A Study of Change Blindness in Immersive Environments

Daniel Martin, Xin Sun, Diego Gutierrez, Belen Masia

Fig. 1: We present a quantitative, systematic study of the phenomenon of change blindness in immersive environments (leftmost illustration). We analyze the effect of different parameters such as the type of manipulation, distance to the observer, complexity, and location with respect to the field of view (top). We then explore the influence of our limited visual memory when many different changes take place at the same time (bottom).

Abstract— Human performance is poor at detecting certain changes in a scene, a phenomenon known as *change blindness*. Although the exact reasons of this effect are not yet completely understood, there is a consensus that it is due to our constrained attention and memory capacity: We create our own mental, structured representation of what surrounds us, but such representation is limited and imprecise. Previous efforts investigating this effect have focused on 2D images; however, there are significant differences regarding attention and memory between 2D images and the viewing conditions of daily life. In this work, we present a *systematic* study of change blindness using immersive 3D environments, which offer more natural viewing conditions closer to our daily visual experience. We devise two experiments; first, we focus on analyzing how different change properties (namely *type*, *distance*, *complexity*, and *field of view*) may affect change blindness. We then further explore its relation with the capacity of our visual working memory and conduct a second experiment analyzing the influence of the *number* of changes. Besides gaining a deeper understanding of the change blindness effect, our results may be leveraged in several VR applications such as redirected walking, games, or even studies on saliency or attention prediction.

Index Terms—Virtual reality, change blindness, visual working memory, attention

1 Introduction

Change blindness is a psychological phenomenon that refers to the human inability to detect changes to an object or scene, even when they would appear to be obvious [16, 17, 25, 35, 36]. Although many factors may come into play, such as age or attention [6, 38], a vast body of literature from psychology and cognitive sciences attributes this effect to limitations in our visual working memory [9, 45]; given our limited brains' capacity [5, 10, 31], significant amounts of visual information must be discarded after being processed[1].

While this phenomenon has been known for decades, most of the efforts to study it have been carried out from a qualitative perspective, and using 2D images. However, it has been shown that there are significant differences between attention- and memory-related experiments using 2D images and immersive 3D environments, since the latter offer more natural viewing conditions, closer to daily visual experience [15, 22]. For instance, spatial learning in 3D leverages proprioceptive and vestibular feedback, while whole-body motion plays a key role in building relationships between objects and the observer [11, 22]. These two important aspects of visual perception are rarely present when viewing 2D images.

In this work, we present a systematic study of the change blindness effect, leveraging natural viewing conditions simulated with VR 3D environments. In particular, we have designed two different experiments: in the first one, we explore the effect of the *type* of manipulation (i.e., adding, removing, relocating, or replacing elements), the *distance* to the manipulation, the *complexity* of the manipulated element, and the influence of the *field of view* (FoV). We have found that, despite the

[1]We refer the reader to the video "*Whodunnit*" for an illustrative example of this effect (https://www.youtube.com/watch?v=ubNF9QNEQLA).

- *Correspondence to: danims@unizar.es*
- *Daniel Martin, Diego Gutierrez, and Belen Masia are with Universidad de Zaragoza, I3A.*
- *Xin Sun is with Adobe Research.*
- *Daniel Martin was intern in Adobe Research for three months during this project.*

Manuscript received xx xxx. 201x; accepted xx xxx. 201x. Date of Publication xx xxx. 201x; date of current version xx xxx. 201x. For information on obtaining reprints of this article, please send e-mail to: reprints@ieee.org. Digital Object Identifier: xx.xxxx/TVCG.201x.xxxxxxx

changes being apparently overt, humans generally miss more than half of the manipulations, even in the more conservative case when they know in advance that there will be changes in the scene. Previous works with 2D images suggest an observers' accuracy to detect such changes ranging from 73% to 92% [16, 17]. Ma et al. [25] reported that for up to a 20% of the changes it took people more than 45 seconds to detect them; working in VR, we found that participants could not find the manipulation within 45 seconds in more than half of the cases. We also found that manipulations on closer or complex elements are detected more easily (as roughly suggested by previous work [46]), while, unlike what has been proposed for 2D images [25], the type of the manipulation did not have a significant effect on its detectability. We also found that manipulations happening inside the users' FoV were more easily detected (despite masking them with a blinking paradigm) than those happening outside such FoV. In any case, approximately 40% of the manipulations in the observers' FoV were not detected, which represents a higher ratio of undetected changes than those reported by previous work [17, 46].

In order to have control over the conditions tested, our main experiment was designed with a single change happening in each trial. In a second experiment, we analyze the influence of the *number* of manipulations in the change blindness phenomenon. Our results suggest that humans are able to correctly detect less than three manipulations, and that their ability to detect four or more changes is limited regardless of the amount of actual changes in the scene, i.e., even with a large numbers of manipulations, humans are consistently not able to detect more than four. Previous work with 2D content has shown that memory performance decreases when four or more changes need to be remembered [2, 14, 31]; our results seem to confirm this, even suggesting that such limitation is slightly stronger in VR. Last, we found a significant correlation between the participants' ability to detect changes and the time being immersed in a virtual scene, with higher detection ratios with longer times. However, such ability is limited, with almost 40% of the changes remaining undetected regardless of time for the time frame we tested. This seems to agree with the hypothesis that, while more time allows for more solid internal representations of the 3D scene, memory remains a limiting factor.

A deeper understanding of the change blindness effect in 3D immersive environments will help predict to what extent certain manipulations will be visible by a human observer, or remain unnoticed. This in turn may lead to improved users' experiences in different VR applications, including redirected walking [40, 41], cinematography [34], interior design [1], saliency detection [7, 48], or gaze prediction [28].

2 Related Work

Change blindness has interested many multidisciplinary researchers for decades. Simons and Levin [35] analyzed many experiments and reported how, while based on diverse methods, they produced strikingly similar results: Most of the visual changes in a scene remained undetected when viewing 2D images. Attwood et al. found later a similar conclusion when viewing museum artifacts in the real world [3].

While many works agree that attention is required to be able to perceive such changes [32, 36], it is generally agreed that change blindness is related to our brains' limited storage capacity, and to the very little visual information that actually seems to be stored from one view or instant to the next. Our visual working memory is used to hold certain information for any posterior task [4], but its storage capacity decreases significantly when more than four elements are to be remembered [12, 24]. Indeed, there is a trade-off between the amount of elements one is able to remember, and the fidelity of their stored representation [2, 5, 31]. Some researchers have tried to devise models to represent how this memory works, and how information is represented in the brain (see the review by Brady and colleagues [9]), proposing that item information is stored in a hierarchical fashion, based on their context and interactions. Experiments by Hollingworth and Henderson further explored the relation between visual memory for objects and scene context [16–18]. From this body of existing literature, Ma et al. [25] introduced a computational method to synthesized changes in 2D images with a certain degree of *blindness*.

