interest that were analyzed with LMMs (i.e., response time, initiation time, scanning time, verification time, and fixation duration) were log-transformed to approximate a normal distribution of the residuals and meet assumptions (except gaze latitude, which was left as is). Other variables were analyzed with GLMMs, that is, search accuracy was modeled with a binomial distribution; fixation count, fixated objects count, target refixations, gaze and head directions were modeled using a Poisson distribution; and gaze and head amplitudes were modeled using a Gamma distribution. In all our models, we included search condition (i.e., upright or inverted) as a fixed effect. We added summed gaze durations on the target object from previous trials as a covariate (scaled and centered values) except for the models on gaze and head movement directions because data was averaged across trials. Each model further contained a random effects structure including random intercepts and random slopes for participants and the target object of the respective trial. Note that the models for initiation time, scanning time, verification time, and fixation durations did not include a random slope for participants to avoid a singular model fit (Barr et al., 2013). To ease understanding of the models' architecture, we report the model formulas here (1–4) in Wilkinson notation (Wilkinson & Rogers, 1973). We used difference contrasts to directly compare inverted to upright trials. We obtained *p*-values for LMMs by estimating degrees of freedom with the Satterthwaite method provided by the lmerTest package (version 3.1-2, Kuznetsova et al., 2017). *p*-values for GLMMs were based on asymptotic Wald tests from the lme4 package (version 1.1-23, Bates et al., 2015). All models were fitted with the restricted maximum-likelihood criterion. For each model, we report unstandardized regression coefficients with the t or z statistic (for LMMs or GLMMs, respectively), and the results of a two-tailed test corresponding to a 5% significance level. Full model results are reported in the online supplementary material, including formula notation and R^2-estimates (Hoffman & Rovine, 2007; Nakagawa & Schielzeth, 2013).

Model (1) was used to model search accuracy, fixation count, fixated objects count, target refixations, gaze latitude, gaze and head amplitudes:

$$y \sim 1 + SceneOrientation + GazeDuration + (1 + SceneOrientation \mid Participant) + (1 + SceneOrientation \mid Target) \quad (1)$$

Model (2) was used to model response time, search initiation time, scanning time, verification time and fixation duration:

$$\log(y) \sim 1 + SceneOrientation + GazeDuration + (1 \mid Participant) + (1 + SceneOrientation \mid Target) \quad (2)$$

Model (3) was used to model response time across trials and their interaction:

$$\log(y) \sim 1 + SceneOrientation * Trial + GazeDuration + (1 + SceneOrientation \mid Participant) + (1 + SceneOrientation \mid Target) \quad (3)$$

And lastly, model (4) was used to model gaze and head directions:

$$y \sim 1 + SceneOrientation * AngleCategory + (1 + SceneOrientation \mid Participant) \quad (4)$$

For estimating memory usage over time, we calculated individual slopes across trials on log-transformed response times for both conditions on participant level, and fed them into an LMM with a fixed effects structure including search condition and trial number as well as their interaction and a random effects structure and summed gaze duration as a covariate as mentioned before. We further conducted Bayesian *t*-tests on the participants' individual slopes to estimate the evidence in favor of a null result. In accordance with Kass and Raftery (1995), we interpreted the resulting Bayes factors (BF_{01}) as either indicating evidence in favor of the null hypothesis ($BF_{01} > 3$), or indicating evidence in favor of the alternative hypothesis ($BF_{01} < 0.3$), or as inconclusive evidence ($BF_{01} > 0.3$ and $BF_{01} < 3$). As an example, a $BF_{01} = 3$ can be interpreted in the sense that the data are three times more likely to stem from the same underlying distribution than from different distributions. Bayesian *t*-tests were computed with the BayesFactor package (version 0.9.12-4.2, Morey et al., 2018) and the default settings, that is, with a Cauchy prior with a width of $r = 0.707$ centered on zero. Figures were created with the ggplot2 package in R (version 3.3.0, Wickham, 2016). Dependent measure variable means are reported with their within-subject standard error calculated with the Rmisc package (version 1.5, Hope, 2013). Data and analysis script can be accessed at https://osf.io/2ntpj/.

Results

Behavioral effects

Participants were descriptively but not significantly less accurate in finding objects in inverted scenes compared to upright scenes (upright: 94.40% ± 3.47%; inverted: 90.80% ± 6.06%; $b = -0.81$, $SE = 0.48$, $z = -1.71$, $p = .088$). Gaze durations on an object before it became the target object had a significant effect on search accuracy ($b = 0.55$, $SE = 0.18$, $z = 3.09$, $p = .002$). As can be seen in Figure 2, participants were substantially slower in finding objects when they were searched in inverted scenes (inverted: 3219 ms ± 1417 ms; upright: 2422 ms ± 950 ms; $b = 0.31$, $SE = 0.03$, $t = 10.42$, $p < .001$), while longer gaze durations again helped in finding objects faster ($b = -0.10$, $SE = 0.01$, $t = -6.57$, $p < .001$). These findings already indicate that the inversion manipulation posed a higher difficulty to participants.

When looking at memory usage across search trials, we do see a general negative slope across trials ($b = -0.02$, *SE*

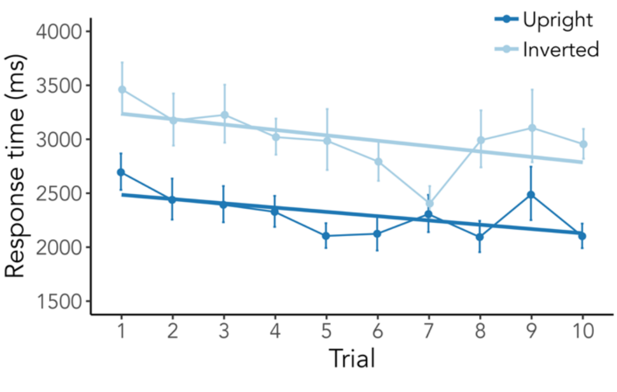

Figure 2. Response times of correct searches from trial 1 to 10 within one scene. Solid straight lines represent regression lines. Solid points indicate means calculated on log-transformed response times which were converted back to their original form for visual purposes, error bars indicate within-subject standard errors.

