



Eye movements to absent objects during mental imagery and visual memory in immersive virtual reality

Sandra Chiquet¹ · Corinna S. Martarelli² · Fred W. Mast¹

Received: 21 January 2020 / Accepted: 1 October 2020 / Published online: 23 October 2020
© The Author(s) 2020

Abstract

The role of eye movements in mental imagery and visual memory is typically investigated by presenting stimuli or scenes on a two-dimensional (2D) computer screen. When questioned about objects that had previously been presented on-screen, people gaze back to the location of the stimuli, even though those regions are blank during retrieval. It remains unclear whether this behavior is limited to a highly controlled experimental setting using 2D screens or whether it also occurs in a more naturalistic setting. The present study aims to overcome this shortcoming. Three-dimensional (3D) objects were presented along a circular path in an immersive virtual room. During retrieval, participants were given two tasks: to visualize the objects, which they had encoded before, and to evaluate a statement about visual details of the object. We observed longer fixation duration in the area, on which the object was previously displayed, when compared to other possible target locations. However, in 89% of the time, participants fixated none of the predefined areas. On the one hand, this shows that looking at nothing may be overestimated in 2D screen-based paradigm, on the other hand, the looking at nothing effect was still present in the 3D immersive virtual reality setting, and thus it extends external validity of previous findings. Eye movements during retrieval reinstate spatial information of previously inspected stimuli.

Keywords Virtual reality · Mental imagery · Visual memory · Eye movements · Eye tracking

1 Introduction

During retrieval people fixate on empty locations that have been associated with task-relevant stimuli during encoding (Altmann 2004; Bone et al. 2019; Brandt and Stark 1997; Johansson et al. 2012; Johansson et al. 2006; Kumcu and Thompson 2018; Laeng et al. 2014; Laeng and Teodorescu 2002; Richardson and Spivey 2000; Scholz et al. 2018; Scholz et al. 2016; Spivey and Geng 2001). For instance, Spivey and Geng (2001) found that when participants were questioned about an object, they gazed back to empty

locations on the screen corresponding to those, where visual information was presented during encoding. To date, many studies have replicated such “looking at nothing” on a two-dimensional (2D) computer screen (Johansson and Johansson 2014; Martarelli et al. 2017; Martarelli and Mast 2011, 2013).

The typical screen-based variant of looking at nothing leads to increased experimental control but is not necessarily a suitable model for mental imagery and visual memory processes in real life. First, participants were exposed to isolated stimuli on a 2D computer screen, which makes them take on the role of external observers of stimuli on a computer screen. This is a key difference between the experimental situation in the laboratory and real life, where people are able to interactively control their view based on head motion and thus have the feeling of being inside the environment (Gorini et al. 2011; Makransky et al. 2019; Moreno and Mayer 2002). Second, eye movements during retrieval are measured while participants are exposed to a blank screen. Only rarely do we view blank screens in real life. Mental imagery and memory processes take place in a

✉ Sandra Chiquet
sandra.chiquet@psy.unibe.ch

Corinna S. Martarelli
corinna.martarelli@fernuni.ch

Fred W. Mast
fred.mast@psy.unibe.ch

¹ Department of Psychology, University of Bern, 3012 Bern, Switzerland

² Swiss Distance University Institute, 3900 Brig, Switzerland

richly structured environment, including depth cues (e.g., texture and perspective) and visual input (e.g., edges, walls, shelves).

Compared to screen-based eye tracking, immersive virtual reality (IVR) has the potential to investigate human visual behavior in a relatively natural context, while still maintaining high experimental control. Thus, virtual reality technology might help to bridge the gap between highly controlled 2D screen-based experimental settings and real life. By means of a head-mounted display (HMD), the visual information presented to the participants, matches some crucial cues of the real world. Indeed, the visual field is updated when the head moves, enabling participants to interact with the environment. As a consequence, eye movements are not constrained to the frame of the computer screen, thus increasing the space of possible locations to look at during retrieval. Moreover, the interactivity contributes to spatial presence which refers to the sensation of being physically present in the environment (Gorini et al. 2011; Makransky et al. 2019; Moreno and Mayer 2002). Rather than being passive observers, in IVR experiments participants become part of the environment. Spatial presence is considered to contribute to the emotional state during the virtual reality experience (Diemer et al. 2015; Riva et al. 2007; Slater 2009) and to affect memory performance (Makowski et al. 2017). It is thus possible, that the feeling of being there in the environment, has an impact on memory encoding and retrieval when compared to being exposed to stimuli on an external computer screen (Kisker et al. 2019; Ventura et al. 2019).

Crucially, IVR offers the possibility to assess eye movements in a structured environment, including the analysis of perceptual information of the environment. It has been shown that additional visual input has consequences on both, memory performance (Scholz et al. 2018) and eye movements in the looking at nothing paradigm (Kumcu et al. 2016, 2018). Thus, providing additional visual information during retrieval (i.e., pedestals) might disrupt looking at nothing behavior in IVR (and in real world).

Despite the advantages of IVR (high control and high ecological validity), there are still relatively few studies using IVR to investigate human visual behavior. This is partly due to challenging or lacking technical equipment. However, recent studies combining eye-tracking and IVR are promising (Clay et al. 2019; Grogorick et al. 2017). For example, Eichert et al. (2018) illustrated that eye movements can be used to investigate and validate traditional experimental findings in immersive virtual environments.

In the current study, we aim to extend external validity of looking at nothing by investigating the phenomenon in IVR. We propose that IVR provides a relatively close approximation of human visual behavior in mental imagery and visual memory, since IVR matches some crucial cues of the

real world. Under a free gaze condition (i.e., participants are allowed to gaze freely) we compared gaze duration in the areas associated with the locations that were inspected during perception (corresponding) with other areas (non-corresponding). We expect longer fixation duration in the corresponding location compared to other locations, because gaze behavior during pictorial recall is expected to reflect the locations where the objects were encoded before.