Fig. 2: Overview of the scene used in our first experiment (see Section 3.1), which depicts a living room interior provided with realistic furniture. Participants were seated while doing the experiment, to avoid translational movements, and were placed in the same virtual position as where the panorama picture has been taken.

While there is a vast literature about change blindness, most of the efforts to reach a better understanding of the phenomenon have been done using traditional 2D images, with the few works using VR focusing on particular *applications*. Steinicke et al. [39] investigated change blindness in stereoscopic displays, showing similar trends as in monocular ones, and suggesting that VR systems could also leverage the effect. However, their studies were limited to static images shown through an HMD. Suma et al. [40] leveraged change blindness to manipulate a 3D environment and thus redirect movement in VR scenarios [21, 33, 42], avoiding potential visual-vestibular inconsistencies. Later, Lohse et al. [23] evaluated using change blindness for haptic remapping in VR, to allow using the same physical object to interact with multiple virtual objects. Despite leading to some interaction errors, the manipulations remained undetected for most of the participants.

Closer to our work, Vasser et al. [46] analyzed whether changes in the background or foreground of virtual environments presented different detectability, with results confirming the previously reported bias favoring foreground objects [29, 44]. Despite the shared goal of analyzing change blindness in VR, a single background/foreground parameter is analyzed. In contrast, we conduct a systematic study on a much wider set of parameters using realistic 3D environments, and provide a through quantitative analysis of the results.

3 Main Experiment

In this first experiment, our goal is to perform a systematic analysis of change blindness in immersive environments. Within this broad goal, we first want to evaluate whether certain types of changes are more prone to be detected than others; to this end, our stimuli feature four types of changes, namely addition, removal, relocation and replacement of an element (i.e., an object, or a set of them). Unlike in traditional media, having immersive content enables making changes both inside or outside the field of view of the observer, and we therefore also analyze whether this has an effect on detectability. Finally, we look into whether the distance to the manipulated element (note that content is viewed stereoscopically) or the complexity of such element have an influence on detectability.

3.1 Stimuli

Participants viewed a realistic, stereoscopic virtual scene created in Unity[2], depicting a living room with realistic furniture (see Figure 2). Within such scene, and in each trial of the experiment, one element is subject to manipulation. Inspired by previous work on change blindness in traditional media (images) [25], we implement and analyze four manipulation operators:

- **Addition** An element is inserted in an empty location of the scene (without intersecting other elements).

[2]The scene for this experiment was *Living Room Interior* by *Blue Dot Studios*, and can be found in the Unity Asset Store.

Fig. 3: We depict a sample manipulation for each of the main four types we have analyzed (see Section 3.1). (*Top left*) Sample addition: We add a fourth basket lamp. (*Top right*) Sample removal: We remove one of the chairs (and adjust the position of the remaining two). (*Bottom left*) Sample replacement: We swap a plant for a floor light. (*Bottom right*) Sample relocation: We move a chest to a different position. Note that manipulations happen at different distances and affect objects with varied complexities.

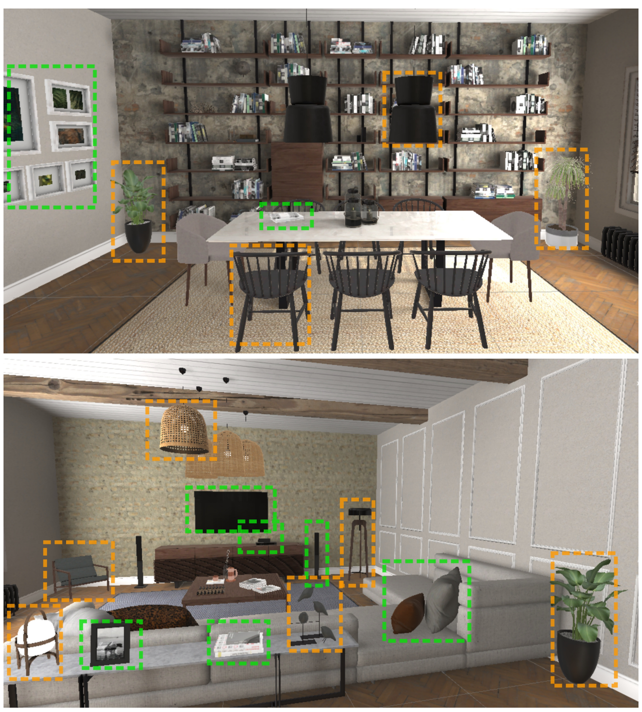

Fig. 4: We depict two viewpoints of the scene we used in our main experiment (Figure 2). For each of them, we outline in orange changes in items of high complexity, while outline changes of lower complexity in green.

- **Removal** An existing element is removed from the scene.
- **Relocation** An existing element is moved to a different, empty location of the scene (without intersecting other elements), or two existing elements exchange their positions.
- **Replacement** An existing element is replaced by a different element, in the same position (e.g., a floor lamp is replaced by a plant).

For each type of manipulation operator, we implement seven different fixed changes, for a total of 28 different possible manipulations (see Figure 3 for an example of each type). The designed manipulations cover a range of distances from the manipulation to the observer, and degrees of complexity of the element. For each of these two factors, we divide our stimuli into two groups: close and far for distance (setting the threshold between both at 1 meter), and complex (i.e., high complexity) and simple (i.e., low complexity) for complexity (by manual classification, see Figure 4 for examples on this feature). In each manipulation, the changed element has two different states, *manipulated* and *original*. Every time the participants rotate their head and the element gets out of their field of view (FoV), its state changes to its counterpart, in an adaptation of the well-known flicker paradigm of traditional media approaches [16, 17, 25] to our immersive 360° setup [46]. We denote this as the *outside FoV* condition, in contrast to the *inside FoV* one introduced in the next paragraph. Note that we are conservative, and do not change the element's state unless it has been within the central 60° of the participant's horizontal FoV.