= 0.01, t = −3.00, p = .003). This indicates that participants did learn something over time within one scene and used this acquired knowledge to speed their search (see Figure 2). This also fits the finding that gaze duration on objects negatively predicted search times (b = −0.08, SE = 0.02, t = −5.28, p < .001). Contrary to our expectations, however, this memory usage did not differ between conditions (b = 0.00, SE = 0.01, t = 0.08, p = .934). This result is further supported by the Bayesian t-test (BF_{01} = 4.27) indicating that participants used memory to an equal extent regardless of whether they searched through upright or inverted scenes.

Eye movement measures of search efficiency

Measuring eye tracking during this immersive task allowed us to decompose the response times into more fine-grained sub-components, that is, initiation time, scanning time, and verification time (Hollingworth & Bahle, 2019; Malcolm & Henderson, 2009). This procedure enabled us to get a deeper insight into which cognitive processes may be affected by the inversion manipulation. As can be seen in Figure 3 and Table 1, scanning time and verification time were longer in inverted scenes, showing that participants had a harder time finding and recognizing objects. This effect was not present for initiation time.

Gaze and head measures

When analyzing fixations, we considered fixation durations, number of fixations within a search trial, the number of objects looked at, and refixations on the target object. While fixation durations did not differ between

Table 1. Modeling results of eye movement measures of search efficiency.

Effect	Estimate b (SE)	t	p
Initiation time model			
Inverted vs. upright	–0.02 (0.06)	–0.37	.715
Gaze duration	0.06 (0.03)	1.96	.050
Scanning time model			
Inverted vs. upright	0.41 (0.05)	8.35	**< .001**
Gaze duration	–0.16 (0.03)	–6.20	**< .001**
Verification time model			
Inverted vs. upright	0.14 (0.02)	5.95	**< .001**
Gaze duration	–0.01 (0.01)	–1.10	.272

Note. Bold p-values indicate significance at an alpha level of 5%.

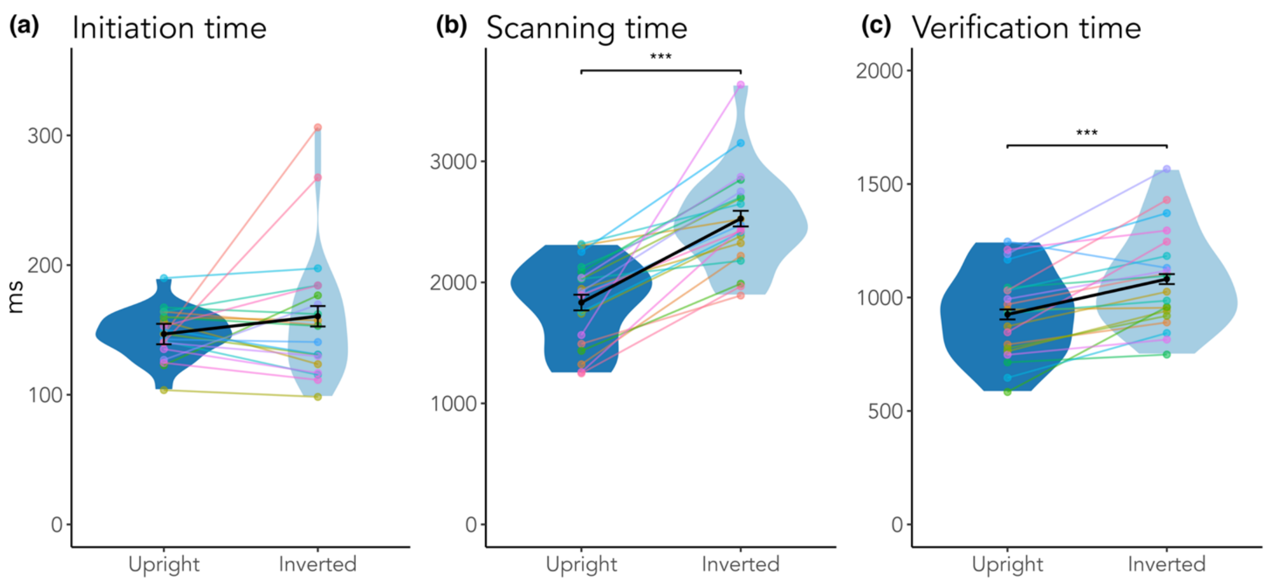

Figure 3. Response time split up into initiation time (a), scanning time (b), and verification time (c). Solid points indicate means, error bars indicate within-subject standard errors, colored points represent individual participants. ***p < .001.

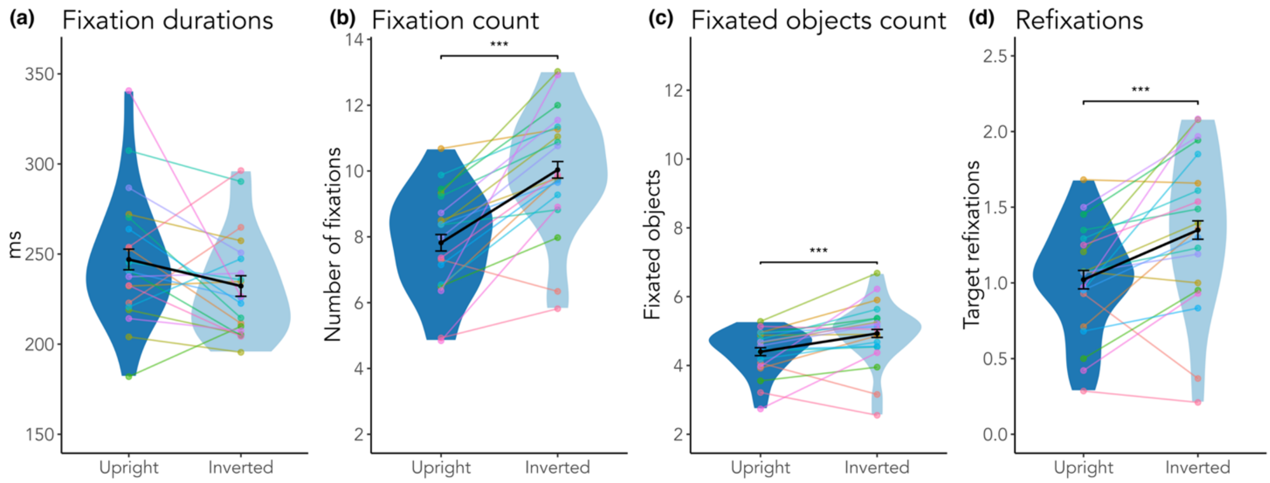

Figure 4. Fixation measures. (a) Fixation durations. (b) Fixation count. (c) Fixated objects count. (d) Refixations. Solid points indicate means, error bars indicate within-subject standard errors, colored points represent individual participants. ****p* < .001.

scene conditions, participants made more fixations in total, fixated more objects on average, and more often refixated the target object during a search trial after having already fixated it previously in inverted scenes (see Figure 4 and Table 2).