2 Method

2.1 Participants

The experiment was carried out at the University of Bern. The results are based on 19 native German-speakers (16 males, $M_{\text{age}} = 21.58$, $SD = 3.15$, range = 16–32). One participant had to be excluded because of recording problems during retrieval. 90% of participants reported having no (60%) or only limited (30%) experience with IVR. 10% mentioned having used head-mounted displays several times. All participants reported normal or corrected-to-normal vision. The sample size was selected a priori and was comparable with sample sizes used in previous studies (e.g., $N = 24$ in Johansson and Johansson 2014; $N = 22$ in Spivey and Geng 2001). Participants were fully informed about the experimental procedure and were given standardised oral instruction. Informed consent was collected from all participants prior to the experiment. There was no financial compensation for study participation. The experiment was approved by the ethics committee of the University of Bern.

2.2 Virtual environment and 3D virtual objects

Using the 3D application software Vizard 5.4 (WorldViz LLC, Santa Barbara, CA) we created a virtual room containing six pedestals distributed along a circular segment of 180 degrees. The pedestals were in the visual field of participants, who were positioned in the center of the virtual environment. Lightning and contrast between background and pedestals was kept constant across participants. A set of 38 computer-generated virtual objects including textures (in 37 cases) were gathered from websites offering free 3D models for VR and AR. The objects are provided in “Appendix A”. In one case the texture was costume-made using the 3D computer graphics software Blender 2.75a (<https://www.blender.org>). Each object belonged to one of five categories (animals, sports equipment, vehicles, technical devices and characters). We quantified object size by designing a virtual bounding box, a regular cuboid which defined the maximum volume of the objects. During the experiment the virtual objects were displayed on top of one of the six pedestals. Example stimuli in the virtual room are shown in

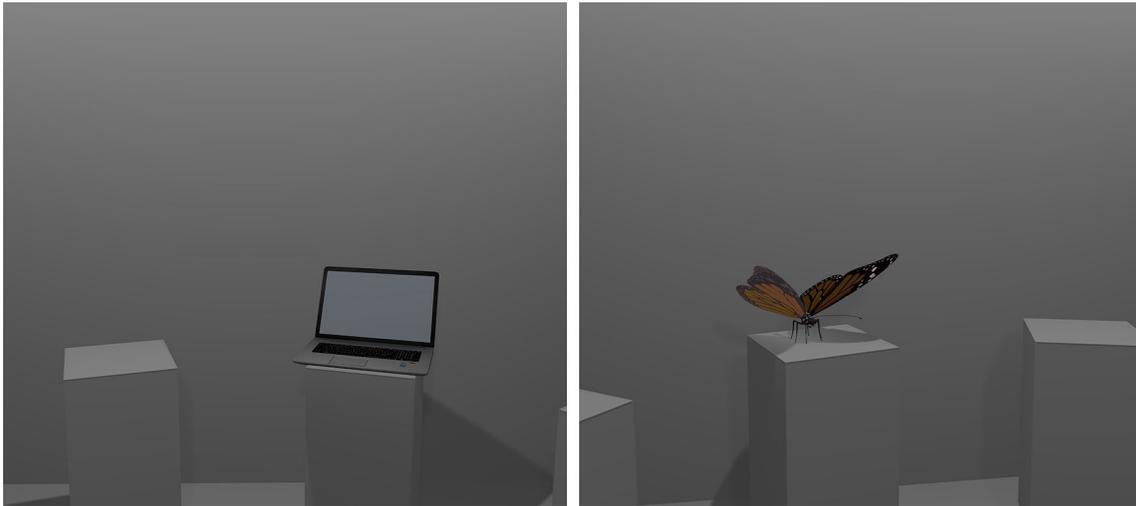


Fig. 1 Participants' view of the virtual room by means of the head-mounted-display (HMD) with two examples of stimuli presented during the perceptual encoding phase

Fig. 1. Each object was paired with a spoken title (i.e., the name of the object) and a spoken statement about the object (e.g., “the butterfly is patterned orange–black”). All titles and statements were spoken by a female native speaker and recorded using Audacity (<https://audacity.sourceforge.net>).

2.3 Equipment

The virtual room and the objects were presented by means of the SMI eye tracking HMD based on the *Oculus Rift Development Kit 2* (DK2) (<https://www.oculus.com/en-us/dk2/>). The DK2 is a head-mounted-display (HMD) with positional camera tracking system with six-degree-of-freedom head tracking (rotation and translation) and a maximum frame rate of 75 Hz at a display resolution of 960*1080 pixel per eye. We used *oculus runtime SDK* for basic HMD functionality, and we used *iViewHMD* for 250 Hz binocular eye tracking. The experiment was programmed and performed using the 3D application software *Vizard 5.4* with a 64bit Windows 10 Pro operating system on a computer with a graphics card (*NVIDIA® GeForce GTX 750 Ti*) sufficient for the DK2 maximal refresh rate. The SMI eye tracking HMD provides a five-point calibration and validation procedure, which each participant performed prior to the experiment proper ($M_{\text{error value}} = 1.16$, $SD = 0.46$, $\text{range} = 0.71\text{--}2.44$). All audio files were presented via loudspeakers.

2.4 Areas of interest and eye-tracking in IVR

To determine fixations, we defined six equally sized 3D areas of interest (AOI) on top of the pedestals. The x (width), y (height) and z (depth) dimensions of the AOI was defined by the object's maximal bounding box. The area where the

object was presented was defined as corresponding AOI, the other areas as non-corresponding AOIs. The space around the AOI was defined as empty space. A schematic representation of the AOIs position and size in relation to the virtual room is provided in Fig. 2.