To act as a baseline for our results, and to analyze the difference between detectability of changes occurring inside or outside the FoV, we selected seven of the previous 28 manipulations and implemented their *inside FoV* counterpart. In this version, whilst the changed element is within the central 60° of the participant's horizontal FoV, a traditional flicker paradigm takes place: The environment turns to a neutral gray for 250ms, during which the manipulation takes place, and the scene then shows the *manipulated* state of the element; after one second, the neutral gray environment is shown for 250ms again, and the change is reversed, showing the *original* state of the element. This is repeated whilst the element is within the central 60° of the participant's horizontal FoV. We select a flicker time of 250ms because it is within the range of blink duration (usually ranging from 100 to 400 ms [8, 21, 30, 43], with a reported average duration of around 300 ms [13]).

3.2 Hardware

Our stimuli were presented on a Meta Quest 2 head-mounted display (HMD) with a horizontal FoV of 104° and a vertical FoV of 98°, a resolution of 1832x1920 pixels per eye, and a frame rate of 90 fps. Participants could select the position of the manipulation with a Meta Quest 2 controller by pointing and pressing a trigger. The Meta Quest 2 are provided with inside-out tracking. The participants reported no perceived errors in head or hand tracking. However, since the Meta Quest 2 is a standalone HMD, we logged tracking data (i.e., head position and orientation) and data about the pointed elements, and checked after each experiment for potential data loss or errors.

3.3 Participants

Twenty-two participants (11 identified themselves as female, 11 identified themselves as male, and none of them identified as non-binary,

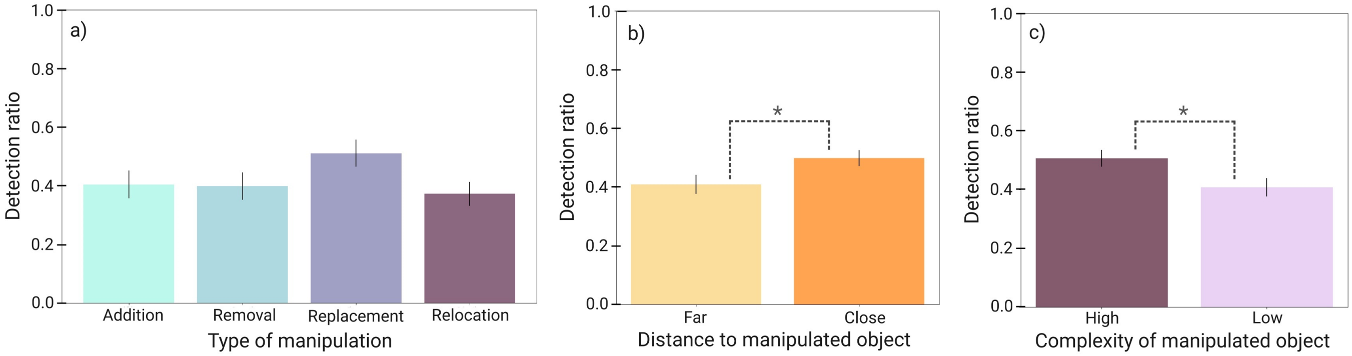

Fig. 5: We have analyzed the effect of different manipulation features (x axes) on the detection ratio (y axes). (a) Detection ratios for each of our four outside FoV manipulation types, namely addition, removal, replacement, and relocation. Our statistical analyses do not yield any significant difference between any types. (b) Effect of distance to manipulated element. Manipulations on closer items are more easily detected than on further ones, in agreement with previous literature [46]. (c) Effect of the item complexity. Manipulations on more complex elements (i.e., objects with more complex geometries) are more easily detected. Bars depict the standard error of the mean (SEM). The asterisk indicates significant ($\alpha < 0.05$) differences.

other, or preferred not to say; average age 27.18 years old, $STD = 8.53$) voluntarily took part in the study and provided written consent for participation. They were naïve about the final purpose of the experiment, and they all reported normal or corrected-to-normal vision. 81.80% of them had already used an HMD before, although with different frequencies (44.80% occasionally, 27.80% often, and 27.80% very often). The experimental procedure was approved by our University Ethics Committee.

3.4 Procedure

Participants sat on a rotating stool, to prevent translational movements and keep all of them in the same position within the virtual environment. The experiment routine was implemented in Unity.

We followed a within-subjects experimental design. Each participant conducted a total of 35 trials, including the 28 *outside-FoV* manipulations and the additional seven *inside-FoV* ones (see Section 3.1). The order of trials was randomized between participants. In each trial, the participant is asked to explore the scene and find the manipulated element. Once found, they should point at it with the controller and press the trigger to confirm the choice. Participants have 45 seconds to find the manipulated element, otherwise the trial yields a timeout. This maximum duration was empirically set, in order to ensure that participants had enough time to explore the whole scene at least twice [37], while keeping the experiment length reasonable to avoid fatigue (i.e., below 30 minutes per participant). Regardless of whether the participant detects the element or there is a timeout, they have to press the trigger to begin the next trial; this allows participants to rest and reposition themselves between trials if necessary.

The whole experiment (including previous and posterior questionnaires) took between 25 and 30 minutes to complete. Before the experiment, users gave written consent for participation and filled a pseudo-anonymous demographic questionnaire, and a pre-experiment sickness questionnaire (SSQ) [19]. Then, the experimenter gave the instructions for the experiment, and the participant was presented a simple tutorial session, to familiarize themselves with the procedure. Once the participant completed the tutorial, the experiment would commence. After 18 trials, the participants had to take a short break (at least 10 seconds). At the end of the experiment, after completing the 35 trials, the participant completed another post-experiment sickness questionnaire and a short debriefing. We refer the reader to the supplementary video for a short overview of the experiment.

4 Main Experiment: Results

We measure the detectability of a condition as the ratio between correctly detected and total number of manipulations in that condition, and analyze this detectability for the four different *types* of manipulations, and for the two levels of both *distance* and *complexity* factors, as described in the previous section.

Collected data We collected data for a total of 770 trials. Within the answered trials, there are correct answers (the participant correctly identified the element that had changed), incorrect answers (the participant chose an element, but it was not the one that had changed), and timeouts. For the analysis shown here, we leave the incorrect trials out and focus only on the trials which were either responded correctly (we are sure the participant identified the change), or timed out (we are sure the participant did not see the change). The supplementary includes the full data. We discard trials in which the participant provided a very fast answer; specifically, we set a threshold τ for the total time that the manipulated element or its counterpart were within the FoV (including the time to select it). We select it based on the interquartile range, such that $\tau = AVG$ - $1.1 * IQR = 4.03s$; this leads to the removal of 32 trials.