Table 2. Modeling results of fixation measures.

Effect	Estimate *b* (*SE*)	*t/z*	*p*
	Fixation duration model		
Inverted vs. upright	–0.02 (0.03)	–0.49	.623
Gaze duration	–0.00 (0.01)	–0.13	.898
	Fixation count model		
Inverted vs. upright	0.32 (0.05)	6.38	**< .001**
Gaze duration	–0.13 (0.01)	–10.64	**< .001**
	Fixated objects count model		
Inverted vs. upright	0.14 (0.04)	3.63	**< .001**
Gaze duration	–0.10 (0.02)	–6.53	**< .001**
	Target refixation model		
Inverted vs. upright	0.37 (0.09)	4.02	**< .001**
Gaze duration	–0.24 (0.03)	–6.86	**< .001**

Note. Bold *p*-values indicate significance on an alpha level of 5%.

Looking at saccade directions, we were interested in whether we can observe changes in movement behavior between upright and inverted scenes. First, we looked at the joint distributions of gaze and head movements in relative and absolute angles as described in David et al. (2022; see Figure 5). While relative saccade angles are in reference to the previously executed saccade (i.e., a saccade at 0° is a forward saccade moving in the same direction as the previous saccade, while a saccade at 180° can be classified as a backward saccade moving against the direction of the previous saccade), absolute angles are in reference to the longitudinal axis (i.e., 0° is right, 180° is left, 90° is up, and 270° is down). Regarding the relative directions in Figure 5 (a), we see that gaze movements went both forwards and backwards, while head movements were mainly directed forwards. Looking at absolute directions in Figure 5 (b), we see that both gaze and head mainly moved in the horizontal plane as indicated by the lighter green along the horizontal axis. To analyze the data in a quantitative way, we categorized absolute gaze and head directions into left (180°), right (0°), up (90°), and down (270°) with a range of 90° spanning each direction (see Figure 6; Anderson et al., 2020; David et al., 2022; Foulsham et al., 2008). For example, gaze directions categorized as *left* fell between 135° and 225°.

There was a main effect of scene orientation ($b = –0.31$, $SE = 0.05$, $z = –6.82$, $p < .001$) on gaze direction (Figure 6 a) in that participants moved their gaze more in inverted than upright scenes. We further found a difference between

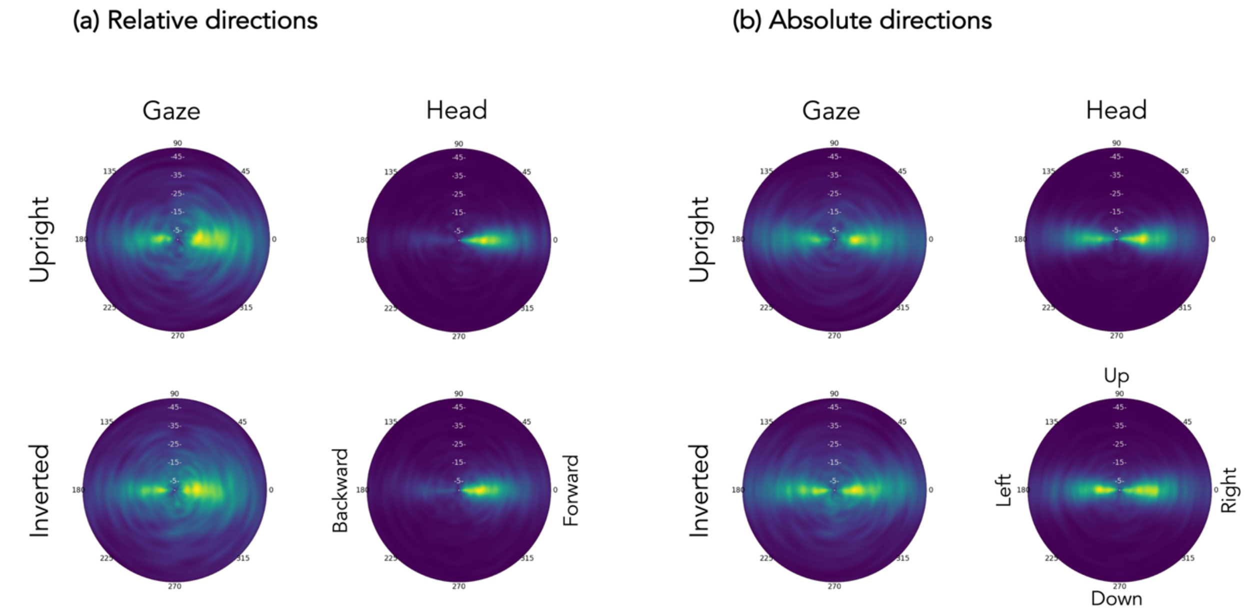

Figure 5. Joint distributions of relative (a) and absolute (b) directions of gaze and head movements and their amplitudes as a function of scene orientation. The lower right plot in each (a) and (b) includes labels for the direction. The lighter the color, the more gaze or head movements were executed into the direction. Radial ticks represent degrees, while ticks from inside out represent saccade amplitudes.