2.5 Procedure

The experiment was divided into two phases: the perceptual encoding phase and the retrieval phase. Participants were seated on a comfortable chair and were informed about the upcoming procedure. They put on the HMD, and the initial position of the HMD was recalibrated to ensure that the participant was perfectly centered in the virtual room. Prior to the start of the perceptual encoding phase, we used a five-point calibration and validation procedure, provided by the *iView* HMD. During the perceptual encoding phase, 38 virtual objects were displayed one by one on top of one of the six pedestals distributed along a circular segment of 180 degrees in the virtual room. The order and the spatial position of the object (i.e., on which pedestal the object was presented) was fully randomized. Simultaneously, participants heard the name of the presented object and were instructed to learn the objects in detail. The stimulus title was used in the following retrieval phase for trial induction. Each stimulus appeared for 6 s, preceded by a 0.5 s fixation square straight ahead on the wall in order to control for the initial gaze direction across participants and trials. The starting gaze position was thus always the same. In the retrieval phase, participants were given two tasks: to visualize the object, which they had encoded before (image generation task), and to evaluate (true/false) a statement about visual details of the object (image inspection task). The order was

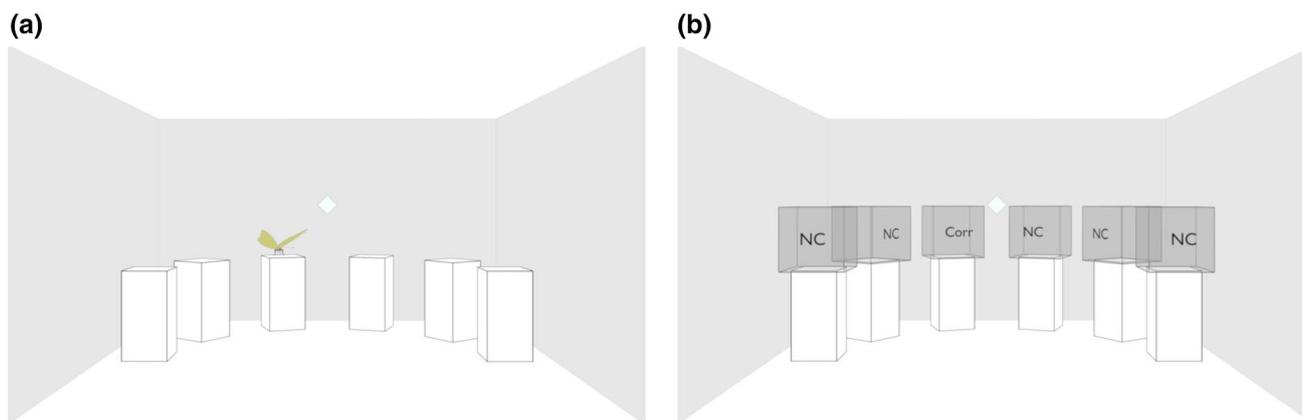


Fig. 2 Schematic representation of the virtual room during encoding (a) and during retrieval (b) with the AOIs (in dark grey; not visible for the participants) on top of the pedestals. The area where the object

was presented during encoding was defined as the corresponding AOI (Corr), the other areas as non-corresponding AOIs (NC) during retrieval

again fully randomized. As indicated by a cue (participants heard one of the 38 object titles, e.g., “butterfly”), participants first visualized the object as vividly as possible and pressed the answer buttons (left and right at the same time), once they had generated the mental image. Then they heard a specific statement about the object (e.g., “the butterfly is patterned orange-black”) and made their response as true or false (self-paced) by pressing a response-button (right click for “true”, left click for “false”). As soon as participants made the true/false judgment, the trial ended, and a new object name was auditorily presented, preceded by a 0.5 s fixation square straight ahead on the wall until all 38 objects were tested.

2.6 Statistical analysis

Data were recorded at a sampling rate of 60 Hz. Every sample was associated with three direction vectors ($V = v_x, v_y, v_z$) one for each eye, and one binocular signal [read Lohr et al. (2019) for more information about the binocular Signal] and the information whether the binocular direction vector overlapped with one of the AOIs or not. We recorded $M = 74,763$ ($SD = 14,122$) data points per participant. We were interested in the total fixation duration in the AOIs and in the empty space (fixation duration in space excluding the AOIs). A fixation was defined as the maintenance of visual gaze on the same AOI for at least 100 ms. That is, collisions of the gaze vector with a pre-defined AOI for less than 100 ms were excluded. This correction led to an exclusion of 4.69% samples of the perceptual encoding dataset and 2.40% samples of the retrieval dataset. The fixation proportion was calculated for each AOI and for empty space per trial and per participant, by phase (perceptual encoding phase, retrieval phase) and by task (image generation and image inspection; only in the retrieval phase). In addition, we calculated the

mean fixation proportion of all non-corresponding areas, in order to account for the weighted probability fixating one of the five non-corresponding AOIs. Fixation proportion in the AOIs are continuous, nonnegative numbers, bounded at zero and one, which can be equal to zero, if no fixation was observed in a certain AOI during a trial or equal to one, if a certain AOI was fixated during the entire trial duration. Many studies have analysed fixation proportion using frameworks assuming normal distribution and unbound and homogenous variances of the data. However, preliminary analysis (check for kurtosis and skewedness, kernel density plot, Q–Q plots) indicate that the continuous component of the response variable is appropriately described by the beta distribution. Because of excessive zero and one outcomes, data structure should follow a mixed distribution, in which a beta variable is mixed with a probability mass accounting for zero and one. Therefore, we used a zero–one-inflated beta model framework which considers a beta distribution for the continuous proportion outcome in the closed $(0, 1)$ interval and a Bernoulli distribution for the binary $\{0, 1\}$ outcome (Ospina and Ferrari 2012).

Data analysis was performed in R (3.5.1; R Core Team). We followed a Bayesian approach using the *brms* package for Bayesian (non-)linear mixed models (Bürkner 2017). We chose Bayesian inference since it allows for estimating the relative credibility of parameters given the data, which is not the case in a frequentist data analysis approach (Kruschke 2015; Lee et al. 2005; Wagenmakers 2007). Furthermore, the *brms* package has the capability to easily handle mixed distributed data. We used Markov Chain Monte Carlo (MCMC) method with 4 chains of 2000 iterations to calculate posterior parameter estimates. Based on the mixed data structure, we analyzed fixation proportion by means of generalized linear mixed models (GLMMs) using a

zero–one-inflated beta model. The model notation including prior distributions is available in "Appendix B".