Analysis Since we are dealing with nonparametric, unbalanced data and multiple factors, we apply the Aligned Rank Transform (ART) for non-parametric factorial ANOVA [20,47]. ART adds a preprocessing step to the data prior to computing averaged ranks, after which common ANOVA procedures can be used. We additionally conducted post-hoc pairwise Bonferroni-corrected comparisons. We achieved a statistical power of 0.9251 ($\alpha = 0.05$, sample size = 21, and effect size f = 0.25). Additional details on the post-hoc pairwise t-test can be found in Table 1 to Table 4. We have additionally conducted multiple comparisons of means with a 95% family-wise confidence level (Tukey HSD test). The yielded confidence intervals can be found in the supplementary material.

4.1 Results

Our results show a *detection ratio* of 46.22% across all trials, indicating that despite being specifically asked to look for manipulations for 45 seconds, more than half of the manipulations remained undetected. In the case of the *outside FoV* condition, this detection ratio was 41.97%, while it increased to 61.66% in the *inside FoV* condition. Results reported hereafter are referred to the *outside FoV* condition, except when specifically indicated otherwise.

For those detected, the average *time to detect* was $t_d = 23.51$s (STD = 10.43s). We also compute the average time to detect the manipulation *after* the object changed had been in their FoV, which is $t_{d,a} = 18.46$s (STD = 11.21s). To assess both whether the participants actually spent time with the manipulation inside their FoV, and whether they alternated between both states of the manipulated object, we compute: (i) the average aforementioned time, $t_{in} = 10.79s$ (STD = 7.83s); and (ii) the average number of times the manipulation entered the participant's FoV,

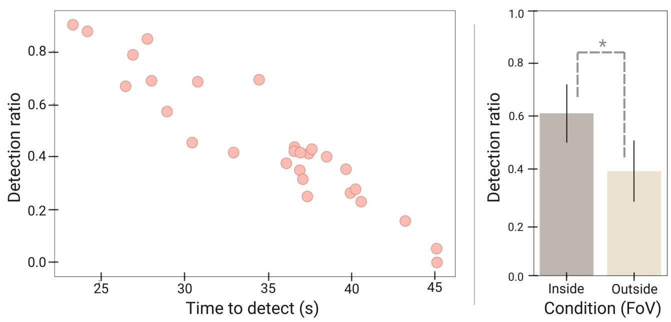

Fig. 6: *Left:* Scatter plot depicting the correlation between *time to detect* t_d (x axis) and *detection ratio* (y axis). Note that the lower the detection ratio, the longer it takes for participants to detect such changes. *Right:* Detection ratio (y axis) for the *inside FoV* and *outside FoV* conditions. Our results indicate that manipulations happening in the field of view are more easily detected, which is consistent with literature indicating that change blindness is inherently related with working memory, and that detection ability degrades as more elements are tried to be remembered. Bars depict the standard error of the mean (SEM). The asterisk marks significant ($\alpha < 0.05$) differences.

which is also the number of times the state of the object was changed, n_{ch} = 5.12 (STD = 2.42). Both clearly indicate that the changes were visible to the participants, yet remained undetected in a number of cases.

We further looked into the relationship between the detection ratio, and the time required to detect the change. To do this, we compute the detection ratio per manipulation, and the corresponding average t_d. The results show a strong negative correlation between both: $r = -0.9382$, $p < 0.0001$ for Pearson's correlation test (Spearman's and Kendall's correlation tests also reflect this trend and can be found in the supplementary); we additionally depict such correlation in Figure 6, left.

Our pre- and post-experiment sickness questionnaires showed that no participant experimented moderate or severe sickness symptoms; we refer the reader to Section S.3 in the supplementary material for the detailed results.

Effect of manipulation type, distance and complexity We analyze the influence of the *type* of manipulation (addition, removal, relocation or replacement), its *distance* to the user (close or far), and the *complexity* of the element being manipulated (high or low) on the detection ratio (see Figure 5). We found a significant effect of both *complexity* ($F = 5.174, p < 0.05$) and *distance* ($F = 4.664, p < 0.05$) on detection ratio. On the other hand, we found no significant effect of the *type* of manipulation ($F = 0.166, p > 0.05$).

We have also analyzed first-order interactions: While the effect of interactions between type and distance on detection ratio was not significant ($F = 0.462, p > 0.05$), interactions between type and complexity, and distance and complexity were significant ($F = 5.716, p < 0.01$ and $F = 10.866, p < 0.001$, respectively). While for the former there is no clear pattern, possibly due to the multiple combinations and the particular nature of each manipulation, for the latter (*distance***complexity*) we found that manipulations on simple items that happen far from the participant have significantly lower detectability, as one could expect.

Additional details on the detection ratio of all the combinations of such properties can be found in Table 5.

Effect of the field of view While our previous analyses have focused on manipulations that happened outside the participants' FoV (*outside FoV* condition), we analyze here the differences between the same manipulations taking place inside and outside the field of view. To do this, and as explained in Section 3.1, in our experimental design we additionally implemented an *inside FoV* counterpart for seven out of the 28 manipulations. The following data thus stems from the analysis of those fourteen manipulations, seven taking place inside the FoV and

Table 1: Results for the statistical analysis (Aligned Rank Transform (ART) for non-parametric factorial ANOVA analysis ($\alpha = 0.05$) as proposed by Wobbrock et al. [47], see Section 4 in the main document) on the effect of type of manipulation, distance to the manipulated object, and complexity, from our first experiment.

		F	Pr (>F)	
1	Type	0.16651	0.91892657	
2	Distance	4.66489	0.03137207	*
3	Complexity	5.17436	0.02344709	*
4	Type:Distance	0.46214	0.70887683	
5	Type:Complexity	5.71599	0.00077128	***
6	Distance:Complexity	10.86615	0.00106569	**
7	Type:Distance:Complexity	12.13378	0.00054963	***

Table 2: Results for the post-hoc comparisons (Detected ≈ Type + Distance). P-values are Bonferroni-corrected. (*Type:* AD: Addition, RE: Removal, RL: Relocation, RP: Replacement. *Distance:* CL: Close, FA: Far). Note that no interaction is significant (p > 0.05).