right and left gaze movements (b = 0.16, SE = 0.02, z = 8.47, p < .001), indicating that there was a bias for rightward movements. This bias was not significantly present for up and down movements (b = 0.05, SE = 0.03, z = 1.43, p = .154). There was a non-significant interaction between right and left movements in upright and inverted scenes (b = 0.07, SE = 0.04, z = 1.75, p = .080), and an interaction between up and down movements and scene orientation (b = –0.39, SE = 0.07, z = –5.94, p < .001), demonstrating that scene orientation did have an impact on

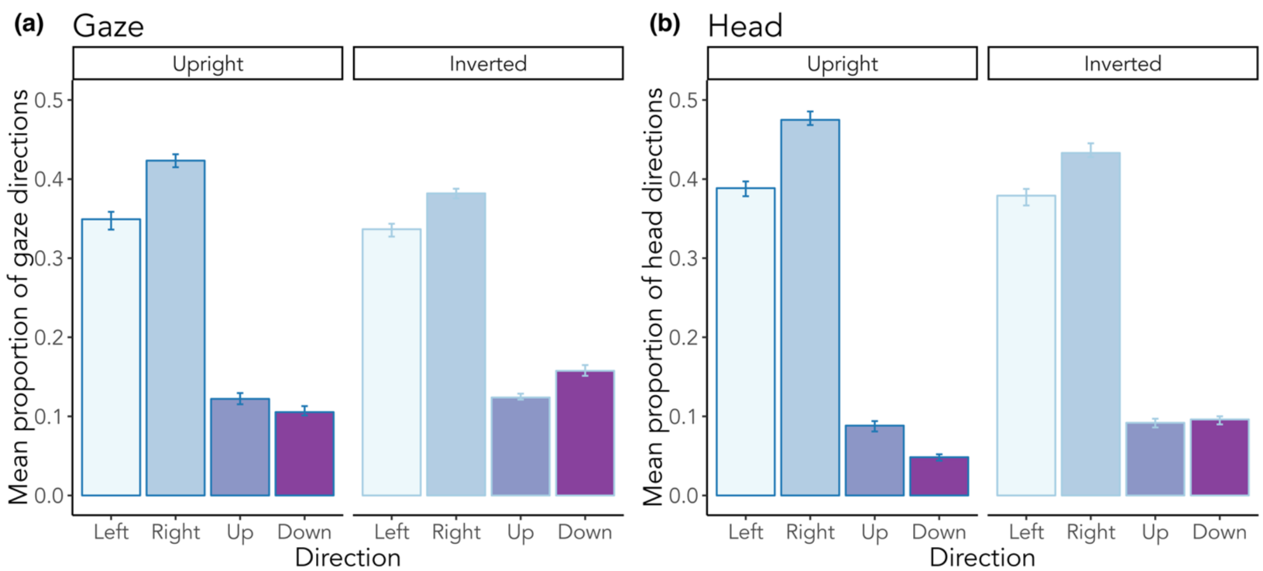

Figure 6. Mean proportion of absolute (a) gaze and (b) head directions as a function of scene orientation. Error bars indicate standard errors.

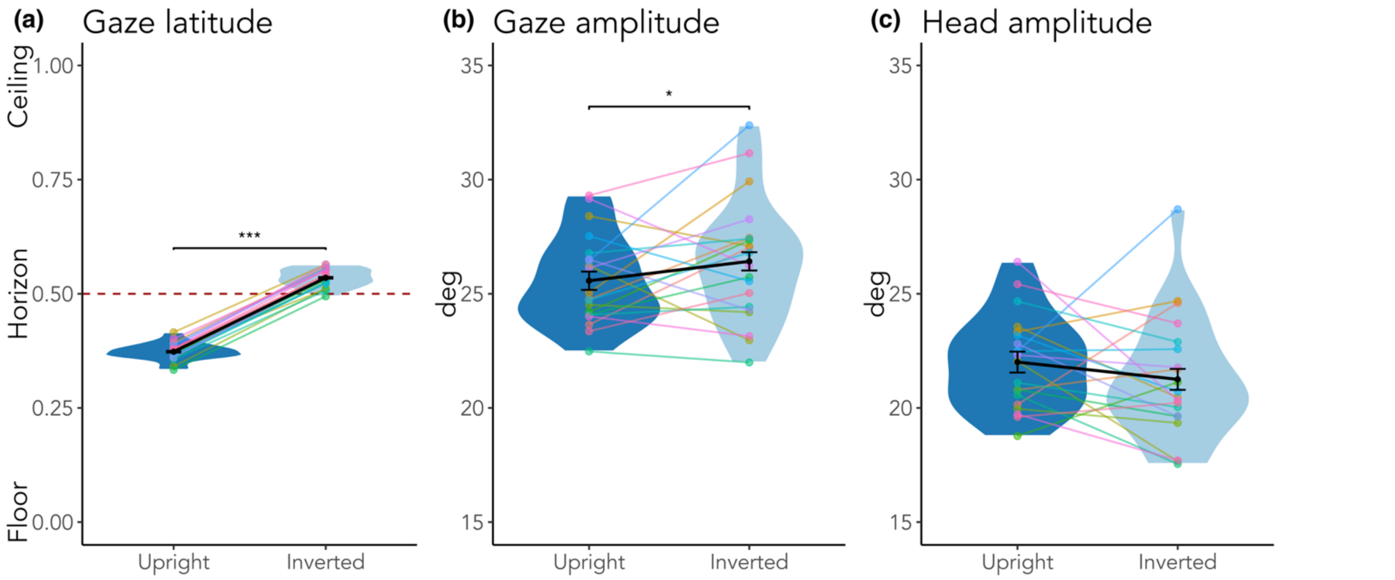

Figure 7. Gaze latitude (a), and amplitudes for (b) gaze and (c) head movements. Solid points indicate means, error bars indicate within-subject standard errors, colored points represent individual participants. In (a), the dashed red line represents the horizon. $^{*}p < .05$, $^{***}p < .001$.

gaze movements in general and especially on up and down movements. These findings basically replicate when looking at isolated head movements (Figure 6 b). Again, more head movements were observed in inverted than in upright scenes ($b = -0.39$, $SE = 0.05$, $z = -8.34$, $p < .001$), as well as a bias for rightward movements ($b = 0.17$, $SE = 0.02$, $z = 9.41$, $p < .001$). In addition, scene inversion again non-significantly interacted with right-left-movements ($b = 0.07$, $SE = 0.04$, $z = 1.90$, $p = .057$) and significantly with up-down-movements ($b = -0.65$, $SE = 0.08$, $z = -7.66$, $p < .001$). Regarding gaze latitude (i.e., whether participants looked above or below the horizon in relation to them and independent of scene orientation; Figure 7 a), there was a strong difference, showing that participants spent more time looking above the horizon when scenes were inverted ($b = 0.16$, $SE = 0.01$, $t = 22.83$, $p < .001$). Gaze durations on the target did not play a role ($b = -0.00$, $SE = 0.00$, $t = -0.63$, $p = .530$).