3 Results

3.1 Fixation duration during perceptual encoding (manipulation check)

As manipulation check and validation check of measuring gaze behavior in IVR, we computed fixation time in the corresponding AOIs compared to the non-corresponding AOIs and to empty space during perceptual encoding. For each stimulus, participants spent on average 4.56 s (SE=0.05) in the area, where the objects were presented (corresponding AOI), whereas they fixated the non-corresponding AOIs 0.43 s (SE=0.02) and the empty space 0.73 s (SE=0.03). This confirms proper encoding of the stimuli, which appeared for 6 s each, and it validates the measurement of eye movements (i.e., fixation duration) in IVR by means of an HMD with an integrated eye-tracker.

3.2 Behavioral data

Participants were correct in evaluating the statements about visual details of the objects in 67% of all trials. After removing extreme values of response times (RTs $> M + 3 \times SD$ & RTs $< M - 3 \times SD$ for each participant and task; 1.66% of all trials), the mean response time (RT) per trial during image generation was $M = 3.27$ s (SE = 1.71) and during image inspection $M = 4.24$ s (SE = 1.95).

3.3 Fixation proportion during retrieval

During retrieval, participants spent most of the time in empty space (image generation task: $M = 2.92$ s, SE = 6.33, image inspection task: $M = 3.80$ s, SE = 6.95 per stimulus). That is, during the largest proportion of the time they fixated none of the AOIs. In order to follow up this observation, we run an exploratory visualization of fixations (i.e., overlap of binocular direction vectors with either AOIs, fixation square or walls in the virtual environment) using the shiny package for interactive web applications in R (Chang et al. 2019). The application is available on the platform shinyapp.io (https://chiquet.shinyapps.io/shiny_3D_fixdisp/). The visualizations indicate a tendency to fixate pedestals in the middle of the environment and locations centered around the fixation square on the wall. Since we were interested in the fixation duration in the corresponding AOI compared to the non-corresponding AOIs we removed fixations on the empty space from further analysis.

To compare the time spent in the corresponding area with the mean time spent in the five non-corresponding areas,

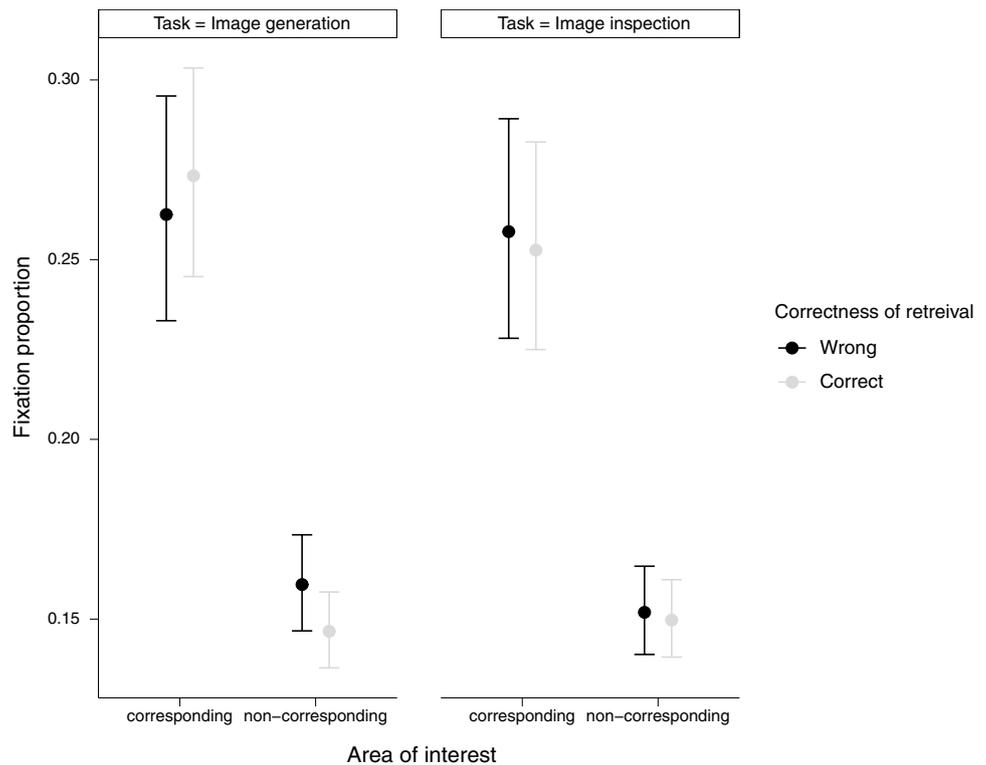
we estimated fixation proportion during retrieval, as a function of the AOI, the task and correctness using a Bayesian zero–one-inflated beta model. AOI as dummy-coded effect of the corresponding AOI compared to the mean of the non-corresponding AOIs, task as dummy-coded effect of image generation task compared to image inspection task, correctness as dummy-coded effect of correct compared to wrong retrievals and their interactions were modeled as fixed effects. Furthermore, subject and trial (i.e., object) were included as random intercepts, in order to control for possible differences between subjects and objects. In addition we specified the model to estimate precision of the beta distribution, the zero–one inflation probability and the conditional one-inflation probability as a function of AOI. As expected, participants spent more time in the corresponding AOI compared to non-corresponding AOIs. The estimated posterior distribution of the GLMM revealed no effect of task and correctness, showing similar fixation proportion during image generation and image inspection and for correct and wrong retrievals, respectively. The model also revealed no interactions, indicating similar effects in both tasks and for correctly and wrongly retrieved objects. This finding did not change, when adjusted for response time, speaking to the robustness of the results. Model parameters of the beta distributed means are reported in Table 1 and visualized in Fig. 3. Bayes factors of all main results are available in "Appendix C" and the parameters of the beta distribution's precision, the zero–one inflation and the conditional one inflation are available in "Appendix D".