	AD:CL	AD:FA	RP:CL	RP:FA	RE:CL	RE:FA	RL:CL
AD:FA	1.00	-	-	-	-	-	-
RP:CL	1.00	1.00	-	-	-	-	-
RP:FA	1.00	1.00	1.00	-	-	-	-
RE:CL	1.00	1.00	0.99	1.00	-	-	-
RE:FA	1.00	1.00	1.00	1.00	1.00	-	-
RL:CL	1.00	1.00	0.10	0.23	1.00	0.14	-
RL:FA	1.00	1.00	1.00	1.00	1.00	1.00	0.13

Table 3: Results for the post-hoc comparisons (Detected ≈ Type + Complexity). P-values are Bonferroni-corrected. (*Type:* AD: Addition, RE: Removal, RL: Relocation, RP: Replacement. *Complexity:* SI: Simple (i.e., low complexity), CO: Complex (i.e., high complexity)). Note that some pairs of interactions are significantly different (p < 0.05).

	AD:CO	AD:SI	RP:CO	RP:SI	RE:CO	RE:SI	RL:CO
AD:SI	1.00	-	-	-	-	-	-
RP:CO	1.00	1.00	-	-	-	-	-
RP:SI	0.54	0.09	0.62	-	-	-	-
RE:CO	1.00	1.00	0.99	1.00	-	-	-
RE:SI	1.00	1.00	1.00	0.03	0.93	-	-
RL:CO	1.00	0.51	1.00	1.00	1.00	0.28	-
RL:SI	1.00	1.00	1.00	0.01	0.28	1.00	0.10

Table 4: Results for the post-hoc comparisons (Detected ≈ Distance + Complexity). P-values are Bonferroni-corrected. (*Distance:* CL: Close, FA: Far. *Complexity:* SI: Simple (i.e., low complexity), CO: Complex (i.e., high complexity)). Note that some pairs of interactions are significantly different (p < 0.01).

	CL:CO	CL:SI	FA:CO
CL:SI	4.4e-5	-	-
FA:CO	1.00	1.3e-4	-
FA:SI	1.00	4.4e-5	1.00

seven outside it. Note that the trial order is fully randomized, counterbalancing inside and outside FoV counterparts. Detection ratio for manipulations in the *inside FoV* condition was 61.66%, significantly higher than the 39.48% detection ratio in *outside FoV* condition ($F = 13.73, p < 0.001$). While there is a significant difference between both conditions (Figure 6, right), there are significant variations in detectability differences between conditions for different manipulations, which would require further investigation.

4.2 Discussion

Our results suggest that the *type* of manipulation had no significant effect on the detection ratio, and actually show very similar ratios across types, with only replacement ones being slightly higher (Figure 5). This

Table 5: Detection ratio for different combination of manipulation properties. The first eight rows correspond to combining *Complexity* and *Type*, while the last four rows correspond to combining *Complexity* and *Distance*. Manipulations on complex elements or in elements that are close are usually easier to be detected (see Figure 5). Besides, some particular combinations yield manipulations that are harder to be detected (e.g., changing one simple item to another is far easier to detect than relocating a complex item). We mark significant differences as follow: Each pair of properties marked with a blue shape is significantly different from its red counterpart (e.g., Simple + Removal is significantly different from Simple + Replacement).

	Manipulation properties	Detection ratio
○ (red)	Simple + Replacement	67.85%
	Complex + Relocation	60.00%
	Complex + Removal	50.81%
	Complex + Replacement	43.05%
	Complex + Addition	42.66%
○ (blue)	Simple + Removal	31.42%
○ (blue)	Simple + Addition	31.38%
○ (blue)	Simple + Relocation	26.92%
× (blue)	Complex + Far	52.89%
× (blue)	Simple + Close	50.93%
× (blue)	Complex + Close	48.82%
× (red)	Simple + Far	21.83%

is in contrast with some previous work in traditional media [25], where the manipulation operator was considered to be relevant. Since they employ a paradigm in traditional media (a flickering gray screen separating the two states of the manipulated object) that is different from our *outside FoV* condition (the change between states happens when the object is outside the FoV), it may be that in their case, in the absence of head movement and other portions of the scene being seen, the differences between relocation, replacement and addition/removal become larger. Additionally, we should note that addition and removal share a higher similarity between them than the rest (one is the counterpart of the other), and so do their corresponding results.

We also observe that both distance and complexity do play a significant role in detectability. This is consistent with previous literature suggesting that visual memory works in a contextual fashion [16], and thus we store information about context and interactions between scene and elements. In this line, previous work has also reported that changes in the foreground are usually easier to detect than changes in the background [46], which is consistent with our observations. While the results we find in this regard can be explained and are consistent with previous literature, the division in two complexity levels was done manually and based on observation; a more in-depth study would thus be required to draw strong conclusions in this regard.

When analyzing how those factors interact, we found that participants had better performance on manipulations of complex elements, in most cases regardless of the type or distance. However, we found that they performed better when detecting replacements of simpler elements. While this needs further research, we hypothesize that it may be due to the fact that, while replacing does not change the spatial representation of our memory, simple elements are probably stored with much less detail, therefore passing unnoticed when changed to other simple items.

We added manipulations inside the FoV for a subset of our trials, to act as a baseline for comparison. As expected, they are more easily detected than changes outside the FoV. This result can be explained by the literature: A number of works have shown that human memory degrades as more information has to be maintained [12]. Still, it is worth noting that even in this *inside FoV* condition, almost 40% of the manipulations remained undetected. This figure is still larger than those found by previous works with 2D images in traditional media, which report that observers missed from 8% to 27% of the manipulations [16, 17]. The more complex nature of our 360° environment could be an explanation for this difference.

Fig. 7: Overview of the scene used in our validation experiment (see Section 4.3), which depicts a living room interior provided with some furniture. Participants saw three different variations of the scene, each of them having five different manipulations, selected from three different sets of manipulations (with different detection ratios, following our main experiment's results, see Table 5).

Table 6: Results from our validation experiment. We show the average ratio of correctly detected changes, missed changes, and incorrect responses, for each variation (see Section 4.3). Participants' ability to detect manipulations was higher for the variations with sets of higher detection ratios (and thus they missed less changes). The ratio of incorrect answers remained stable for the more complex variations, while slightly decreased for the simplest one.

Scene var.	Correct ratio	Missed ratio	Incorrect ratio
#1 - Complex	0.40	0.60	0.20
#2 - Medium	0.52	0.48	0.20
#3 - Simple	0.68	0.32	0.04

Since the experiment is performed with one scene, we consider whether learning and habituation effects could enhance participants' performance over time. While we found no significant effect of trial order on detection ratio, we empirically observed a slow yet steady increase in detection ratio, which stabilized at 50.57% after the first fifteen trials, suggesting a slight tendency to improvement (average for the first fifteen was 40.12%).