Lastly, we investigated amplitudes for gaze and head movements (Figure 7 b and c), and found a small effect of scene inversion on gaze but not head movement amplitudes ($b = -0.07$, $SE = 0.03$, $t = -1.97$, $p = .049$; $b = 0.03$, $SE = 0.04$, $t = 0.78$, $p = .439$, respectively). There was no effect of gaze duration on the target for either amplitudes of gaze or head movements ($b = -0.01$, $SE = 0.01$, $t = -0.46$, $p = .648$; $b = -0.01$, $SE = 0.01$, $t = -0.97$, $p = .331$, respectively).

Discussion

With this study, we showed that it is feasible to use scene inversion as a manipulation in VR to investigate cognitive functions such as visual search. In our case, we were interested in whether scene inversion in VR exerts the same disruptive effects on semantic guidance as has been observed in studies using scene inversion in computer-based 2D setups (Brockmole & Henderson, 2006; Hayes & Henderson, 2022; Koehler & Eckstein, 2015). Due to the disruption, we expected to see more memory usage through inverted scenes compared to the usual already very efficient search through upright scenes guided by scene grammar knowledge (Boettcher et al., 2018; David et al., 2021; Draschkow & Võ, 2017; Helbing et al., 2022; Võ, 2021; Võ et al., 2019).

In line with the findings by Koehler and Eckstein (2015), where search performance dropped in inverted scenes, we similarly found that participants needed more time to find target objects in inverted scenes compared to upright scenes, indicating that scene inversion did impede search efficiency. Splitting up response times into more fine-grained subcomponents of the search process, that is, initiation time, scanning time, and verification time (Hollingworth & Bahle, 2019; Malcolm & Henderson, 2009), revealed that all phases of the search process except initiation time were negatively affected by inversion.

Surprisingly, we did not see increased memory usage during search through inverted scenes, even though search through inverted scenes was more difficult as observed by the overall longer search times. Higher task difficulty is also known to increase memory usage (Draschkow et al., 2021). This indicates that participants did not compensate the increased difficulty by using more memory. Instead, semantic guidance—though inverted—might still be accessible to a degree that it can be used to guide search efficiently enough.

In natural behavior there is often little to no need to keep information in visual working memory when the information is still available and only needs to be looked at again (Draschkow et al., 2021; Melnik et al., 2018; Van der Stigchel, 2020; Zhang & Pan, 2022). Since the information (i.e., location of the target) in our task was still readily available in inverted scenes, participants preferred to simply search again (Horowitz & Wolfe, 1998) instead of building up more effortful memory using the world as an "outside memory" (Chemero, 2009; O'Regan, 1992). In both upright and inverted scenes participants could further use their own body movements within the scene to infer spatial relations of object locations (Droll & Hayhoe, 2007; Pan et al., 2013); information which is lacking in traditional 2D computer screen setups and might lead to different memory encoding strategies (Johnsdorf et al., 2023; Zhang & Pan, 2022).

Last but not least, another reason why we did not observe an increase in memory usage might lie in the nature of our VR stimuli. Every scene consisted of eight anchor objects and 20 local objects. However, our VR scenes greatly underestimate the number of objects typically found in real scenes which tend to be much more cluttered (Neider & Zelinsky, 2008; Wolfe et al., 2011). Thus, our scenes may have been too sparse to elicit stronger memory effects. This, however, is unlikely to be the determining factor in causing the absence of a memory effect: Other studies using the same (Helbing et al., 2022) or less complex scenes (Draschkow et al., 2017) observed strong search slope differences between conditions when other aspects of the scenes were manipulated.

Further, we looked at several gaze measures in an exploratory fashion. In line with Hayes and Henderson (2022), we also did not observe an effect on fixation durations between upright and inverted scenes. However, participants made more fixations in total and fixated more objects in inverted scenes, which is likely due to the fact that they also spent more time in total in inverted scenes. In addition, we also observed more target refixations in inverted scenes, which indicates that scene inversion did impede object recognition (Kelley et al., 2003; Lauer et al., 2020). In line with Foulsham and colleagues (2008) and Anderson and colleagues (2020), we found that most gaze saccades occurred along the horizontal axis regardless of scene orientation. We further found a bias for rightward movements for both gaze and head. In addition, head movements were mainly directed forwards, while gaze also exhibited return saccades (David, Beitner, et al., 2020; David et al., 2021, 2022). When investigating saccade movements, we found that scene inversion affected the pattern of absolute directions of both head and gaze movements, that is, the rightward bias slightly decreased and downward movements increased in inverted scenes.

To summarize, we found that scene inversion in VR did impede search efficiency but did not lead to an uptake in memory usage. Eye tracking allowed us to further investigate gaze and head movement behavior during unrestricted natural behavior and to examine how it was affected by the inversion manipulation. We showed that it is feasible and advances research to use more complex manipulations such as scene inversion in VR in combination with eye tracking to investigate cognitive mechanisms and eye movement behavior in ecologically valid settings. Results can then be compared to studies performed in traditional laboratory setups using 2D computer screens. In our case, we replicated effects of scene inversion, that is, detrimental effects on search performance as well as a change in gaze and head movement directions. However, we could not find the expected increase in memory usage brought about by the disruptions caused by scene inversion. Future studies using within-participant designs might be able to shed more light on how scene inversion affects eye movement behavior and memory usage in 3D environments compared to 2D setups. We hope that our study inspires other vision researchers to implement new paradigms in VR in combination with eye tracking to allow for rigorous and highly controlled investigations of daily human behavior in real-world scenarios.