Table 1 Table of logit-transformed regression coefficients (posterior mean, standard error, and 95% credible intervals) of the beta distribution (fixation proportion as a function of AOI, task and correctness of retrieval)

	Estimate	Est.Error	Q2.5	Q97.5
<i>Group-level effects</i>				
Subject (sd)	0.12	0.03	0.08	0.18
Trial (sd)	0.04	0.03	0.00	0.09
<i>Population-level effects</i>				
Intercept	0.13	0.26	-0.39	0.65
β_{NC}	-1.56	0.26	-2.08	-1.03
$\beta_{Im\ In}$	-0.14	0.33	-0.80	0.50
β_{True}	0.30	0.30	-0.30	0.88
$\beta_{NC: Im\ In}$	0.07	0.33	-0.56	0.74
$\beta_{NC: True}$	-0.41	0.30	-0.98	0.18
$\beta_{Im\ In: True}$	-0.44	0.38	-1.17	0.34
$\beta_{NC: Im\ In: True}$	0.53	0.39	-0.24	1.26

Estimates with credible intervals not including zero are indicated in bold. β_{NC} = non-corresponding. $\beta_{Im\ In}$ = image inspection. β_{True} = correct retrievals

Fig. 3 Posterior means and 95% credible intervals for the estimated fixation proportion per stimulus as a function of AOI, task and correctness of retrieval



4 Discussion

The primary goal of the current study was to examine the looking at nothing effect in an immersive virtual environment. Previous research has assessed the looking at nothing phenomenon in highly controlled 2D screen-based settings. For example, Martarelli and Mast (2011, 2013) found that when participants recalled objects from memory, they fixated more often the areas on a blank screen, where the objects were previously seen. Our findings provide the first direct evidence for the looking at nothing phenomenon in an immersive 3D setting. In an immersive virtual environment, we compared fixation proportion on those pedestals where the objects were previously inspected (corresponding area) with the other pedestals (non-corresponding areas). We observed higher proportion of fixation in the corresponding area, compared to non-corresponding areas. This finding extends external validity of previous findings suggesting that eye movements during pictorial recall indeed reflect the locations of where the stimuli were previously inspected (Johansson and Johansson 2014; Martarelli and Mast 2011, 2013; Scholz et al. 2016; Spivey and Geng 2001). However, it turned out, that in 89% of the time, participants fixated

none of the predefined areas. This finding is interesting, given the fact that most of the 2D screen-based studies do not provide any information about fixations outside the AOIs. In those studies, AOIs often cover the whole screen, leaving literally no or not much empty space to look at (e.g., Martarelli and Mast 2011, 2013; Johansson and Johansson 2014). This is a substantial difference to our experiment, which revealed that participants looked most of the time at locations outside the defined AOIs. In our experimental setting there were no restrictions with respect to eye movements and head-movements. Participants were able to look in all directions of the simulated environment. Besides, the dimension of the AOIs was defined by the maximal bounding box of the objects, leading to relatively small areas, compared to the entire space. Thus, the probability of fixations to empty space was relatively high, compared to fixations to AOIs. In previous studies, however, the probability of fixations coded as being outside the AOIs was small or inexistent. It is thus possible that IVR revealed less looking at nothing as a consequence of a narrow spatial dispersion of eye movements. Exploratory visualization of fixations in the 3D space indicates that participants tend to look at pedestals in the middle of the environment and at locations

which are centered around the fixation-square on the wall. This observation is in line with previous studies, suggesting lower dispersed eye-gaze patterns when stimuli are imagined rather than perceived (e.g., Brandt and Stark 1997; Gurtner et al. 2019; Johansson et al. 2006; Johansson et al. 2011).

Our study differed from previous research in yet another aspect, in that we investigated the occurrence of eye movements within a visual context. During retrieval the environment was not completely blank (as in the 2D looking at nothing paradigm) but provided critical visual landmarks (i.e., the pedestals) to look at. It is possible that the visual input in our experiment interfered with gazing back to the locations visited during encoding and thus could—at least partly—explain the high proportion of fixations on the empty space. This explanation is in line with previous research suggesting that perception of visual information within the looking at nothing paradigm interfere with eye movements during retrieval (Johansson and Johansson 2014; Kumcu and Thompson 2016, 2018; Scholz et al. 2016, 2018). For example, Kumcu and Thompson (2016, 2018) showed, that presenting an incongruent cue between encoding and retrieval, disrupt looking at nothing, whereas a congruent cue increase the effect.

4.1 Limitations and future directions

Some technical limitations should be considered when interpreting the findings. First, the number and time points of calibrations could be increased. For the present setup, we calibrated the eye-tracking system only at the beginning of the experiment, that is, before encoding started. It would be more accurate to calibrate the system several times, most importantly before retrieval starts. Clay et al. (2019) recommended repeated calibrations during a VR experiment to ensure exact tracking data. However, we chose an approach using predefined AOIs, which moreover were not bordering on each other. Thus, a single calibration at the beginning should have been sufficient. During encoding, participants fixated on average 76% of the time in the area, where they were supposed to look at, thus validating the measurement.

Second, although IVR offers high real-life resemblance, there are still some differences between viewing virtual worlds vs. real worlds. First, the resolution of HMDs is higher in the center compared to the periphery. If the user wants to see a target object in high resolution, he needs to move his or her head, whereas in real life eye movements

may be sufficient to bring the target object in the central visual field. However, recent technical advances in eye tracking-based foveated rendering can overcome this shortcoming. Second, in real-life viewing, accommodation is linked to vergence, but in IVR the lenses do not need to shift focus in order to maintain a clear image of different depth cues (accommodation). Visual input of varying depth in IVR is presented on a head-mounted display and therefore on a fixed distance (~5 cm) from the eyes. In a recent review, Harris et al. (2019) suggest, that such an artificial presentation of depth cues (vergence-accommodation conflict) may lead to less reliable binocular information and thus to fundamental different visual processing in IVR compared to real life.