Finally, as explained in Section 4.1, we have conducted our analysis with data from participants that either responded correctly or timed out. We discard *incorrect* answers since within our paradigm, there is no way of disambiguating whether an incorrect answer comes from a random response (pure guess) or from mistaking the manipulated element.

4.3 Validation

Our results suggest that certain interactions between manipulations and properties of the manipulated element may ease or hinder the detection of such manipulation. To validate those insights, we have devised a small proof of concept experiment in a different scenario[3], depicted in Figure 7. For this experiment, we have designed three different arrangements of the scene, to avoid habituation effects. In each of the three scene variations, we have included five different manipulations. For the first variation, we have included five manipulations pertaining to the highest detection ratio yielded by our previous results (first two rows of Table 5), with an average detection ratio of 66.28%; the second variation contains manipulations from groups with intermediate detection ratio, with an average detection ratio of 42.81%; while the third variation includes manipulations from the lowest detection ratio (i.e., those with detection under 40%, see Table 5), with an average detection ratio of 31.02%.

This proof of concept experiment procedure is as follows: Each participant is presented the three variations of the scene in a randomized

[3]The scene is based on *Apartment Kit* by Brick Project Studio, available in the Unity Asset Store.

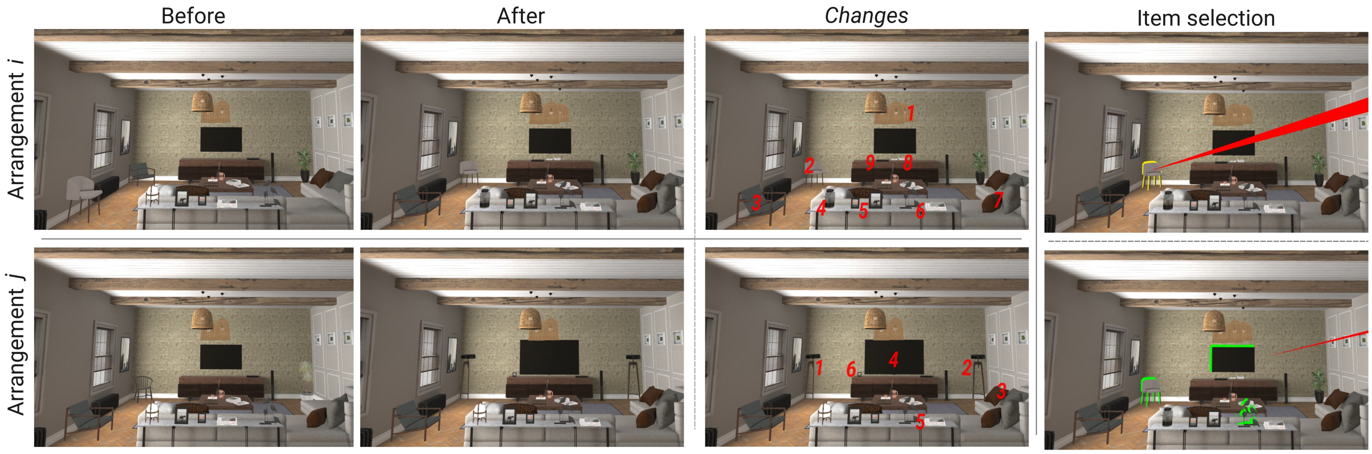

Fig. 8: (*Left*) Each row depicts an initial scene arrangement and the modified scene with the implemented changes following the flickering paradigm (see Section 5). Some elements of the scene were randomly rearranged before each trial, so that every trial was different. (*Right*) Participants select all the elements that they think that have changed with the controller. When pointing at an element, it gets outlined in yellow (*top*), while selecting it with the trigger gets it outlined in green until the end of the trial (*bottom*).

order. For each of the three variations, they can explore the scene for an unlimited amount of time, until they feel they already explored everything they want. Then, they have to press the controller trigger, a grey flicker is presented (following the traditional flicker paradigm), and the changes are applied. After the flicker ends, the participants have to indicate which elements have changed on the scene, by means of pointing with the controller. Once they have marked all the desired elements, they can move to the next scene variation.

Within this paradigm, we would expect that participants detect more changes from the scene variation that uses manipulations from the highest detection ratio set; while their ability to detect the changes would decrease for the variations with manipulations from the lowest detection ratio sets. We thus expect that scenes with a higher score contain manipulations more easy to be detected.

We ran this validation test with five participants, all naïve to the final goal of the experiment. Our results (see Table 6) indeed follow this trend, showing how our insights seem to be also valid within a different scene, while suggesting that a generalizable change blindness metric could be achievable, provided a large enough dataset.

5 Visual Working Memory Capacity

A plethora of works suggest that change blindness is indeed a question of memory (see Section 2), and our previous results (Section 4.2) suggest similar insights. Previous literature has investigated how visual stimuli is stored in our brain, and whether there is some limit in our capacity. They have found that memory tends to degrade when the amount of information to retain increases [12]. Some works suggest that capacity varies substantially depending on the number of stimuli, and ranges from two to four elements [2, 14]. Such limit would thus unavoidably facilitate change blindness. Therefore, while the main experiment (Section 3) is designed such that only *one* change takes place in each trial to ensure control over the different factors, we have devised an additional experiment to analyze whether the number of changes is a relevant parameter and can affect the ability to detect changes.

5.1 Experiment Details

Stimuli Participants viewed another realistic, stereoscopic virtual living room, based on the Living Room Interior by Blue Dot Studios (see Section 3.1). For this experiment, we arranged the furniture in the room in such a way that the participant would be located in one side of the room, close to the wall, and would have all the visual stimuli in front of them (i.e., no elements would be behind them). Participants sat on a rotating stool to prevent translational movements. During the experiment, participants have some seconds to explore the room and memorize it, and then some changes are applied (see Figure 5, left) following the flickering paradigm (see Section 3.1). Participants have to select all the elements where some change has been applied.

To avoid habituation effects, we rearrange some elements on the scene before each trial, so that every arrangement of the scene is different to the previous one. In this experiment, we implemented seventeen different manipulations (see Figure 8 for some examples). Each of this manipulations has a probability of occurring, and thus changes in each trial are selected randomly. In our experiment, we set this probability to $1/3$, so that, on average, each trial would have between five and six manipulations. These numbers have been suggested to be close to the maximum capacity [2, 14]. We let the user visually explore the scene for ten to twelve seconds (randomly selected each trial), to memorize all the information they can.