Ethics and Conflict of Interest

The authors declare that the contents of the article are in agreement with the ethics described in http://biblio.unibe.ch/portale/elibrary/BOP/jemr/ethics.html and that there is no conflict of interest regarding the publication of this paper.

Acknowledgements

This work was funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation)—project number 222641018—SFB/TRR 135, sub-project C7 as well as the Hessisches Ministerium für Wissenschaft und Kunst (HMWK), project "The Adaptive Mind" to M.L-H.V., and by the Main-Campus-doctus scholarship of the Stiftung Polytechnische Gesellschaft Frankfurt a. M. to J.B.

The Wellcome Centre for Integrative Neuroimaging is supported by core funding from the Wellcome Trust (203139/Z/16/Z). D.D. is supported by the NIHR Oxford Health Biomedical Research Centre. The funders had no role in the decision to publish or in the preparation of the manuscript.

J.H. is supported by the Deutschlandstipendium scholarship from the Bundesministerium für Bildung und Forschung (BMBF; Federal Ministry of Education and Research) and Goethe University Frankfurt.

We thank Martin Schultze for helpful feedback on the statistical analyses. We would further like to thank Zoë Bolz, Danislava Chuhovska, Aylin Kallmayer, Lea Karoza, and Levi Kumle for help with testing participants.

Author Contributions

Conceptualization: J.B. & M.L.-H.V.; Methodology: J.B., D.D., M.L.-H.V.; Software: J.B., J.H., E.J.D.; Validation: J.B.; Formal Analysis: J.B.; E.J.D.; Investigation: J.B.; Resources: J.B., J.H., E.J.D.; Data Curation: J.B.; Writing – Original: J.B.; Writing – Review & Editing: J.B., J.H., D.D., E.J.D., M.L.-H.V.; Supervision: D.D., M.L.-H.V., Project Administration: J.B., M.L.-H.V.; Funding Acquisition: M.L.-H.V.

References

Anderson, B. A., & Lee, D. S. (2023). Visual search as effortful work. *Journal of Experimental Psychology: General*. Advance online publication. https://doi.org/10.1037/xge0001334

Anderson, N. C., Bischof, W. F., Foulsham, T., & Kingstone, A. (2020). Turning the (virtual) world around: Patterns in saccade direction vary with picture orientation and shape in virtual reality. *Journal of Vision*, *20*(8), 21. https://doi.org/10.1167/jov.20.8.21

Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, *59*(4), 390–412. https://doi.org/10.1016/j.jml.2007.12.005

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, *68*(3), 255–278. https://doi.org/10.1016/j.jml.2012.11.001

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1). https://doi.org/10.18637/jss.v067.i01

Beighley, S., & Intraub, H. (2016). Does inversion affect boundary extension for briefly-presented views? *Visual Cognition*, *24*(3), 252–259. https://doi.org/10.1080/13506285.2016.1229369

Beitner, J., Helbing, J., Draschkow, D., & Võ, M. L.-H. (2021). Get your guidance going: Investigating the activation of spatial priors for efficient search in virtual reality. *Brain Sciences*, *11*(1), 44. https://doi.org/10.3390/brainsci11010044

Boettcher, S. E. P., Draschkow, D., Dienhart, E., & Võ, M. L.-H. (2018). Anchoring visual search in scenes: Assessing the role of anchor objects on eye movements during visual search. *Journal of Vision*, *18*(13), 11. https://doi.org/10.1167/18.13.11

Brockmole, J. R., & Henderson, J. M. (2006). Using real-world scenes as contextual cues for search. *Visual Cognition*, *13*(1), 99–108. https://doi.org/10.1080/13506280500165188

Chemero, A. (2009). *Radical embodied cognitive science*. MIT Press.

David, E. J., Beitner, J., & Võ, M. L.-H. (2020). Effects of transient loss of vision on head and eye movements during visual search in a virtual environment. *Brain Sciences*, *10*(11), 841. https://doi.org/10.3390/brainsci10110841

David, E. J., Beitner, J., & Võ, M. L.-H. (2021). The importance of peripheral vision when searching 3D real-world scenes: A gaze-contingent study in virtual reality. *Journal of Vision*, *21*(7), 3. https://doi.org/10.1167/jov.21.7.3

David, E. J., Lebranchu, P., Perreira Da Silva, M., & Le Callet, P. (2020). A new toolbox to process gaze and head motion in virtual reality. *In Preparation*.

David, E. J., Lebranchu, P., Perreira Da Silva, M., & Le Callet, P. (2022). What are the visuo-motor tendencies of omnidirectional scene free-viewing in virtual reality? *Journal of Vision*, *22*(4), 12. https://doi.org/10.1167/jov.22.4.12

Draschkow, D. (2022). Remote virtual reality as a tool for increasing external validity. *Nature Reviews Psychology*, 1(8), 433–434. https://doi.org/10.1038/s44159-022-00082-8

Draschkow, D., Kallmayer, M., & Nobre, A. C. (2021). When natural behavior engages working memory. *Current Biology*, *31*(4), 869–874. https://doi.org/10.1016/j.cub.2020.11.013

Draschkow, D., Nobre, A. C., & van Ede, F. (2022). Multiple spatial frames for immersive working memory. *Nature Human Behaviour*, *6*(4), 536–544. https://doi.org/10.1038/s41562-021-01245-y

Draschkow, D., & Võ, M. L.-H. (2017). Scene grammar shapes the way we interact with objects, strengthens memories, and speeds search. *Scientific Reports*, *7*, 16471. https://doi.org/10.1038/s41598-017-16739-x

Droll, J. A., & Hayhoe, M. M. (2007). Trade-offs between gaze and working memory use. *Journal of Experimental Psychology: Human Perception and Performance*, *33*(6), 1352–1365. https://doi.org/10.1037/0096-1523.33.6.1352

Einhäuser, W., Schumann, F., Bardins, S., Bartl, K., Böning, G., Schneider, E., & König, P. (2007). Human eye-head co-ordination in natural exploration. *Network: Computation in Neural Systems*, *18*(3), 267–297. https://doi.org/10.1080/09548980701671094