To further assess how IVR affects looking at nothing, future research is needed in order to compare spatial dispersion of eye movements in IVR with a 2D looking at nothing paradigm. As stated above, fixations can be spatially constrained during retrieval. To what extent such an effect contributed to reduced looking at nothing in the current study and whether amplitude and frequency of eye-fixations in IVR differ from the traditional screen-based paradigm remains open. Moreover, further manipulation of external validity (i.e., approximation of real-life viewing) for example by using augmented reality, may help to clarify, whether looking at nothing in both, IVR and 2D paradigms are representative for eye movements in real life.

Future research will also need to be carried out in order to investigate the impact of additional visual information during retrieval on both eye movements and memory performance. In fact, there is evidence, that visual input can lead to a higher frequency of eye movements to relevant, empty locations. For example, Spivey and Geng (2001), showed stronger looking at nothing effects for conditions where a frame or a grid was provided, compared to the condition using the screen frame only (i.e., blank screen). However, additional visual information from the external world may also interfere with eye movements during retrieval. In the present study, looking at nothing in a completely blank environment was not examined. Adding such a condition but also conditions with varying degrees of visual information would be important in order to compare eye movements in a visual context vs. in a blank environment and to disentangle the relative influence of visual context on retrieval. It remains an open question to what extent eye-movement behavior and

performance was affected by the visual information in the present setting.

Moreover, there is a general need to further investigate the functional role of eye movements in memory performance. Some studies suggest, that looking at nothing during retrieval is related to memory strength (Johansson et al. 2011; Kumcu and Thompson 2016; Scholz et al. 2011; Wantz et al. 2015). Assessing the relation of vividness of mental imagery and looking at nothing may shed light on whether eye movements to nothing are executed to compensate for weak mental representation and thus could be used as a strategy in learning and knowledge organization. Indeed, re-enactment of processes (i.e., eye movements) has been shown to increase retrieval performance (e.g., Scholz et al. 2016). The question, whether this reenactment process translates to head-movements is important for an embodied view of cognition and should be addressed in future studies.

In sum, we investigated the looking at nothing effect during immersion in IVR. Participants showed high percentage of fixations to empty space because there were no restrictions to eye behavior. Yet, the looking at nothing effect could still be found when comparing corresponding and non-corresponding AOIs indicating possible target locations. Thus, eye movements during retrieval reflected spatial information associated with the encoded stimuli.

Even though our results suggest that looking at nothing may be overestimated in previous 2D screen-based settings, we still found the effect of looking back to empty locations during mental imagery and visual memory, thereby confirming ecological validity of previous findings. Given the growing use of IVR in cognitive rehabilitation settings

(Maggio et al. 2019) looking at nothing could thus be used as control for task performance in memory assessment and rehabilitation. The finding creates a starting point for future research in order to gain a more conclusive understanding of perception and visual behavior in 3D space. Thus, IVR could complement research for example in the field of eye-hand coordination and visual guided actions. We conclude, that eye tracking in IVR has a yet unexploited potential to investigate visual processing in relatively natural contexts while still maintaining high experimental control.

Acknowledgement We thank Simon Maurer from the Technology Platform of the Human Sciences Faculty for his valuable help in programming the experiment and designing the virtual environment.

Funding Open access funding provided by University of Bern.

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

Appendix A

A set of 38 computer-generated virtual objects including textures (in 37 cases) were gathered from websites offering free 3D models for VR and AR. In one case the texture

was costume-made using the 3D computer graphics software Blender 2.75a (<https://www.blender.org>). Each object belonged to one of five categories (animals, sports equipment, vehicles, technical devices and characters).

Category

Animals:



Sports:



Equipment:



 Vehicles:


 Characters:


Appendix B

Data analysis was performed in R (3.5.1; R Core Team). We followed a Bayesian approach using the *brms* package for Bayesian (non-)linear mixed models (Bürkner 2017). We used Markov Chain Monte Carlo (MCMC) method with

4 chains of 2000 iterations to calculate posterior parameter estimates.

Model notation for the *brms* model, implementing the zero–one-inflated beta model (1) and the prior distribution (2) via R.

```

model <- bf(time_proportion ~ aoi*task*correctness + (1|code) + (1|object),
            phi ~ aoi,
            zoi ~ aoi,
            coi ~ aoi,
            family = zero_one_inflated_beta()
            )
    
```

```

fit <- brm(
  formula = model,
  data = df,
  prior = priors,
  core = 4,
  iter = 2000,
  seed = 321,
  sample_prior = TRUE
)
    
```

```

priors = c(
  set_prior('normal(-1.6, 2)', class = 'Intercept'),
  set_prior('normal(0,2)', class = 'b'),
  set_prior('student_t(3,0, 0.1)', class = 'sd', group = 'object'),
  set_prior('student_t(3,0, 0.1)', class = 'sd', group = 'code'),
  set_prior('logistic(0,1)', class = 'b', dpar = "coi"),
  set_prior('logistic(0,1)', class = 'Intercept', dpar = 'coi'),
  set_prior('student_t(3, 0, 10)', class = 'b', dpar = 'phi'),
  set_prior('student_t(3, 0, 10)', class = 'Intercept', dpar = 'phi'),
  set_prior('logistic(0,1)', class = 'b', dpar = 'zoi'),
  set_prior('logistic(0,1)', class = 'Intercept', dpar = 'zoi')
)
    
```

Appendix C

We also calculated Bayes factors using the Savage-Dickey density ratio method provided by *brms* package for Bayesian (non-)linear mixed models (Bürkner 2017). Bayes factors should be taken with caution since they are highly sensitive to prior selection and their classification is based on simple rules of thumb set by practical consideration.

Logit transformed regression coefficients (posterior mean, standard error, 95% credible intervals and Bayes factor) of the beta distribution (fixation proportion as a function of AOI, task and correctness of retrieval).