Hardware We resorted to the same hardware setup as in the previous experiment (see Section 3.2).

Participants Thirteen participants (six identified themselves as female, seven identified themselves as male, and none of them identified as non-binary, other, or preferred not to say; average age 26.84 years old, $STD = 6.26$) voluntarily took part in the study and provided written consent for participation. They were naïve to the final purpose of the experiment, and they all reported normal or corrected-to-normal vision. Twelve of them had already used an HMD before, although with different frequencies (six occasionally, three often, and three very often). All the experimental procedure was approved by our local ethics committee (left anonymous for reviewing purposes).

Procedure Each participant completed twenty different trials. The participant was given between approximately 20 seconds [16] to explore the scene. After that, some spheres spawned to avoid participants directly look at a potentially changing element (inspired by inattentional blindness, see Section 2), followed by a grey flicker (as in the flicker paradigm [16, 17]) that lasts for 200 ms., which again is in the range of an average human blink duration. While the grey flicker was happening, some of our seventeen implemented changes were randomly chosen. Each manipulation had a probability of occurring of $1/3$, and thus, on average, each trial had approximately six effective changes. When the flicker ended, the participant had to select all the elements that had changed (i.e., all that they remembered that were different before the flicker). To do so, they used the controller to point any items in the room. When pointed, items got outlined in yellow (see Figure 8, top right). Pressing the trigger selected the pointed item, which got outlined in green (see Figure 8, bottom right). There was no time limit for participants to select those elements. Once the participant had selected all desired elements, they could confirm pressing the A/B

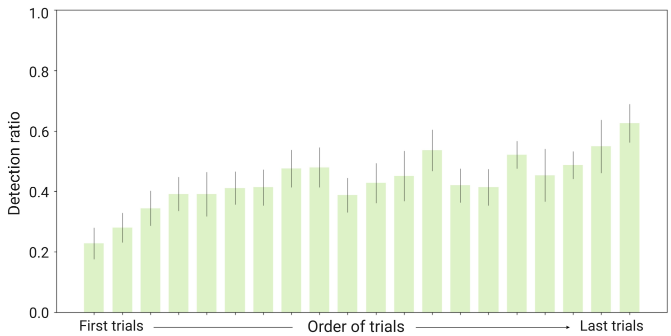

Fig. 9: Participants' detection ratio w.r.t. trial order. Our results suggest that participants are able to detect a larger number of changes over time, probably due to habituation effects. However, such capacity increases at a slow pace: In our experiment, participants were exposed to the same scene for twenty trials (i.e., approx. twenty minutes), and yet their detection ratio increased just up to 60%, suggesting that in several cases, change blindness is unavoidable. Please refer to Section 5.2 for further details.

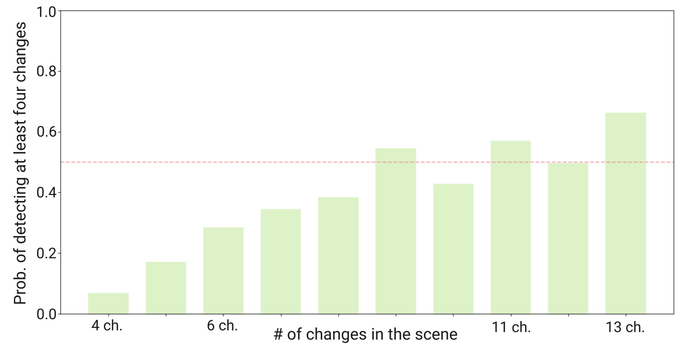

Fig. 10: Probability of participants detecting at least four changes ($d = 4$) depending on the amount of changes on the scene c. We selected four since previous literature suggests that having more than four changes causes memory's performance to drastically decrease [2, 14]. The red line represents chance level (probability of 50%. Note how the probability of detecting four changes remains below chance level even when a significant amount of changes are present in the scene. We refer the reader to Table 7 for additional details on different amount of detected changes d.

button. Then, the trial was completed, and they could move on to the next one by pressing the trigger button.

Additionally, we asked participants both before and after the experiment to fill a sickness questionnaire, to measure whether any sickness symptoms had arisen throughout the experiment. Additional information on this can be found in the supplementary material.

5.2 Results and Discussion

Following the experimental procedure introduced in the previous section, we obtain the results of twenty different trials per participant. For each of those trials, we store the arrangement of the scene, the applied manipulations, and the items that the participant selected as manipulated. For a total of thirteen participants, we ended up with 260 different trials. On average, each trial had 6.33 manipulations on the scene, ranging from 2 to 13 and following a normal distribution (Shapiro-Wilk test, $p-value < 0.001$).

Our results show that, on average, participants selected 3.38 ($STD = 1.89$) manipulations. Within those, only a 79.29% of the selected ones were elements that had actually been manipulated, while the remaining were elements in which no change had been applied (i.e., they incorrectly selected them). Besides, 3.68 ($STD = 2.19$) manipulations remained unnoticed on average, representing 58.14% of the total average manipulations. An average capacity of detecting 2.65 changes is indeed on par with the state of the art on visual working memory on traditional media, which reports such capacity to be between two and four elements [16]. Participants took on average 24.28 ($STD = 10.20$) seconds to select the manipulations.

We have analyzed whether the number of changes presented affected participants' ability to detect changes. However, when more changes are conducted in a scene, the probability of participants randomly selecting a correct change increase. To take into account such dependency, we have resorted to Bayesian inference to compute the corrected probabilities $P(d|c)$ of participants detecting at least $d = \{1,..,c\}$ changes in a scene where $c \in [2,10]$ changes took place. Particularly, we compute:

$$P(d|c) = \frac{P(c|d) \cdot P(d)}{P(c)},\ d = \{1,..,c\},\ c \in [2,10], \tag{1}$$

where $P(c|d)$ represents the probability of a scene having c changes given that the number of detected changes was at least d, $P(d)$ is the probability of detecting at least d changes in any scene, and $P(c)$ is the probability of the scene having c changes, as given by a normal distribution (see previous subsection). Results of such computation can be seen in Table 7.