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. https://doi.org/10.1016/j.visres.2008.05.018

Freedman, E. G. (2008). Coordination of the eyes and head during visual orienting. *Experimental Brain Research*, *190*(4), 369–387. https://doi.org/10.1007/s00221-008-1504-8

Gutiérrez, J., David, E. J., Coutrot, A., Da Silva, M. P., & Le Callet, P. (2018). Introducing UN Salient360! Benchmark: A platform for evaluating visual attention models for 360° contents. *2018 Tenth International Conference on Quality of Multimedia Experience (QoMEX)*, 1–3. https://doi.org/10.1109/QoMEX.2018.8463369

Gutiérrez, J., David, E. J., Rai, Y., & Le Callet, P. (2018). Toolbox and dataset for the development of saliency and scanpath models for omnidirectional/360° still images. *Signal Processing: Image Communication*, *69*, 35–42. https://doi.org/10.1016/j.image.2018.05.003

Hayes, T. R., & Henderson, J. M. (2022). Scene inversion reveals distinct patterns of attention to semantically interpreted and uninterpreted features. *Cognition*, *229*, 105231. https://doi.org/10.1016/j.cognition.2022.105231

Helbing, J., Draschkow, D., & Võ, M. L.-H. (2020). Search superiority: Goal-directed attentional allocation creates more reliable incidental identity and location memory than explicit encoding in naturalistic virtual environments. *Cognition*, *196*, 104147. https://doi.org/10.1016/j.cognition.2019.104147

Helbing, J., Draschkow, D., & Võ, M. L.-H. (2022). Auxiliary scene-context information provided by anchor objects guides attention and locomotion in natural search behavior. *Psychological Science*, *33*(9), 1463–1476. https://doi.org/10.1177/09567976221091838

Hoffman, L., & Rovine, M. J. (2007). Multilevel models for the experimental psychologist: Foundations and illustrative examples. *Behavior Research Methods*, *39*(1), 101–117. https://doi.org/10.3758/BF03192848

Hollingworth, A., & Bahle, B. (2019). Eye tracking in visual search experiments. In S. Pollmann (Ed.), *Spatial learning and attention guidance* (Vol. 151, pp. 23–35). Humana Press. https://doi.org/10.1007/7657_2019_30

Hope, R. M. (2013). *Rmisc: Ryan Miscellaneous* (1.5) [R package]. https://cran.r-project.org/package=Rmisc

Horowitz, T. S., & Wolfe, J. M. (1998). Visual search has no memory. *Nature*, *394*(6693), 575–577. https://doi.org/10.1038/29068

Johnsdorf, M., Kisker, J., Gruber, T., & Schöne, B. (2023). Comparing encoding mechanisms in realistic virtual reality and conventional 2D laboratory settings: Event-related potentials in a repetition suppression paradigm. *Frontiers in Psychology*, *14*, 1051938. https://doi.org/10.3389/fpsyg.2023.1051938

Kass, R. E., & Raftery, A. E. (1995). Bayes factors. *Journal of the American Statistical Association*, *90*(430), 773–795. https://doi.org/10.2307/2291091

Kelley, T. A., Chun, M. M., & Chua, K.-P. (2003). Effects of scene inversion on change detection of targets matched for visual salience. *Journal of Vision*, *3*(1), 1. https://doi.org/10.1167/3.1.1

Kisker, J., Gruber, T., & Schöne, B. (2021). Experiences in virtual reality entail different processes of retrieval as opposed to conventional laboratory settings: A study on human memory. *Current Psychology*, *40*(7), 3190–3197. https://doi.org/10.1007/s12144-019-00257-2

Kliegl, R., Wei, P., Dambacher, M., Yan, M., & Zhou, X. (2011). Experimental effects and individual differences in linear mixed models: Estimating the relationship between spatial, object, and attraction effects in visual attention. *Frontiers in Psychology*, *1*, 238. https://doi.org/10.3389/fpsyg.2010.00238

Koehler, K., & Eckstein, M. P. (2015). Scene inversion slows the rejection of false positives through saccade exploration during search. *Proceedings of the 37th Annual Meeting of the Cognitive Science Society*, 1141–1146. https://cogsci.mindmodeling.org/2015/papers/0202/paper0202.pdf

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest package: Tests in linear mixed effects models. *Journal of Statistical Software*, *82*(13). https://doi.org/10.18637/jss.v082.i13

Lauer, T., Willenbockel, V., Maffongelli, L., & Võ, M. L.-H. (2020). The influence of scene and object orientation on the scene consistency effect. *Behavioural Brain Research*, *394*, 112812. https://doi.org/10.1016/j.bbr.2020.112812

Li, C.-L., Aivar, M. P., Kit, D. M., Tong, M. H., & Hayhoe, M. M. (2016). Memory and visual search in naturalistic 2D and 3D environments. *Journal of Vision*, *16*(8), 9. https://doi.org/10.1167/16.8.9

Malcolm, G. L., & Henderson, J. M. (2009). The effects of target template specificity on visual search in real-world scenes: Evidence from eye movements. *Journal of Vision*, *9*(11), 8. https://doi.org/10.1167/9.11.8

Melnik, A., Schüler, F., Rothkopf, C. A., & König, P. (2018). The world as an external memory: The price of saccades in a sensorimotor task. *Frontiers in Behavioral Neuroscience*, *12*, 253. https://doi.org/10.3389/fnbeh.2018.00253

Meng, M., & Potter, M. C. (2008). Detecting and remembering pictures with and without visual noise. *Journal of Vision*, *8*(9), 7. https://doi.org/10.1167/8.9.7

Morey, R. D., Rouder, J. N., Jamil, T., Urbanek, S., Forner, K., & Ly, A. (2022). *Package "BayesFactor"* (0.9.12-4.2). https://CRAN.R-project.org/package=BayesFactor

Nakagawa, S., & Schielzeth, H. (2013). A general and simple method for obtaining R^2 from generalized linear mixed-effects models. *Methods in Ecology and Evolution*, *4*(2), 133–142. https://doi.org/10.1111/j.2041-210x.2012.00261.x