	Estimate	Est. error	Q2.5	Q97.5	BF01
<i>Group-level effects</i>					
Subject (sd)	0.12	0.03	0.08	0.18	
Trial (sd)	0.04	0.03	0.00	0.09	
<i>Population-level effects</i>					
Intercept	0.13	0.26	-0.39	0.65	
β_{NC}	-1.56	0.26	-2.08	-1.03	0
$\beta_{Im In}$	-0.14	0.33	-0.80	0.50	6.01
β_{True}	0.30	0.30	-0.30	0.88	4.21
$\beta_{NC: Im In}$	0.07	0.33	-0.56	0.74	6.27
$\beta_{NC: True}$	-0.41	0.30	-0.98	0.18	2.82
$\beta_{Im In: True}$	-0.44	0.38	-1.17	0.34	2.65
$\beta_{NC: Im In: True}$	0.53	0.39	-0.24	1.26	2.01

β_{NC} = non-corresponding. $\beta_{Im In}$ = image inspection. β_{True} = correct retrievals

Appendix D

To compare the time spent in the corresponding area with the mean time spent in the five non-corresponding areas, we estimated fixation proportion during retrieval, as a function of the AOI, the task and correctness using a Bayesian zero–one-inflated beta model. The model considers a beta distribution for the continuous proportion outcome in the closed (0, 1) interval and a Bernoulli distribution for the binary {0, 1} outcome.

Model coefficients of the beta distribution's precision (ϕ) on the scale of the log-link function, the zero–one inflation (zoi) on the logit scale and the conditional one inflation (coi) on the logit scale for AOI as dummy-coded effect of the corresponding AOI compared to the mean of the non-corresponding (NC) AOIs.

	Estimate	Est. error	Q2.5	Q97.5
Phi_Intercept	1.00	0.11	0.77	1.22
Zoi_Intercept	1.76	0.10	1.58	1.96
Coi_Intercept	−0.29	0.09	−1.47	−1.11
Phi_AOI NC	2.75	0.13	2.50	3.00
Zoi_AOI NC	−3.24	0.13	−3.48	−3.00
Coi_AOI NC	−3.79	0.99	−6.04	−2.22

References

- Altmann GTM (2004) Language-mediated eye movements in the absence of a visual world: the “blank screen paradigm”. *Cognition* 93:79–87. <https://doi.org/10.1016/j.cognition.2004.02.005>
- Bone MB, St-Laurent M, Dang C, McQuiggan DA, Ryan JD, Buchsbaum BR (2019) Eye movement reinstatement and neural reactivation during mental imagery. *Cereb Cortex* 29(3):1075–1089. <https://doi.org/10.1093/cercor/bhy014>
- Brandt SA, Stark LW (1997) Spontaneous eye movements during visual imagery reflect the content of the visual scene. *J Cogn Neurosci* 9(1):27–38. <https://doi.org/10.1162/jocn.1997.9.1.27>
- Bürkner P-C (2017) brms: an R package for Bayesian multilevel models using Stan. *J Stat Softw* 80(1):1–28. <https://doi.org/10.18637/jss.v080.i01>
- Chang W, Cheng J, Allaire J, Xie Y, McPherson J (2019) shiny: Web application framework for R. <https://cran.r-project.org/package=shiny>
- Clay V, König P, König S (2019) Eye tracking in virtual reality. *J Mov Res* 12(1):3. https://doi.org/10.1007/978-3-319-08234-9_170-1
- Diemer J, Alpers GW, Peperkorff HM, Shibani Y, Mühlberger A (2015) The impact of perception and presence on emotional reactions: a review of research in virtual reality. *Front Psychol*. <https://doi.org/10.3389/fpsyg.2015.00026>
- Eichert N, Peeters D, Hagoort P (2018) Language-driven anticipatory eye movements in virtual reality. *Behav Res Methods* 50(3):1102–1115. <https://doi.org/10.3758/s13428-017-0929-z>
- Gorini A, Capideville CS, De Leo G, Mantovani F, Riva G (2011) The role of immersion and narrative in mediated presence: the virtual hospital experience. *Cyberpsychol Behav Soc Netw* 14(3):99–105. <https://doi.org/10.1089/cyber.2010.0100>
- Grogorick S, Michael S, Elmar E, Magnor M (2017) Subtle gaze guidance for immersive environments. In: SAP, p 7. <https://doi.org/10.1145/3119881.3119890>
- Gurtner LM, Bischof WF, Mast FW (2019) Recurrence quantification analysis of eye movements during mental imagery. *J Vis* 19(1):1–17. <https://doi.org/10.1167/19.1.17>
- Harris DJ, Buckingham G, Wilson MR, Vine SJ (2019) Virtually the same? How impaired sensory information in virtual reality may disrupt vision for action. *Exp Brain Res* 237(11):2761–2766. <https://doi.org/10.1007/s00221-019-05642-8>
- Johansson R, Holsanova J, Dewhurst R, Holmqvist K (2012) Eye movements during scene recollection have a functional role, but they are not reinstatements of those produced during encoding. *J Exp Psychol Hum Percept Perform* 38(5):1289–1314. <https://doi.org/10.1037/a0026585>
- Johansson R, Holsanova J, Holmqvist K (2006) Pictures and spoken descriptions elicit similar eye movements during mental imagery, both in light and in complete darkness. *Cogn Sci* 30(6):1053–1079. https://doi.org/10.1207/s15516709cog0000_86
- Johansson R, Holsanova J, Holmqvist K (2011) The dispersion of eye movements during visual imagery is related to individual differences in spatial imagery ability. In: Proceedings of the annual meeting of the cognitive science society, p 33. <https://escholarship.org/content/qt70782950/qt70782950.pdf>
- Johansson R, Johansson M (2014) Look here, eye movements play a functional role in memory retrieval. *Psychol Sci* 25(1):236–242. <https://doi.org/10.1177/0956797613498260>
- Kisker J, Gruber T, Schöne B (2019) Experiences in Virtual Reality entail different processes of retrieval as opposed to conventional laboratory settings: a study on human memory. *Curr Psychol*. <https://doi.org/10.1007/s12144-019-00257-2>
- Kruschke J (2015) Doing bayesian data analysis: a tutorial with R, JAGS, and Stan. Elsevier, Amsterdam
- Kumcu A, Thompson RL (2016) Spatial interference and individual differences in looking at nothing for verbal memory. In: Proceedings of the 38th annual conference of the cognitive science society
- Kumcu A, Thompson RL (2018) Less imageable words lead to more looks to blank locations during memory retrieval. *Psychol Res*. <https://doi.org/10.1007/s00426-018-1084-6>
- Laeng B, Bloem IM, D'Ascenzo S, Tommasi L (2014) Scrutinizing visual images: the role of gaze in mental imagery and memory. *Cognition* 131(2):263–283. <https://doi.org/10.1016/j.cognition.2014.01.003>
- Laeng B, Teodorescu DS (2002) Eye scanpaths during visual imagery reenact those of perception of the same visual scene. *Cogn Sci*. [https://doi.org/10.1016/S0364-0213\(01\)00065-9](https://doi.org/10.1016/S0364-0213(01)00065-9)
- Lee MD, Wagenmakers E-J, Trafimow D (2005) Bayesian statistical inference in psychology: comment on trafimow. *Psychol Rev* 112(3):662–668. <https://doi.org/10.1037/0033-295X.112.3.662>
- Lohr DJ, Friedman L, Komogortsev OV (2019) Evaluating the data quality of eye tracking signals from a virtual reality system: case study using SMI's eye-tracking HTC vive. <https://arxiv.org/abs/1912.02083>
- Maggio MG, Maresca G, De Luca R, Stagnitti MC, Porcari B, Ferrara MC, Calabrò RS (2019) The growing use of virtual reality in cognitive rehabilitation: fact, fake or vision? A scoping review. *J Natl Med Assoc* 111(4):457–463. <https://doi.org/10.1016/j.jnma.2019.01.003>
- Makowski D, Sperduti M, Nicolas S, Piolino P (2017) “Being there” and remembering it: presence improves memory encoding. *Conscious Cogn* 53:194–202. <https://doi.org/10.1016/j.concog.2017.06.015>
- Makransky G, Terkildsen TS, Mayer RE (2019) Adding immersive virtual reality to a science lab simulation causes more presence but less learning. *Learn Instr* 60:225–236. <https://doi.org/10.1016/j.learninstruc.2017.12.007>