Neider, M. B., & Zelinsky, G. J. (2008). Exploring set size effects in scenes: Identifying the objects of search. *Visual Cognition*, *16*(1), 1–10. https://doi.org/10.1080/13506280701381691

Nyström, M., & Holmqvist, K. (2010). An adaptive algorithm for fixation, saccade, and glissade detection in eyetracking data. *Behavior Research Methods*, *42*(1), 188–204. https://doi.org/10.3758/BRM.42.1.188

O'Regan, J. K. (1992). Solving the "real" mysteries of visual perception: The world as an outside memory. *Canadian Journal of Psychology / Revue Canadienne de Psychologie*, *46*(3), 461–488. https://doi.org/10.1037/h0084327

Pan, J. S., Bingham, N., & Bingham, G. P. (2013). Embodied memory: Effective and stable perception by combining optic flow and image structure. *Journal of Experimental Psychology: Human*

Perception and Performance, *39*(6), 1638–1651. https://doi.org/10.1037/a0032070

Parsons, T. D. (2015). Virtual reality for enhanced ecological validity and experimental control in the clinical, affective and social neurosciences. *Frontiers in Human Neuroscience*, *9*, 660. https://doi.org/10.3389/fnhum.2015.00660

R Core Team. (2022). *R: A language and environment for statistical computing* (4.2.2). [Programming language R]. R Foundation for Statistical Computing. https://www.R-project.org/

Rock, I. (1974). The perception of disoriented figures. *Scientific American*, *230*(1), 78–85. https://doi.org/10.1038/scientificamerican0174-78

RStudio Team. (2022). *RStudio: Integrated development environment for R* (2022.7.1.554). [Computer software]. Posit Software. https://posit.co/

Rubo, M., Messerli, N., & Munsch, S. (2021). The human source memory system struggles to distinguish virtual reality and reality. *Computers in Human Behavior Reports*, *4*, 100111. https://doi.org/10.1016/j.chbr.2021.100111

Schöne, B., Sylvester, R. S., Radtke, E. L., & Gruber, T. (2021). Sustained inattentional blindness in virtual reality and under conventional laboratory conditions. *Virtual Reality*, *25*(1), 209–216. https://doi.org/10.1007/s10055-020-00450-w

Schöne, B., Wessels, M., & Gruber, T. (2017). Experiences in virtual reality: A window to autobiographical memory. *Current Psychology*, *38*, 715–719. https://doi.org/10.1007/s12144-017-9648-y

Shore, D. I., & Klein, R. M. (2000). The effects of scene inversion on change blindness. *The Journal of General Psychology*, *127*(1), 27–43. https://doi.org/10.1080/00221300009598569

Stahl, J. S. (1999). Amplitude of human head movements associated with horizontal saccades. *Experimental Brain Research*, *126*(1), 41–54. https://doi.org/10.1007/s002210050715

Tarr, M. J., & Warren, W. H. (2002). Virtual reality in behavioral neuroscience and beyond. *Nature Neuroscience*, *5*(11, Suppl.), 1089–1092. https://doi.org/10.1038/nn948

Taubert, J., Apthorp, D., Aagten-Murphy, D., & Alais, D. (2011). The role of holistic processing in face perception: Evidence from the face inversion effect. *Vision Research*, *51*(11), 1273–1278. https://doi.org/10.1016/j.visres.2011.04.002

Valentine, T. (1988). Upside-down faces: A review of the effect of inversion upon face recognition. *British Journal of Psychology*, *79*(4), 471–491. https://doi.org/10.1111/j.2044-8295.1988.tb02747.x

Van der Stigchel, S. (2020). An embodied account of visual working memory. *Visual Cognition*, *28*(5–8), 414–419. https://doi.org/10.1080/13506285.2020.1742827

Võ, M. L.-H. (2021). The meaning and structure of scenes. *Vision Research*, *181*, 10–20. https://doi.org/10.1016/j.visres.2020.11.003

Võ, M. L.-H., Boettcher, S. E. P., & Draschkow, D. (2019). Reading scenes: How scene grammar guides attention and aids perception in real-world environments. *Current Opinion in Psychology*, *29*, 205–210. https://doi.org/10.1016/j.copsyc.2019.03.009

Võ, M. L.-H., & Wolfe, J. M. (2012). When does repeated search in scenes involve memory? Looking at versus looking for objects in scenes. *Journal of Experimental Psychology: Human Perception and Performance*, *38*(1), 23–41. https://doi.org/10.1037/a0024147

Võ, M. L.-H., & Wolfe, J. M. (2013). The interplay of episodic and semantic memory in guiding repeated search in scenes. *Cognition*, *126*(2), 198–212. https://doi.org/10.1016/j.cognition.2012.09.017

von Noorden, G. K., & Campos, E. C. (2002). *Binocular vision and ocular motility: Theory and management of strabismus* (6th ed.). Mosby.

Walther, D. B., Caddigan, E., Fei-Fei, L., & Beck, D. M. (2009). Natural scene categories revealed in distributed patterns of activity in the human brain. *Journal of Neuroscience*, *29*(34), 10573–10581. https://doi.org/10.1523/JNEUROSCI.0559-09.2009

Wickham, H. (2016). *ggplot2: Elegant graphics for data analysis*. (2nd ed.). [R Package]. Springer. https://doi.org/10.1007/978-3-319-24277-4

Wilkinson, G. N., & Rogers, C. E. (1973). Symbolic description of factorial models for analysis of variance. *Applied Statistics*, *22*(3), 392–399. https://doi.org/10.2307/2346786

Wolfe, J. M., Alvarez, G. A., Rosenholtz, R., Kuzmova, Y. I., & Sherman, A. M. (2011). Visual search for arbitrary objects in real scenes. *Attention, Perception, & Psychophysics*, *73*(6), 1650–1671. https://doi.org/10.3758/s13414-011-0153-3

Yin, R. K. (1969). Looking at upside-down faces. *Journal of Experimental Psychology*, *81*(1), 141–145. https://doi.org/10.1037/h0027474

Zhang, H., & Pan, J. S. (2022). Visual search as an embodied process: The effects of perspective schange and external reference on search performance. *Journal of Vision*, *22*(10), 13. https://doi.org/10.1167/jov.22.10.13