- Martarelli CS, Chiquet S, Laeng B, Mast FW (2017) Using space to represent categories: insights from gaze position. *Psychol Res* 81(4):721–729. <https://doi.org/10.1007/s00426-016-0781-2>
- Martarelli CS, Mast FW (2011) Preschool children’s eye-movements during pictorial recall. *Br J Dev Psychol* 29(3):425–436. <https://doi.org/10.1348/026151010X495844>
- Martarelli CS, Mast FW (2013) Eye movements during long-term pictorial recall. *Psychol Res* 77(3):303–309. <https://doi.org/10.1007/s00426-012-0439-7>
- Moreno R, Mayer RE (2002) Learning science in virtual reality multimedia environments: role of methods and media. *J Educ Psychol* 94(3):598–610. <https://doi.org/10.1037/0022-0663.94.3.598>
- Ospina R, Ferrari SLP (2012) A general class of zero-or-one inflated beta regression models. *Comput Stat Data Anal* 56(6):1609–1623. <https://doi.org/10.1016/j.csda.2011.10.005>
- Richardson DC, Spivey MJ (2000) Representation, space and hollywood squares: looking at things that aren’t there anymore. *Cognition*. [https://doi.org/10.1016/S0010-0277\(00\)00084-6](https://doi.org/10.1016/S0010-0277(00)00084-6)
- Riva G, Mantovani F, Capideville CS, Preziosa A, Morganti F, Villani D, Alcañiz M (2007) Affective interactions using virtual reality: the link between presence and emotions. *Cyberpsychol Behav* 10(1):45–56. <https://doi.org/10.1089/cpb.2006.9993>
- Scholz A, Klichowicz A, Krems JF (2018) Covert shifts of attention can account for the functional role of “eye movements to nothing”. *Mem Cogn* 46(2):230–243. <https://doi.org/10.3758/s13421-017-0760-x>
- Scholz A, Mehlhorn K, Bocklisch F, Krems JF (2011) Looking at nothing diminishes with practice. In: Proceedings of the 33rd annual meeting of the cognitive science society, pp 1070–1075
- Scholz A, Mehlhorn K, Krems JF (2016) Listen up, eye movements play a role in verbal memory retrieval. *Psychol Res* 80(1):149–158. <https://doi.org/10.1007/s00426-014-0639-4>
- Slater M (2009) Place illusion and plausibility can lead to realistic behaviour in immersive virtual environments. *Philos Trans R Soc B Biol Sci* 364(1535):3549–3557. <https://doi.org/10.1098/rstb.2009.0138>
- Spivey MJ, Geng JJ (2001) Oculomotor mechanisms activated by imagery and memory: eye movements to absent objects. *Psychol Res* 65(4):235–241
- Ventura S, Brivio E, Riva G, Baños RM (2019) Immersive versus non-immersive experience: exploring the feasibility of memory assessment through 360° technology. *Front Psychol*. <https://doi.org/10.3389/fpsyg.2019.02509>
- Wagenmakers E-J (2007) A practical solution to the pervasive problems of *p* values. *Psychon Bull Rev* 14(5):779–804. <https://doi.org/10.3758/BF03194105>
- Wantz AL, Martarelli CS, Mast FW (2015) When looking back to nothing goes back to nothing. *Cogn Process* 17(1):105–114. <https://doi.org/10.1007/s10339-015-0741-6>

Publisher’s Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.