We found that the probability of detecting at least one change rapidly increased w.r.t. the number of changes in the scene, but the probability of detecting more changes (i.e., $d >= 2$) increased in a much slower fashion, reaching a point where even with large numbers of changes c in a scene, the probability of detecting $d >= 4$ changes was close to chance level (we specifically depict the case of $d >= 4$ in Figure 10 since previous literature [2, 14] suggests that having more than four changes drastically affects memory's performance, and refer the reader to Table 7 for other values of d.) Besides, we found that the probability of detecting all the changes in a scene was very small, and decreased at such a rapid pace that when more than four changes were presented in the scene, it is virtually impossible that observers are able to detect them all.

We have also studied whether participants' ability to detect changes improves over time, caused by some habituation effect. We analyzed detection ration w.r.t. to trial order, and found a positive correlation ($r = 0.82$, $p < 1e-6$ for Pearson's correlation) between both factors, with participants detecting less than 40% of the changes in the first trials, and up to around a 60% in the last ones (see Figure 9). We also conducted a generalized linear mixed model (GLMM) to assess such dependency, and found trial order to have a significant effect on detection ratio ($tStat = 5.82, p < 1e-7$). Note that the ratio of missed changes (i.e., changes that happened but participants did not select) is complementary to the previous one ($detected + missed = total$). As such, all statistical operations yield analogous results (i.e., same Pearson's coefficient and p-value but with a negative correlation, and identical p-value and statistic for the GLMM).

6 Conclusion

In this work, we have performed a systematic, quantitative analysis of the phenomenon of change blindness, varying parameters such as type, distance, complexity, field of view, and number of manipulations. Different from previous works, we have performed our experiments using virtual reality environments, where the viewing conditions differ significantly from 2D images shown on traditional displays.

Some of our results are in agreement with previously reported results using 2D images; for instance, the limitations of our visual memory seem to impose a threshold in the amount of changes we can detect. Although performance increases as observers are exposed to the same stimuli for longer periods of time (which we hypothesize is due to an habituation effect that allows a more robust mental representation of the scene) we found that this limit is rarely surpassed, and seems to be independent of the viewing conditions and the amount of actual changes performed in the scene. Other results differ from previous experiments reported in the literature: for instance, we found that the type of manipulation does not have a significant effect on its detectability,

Table 7: Probability of detecting at least d changes in a scene where a total of c changes have been implemented, computed as stated in Section 5.2. Boldface values are $d = c$ (i.e., detecting all changes), while underlined values are below chance level (50%). Note how detecting a few changes gets easier when sufficient number of changes are implemented in a scene, but detecting several changes ($d >= 4$) gets hard regardless of the amount of changes in the scene c.

Det. ch. d >=	**Number of changes c in the scene**								
	2	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
2	**0.11**	0.36	0.59	0.66	0.81	0.69	0.92	0.91	0.86
3	-	**<1e-4**	0.34	0.41	0.57	0.65	0.62	0.73	0.64
4	-	-	**0.07**	0.17	0.29	0.35	0.38	0.55	0.43
5	-	-	-	**0.07**	0.02	0.23	0.26	0.36	0.36
6	-	-	-	-	**<1e-4**	0.04	0.10	0.23	0.21
7	-	-	-	-	-	**<1e-4**	0.05	0.05	<1e-4
8	-	-	-	-	-	-	**0.02**	<1e-4	<1e-4
9	-	-	-	-	-	-	-	**<1e-4**	<1e-4
10	-	-	-	-	-	-	-	-	**<1e-4**
c	**0.11**	**<1e-4**	**0.06**	**0.07**	**<1e-4**	**<1e-4**	**0.02**	**<1e-4**	**<1e-4**

whereas previous work on 2D images assumes such effect indeed exists. In those works, changes are analyzed from an image editing perspective, where the importance of the change depends on the absolute variation between the original and the manipulated image, rather than on the change itself [25]. In our 3D environments higher level factors such as context, semantics, and similarity to real environments come into play, while the participant can leverage other additional cues (such as motion parallax or depth perception) not present in 2D images. As a result, direct comparison between change blindness in 2D images and 3D environments is not trivial.

We believe that our work is a timely effort towards designing compelling experiences in VR that can be leveraged in VR applications such as redirected walking, games, or studies on saliency and gaze prediction.

6.1 Limitations and future work

The exact mechanisms that trigger the change blindness effect are still not completely understood; in that regard, there is still plenty of space for exciting future work on the topic of change blindness. Studying lower-level features such as color or contrast, or additional properties such as semantics or context would help improve our knowledge of this effect. Similarly, designing and evaluating a larger set of manipulations would generate even more robust results, eventually leading to a potential dataset useful for further studies.

All the experiments conducted throughout this work use indoor scenes. While our results suggest that our insights are robust, outdoor layouts may lead to different outcomes. Generally, such environments are much larger than indoor scenes, and may contain less structure but more elements. This would lead to a less precise visual working memory representation (as explained in Section 5), and thus detection and memory performance would be likely to decrease. Further investigation would be needed to understand the dependency of change blindness with the type of environment.

Analyzing this effect when observers are under higher cognitive loads (i.e., when performing a particular task) also remains to be explored. For instance, Suma et al. [40] found that manipulations of a 3D scene were unnoticed for 76 out of 77 participants engaged in a walking task, suggesting that such confounding factors play an important role. In this regard, our results are conservative, since users were informed that changes were going to happen; this facilitates generalization to a larger scope of applications. Last, another interesting avenue of future work is exploring the interactions with multimodal effects [27]. For instance, it is known that unexpected sounds may elicit temporary blindness [26], but its potential relation with change blindness has not been thoroughly explored.

Exploring potential applications of this phenomenon also merits further research. Techniques like redirected walking could potentially improve by eliciting a desired trajectory by manipulating some elements of a scene without the user noticing, similar to the work by Suma et al. [40]. Indeed, exploring how much redirection can be applied depending on the manipulation remains an open problem. Other applications like education or training could also benefit by adjusting the cognitive load to a level that visual working memory is able to maintain. Finally, scene design for narrative experiences or videogames could also take into account this effect, helping prevent particular details from being missed by the observer.

Acknowledgments

This work has received funding from the European Research Council (ERC) under the EU's Horizon 2020 research and innovation programme (project CHAMELEON, Grant no. 682080). This project has also received funding from the PRIME project (MSCA-ITN, grant agreement No. 956585), and from the Spain's Agencia Estatal de Investigación (project PID2019-105004GB-I00). Additionally, Daniel Martin was supported by a Gobierno de Aragon (2020-2024) predoctoral grant.

References

[1] R. Alghofaili, M. S. Solah, H. Huang, Y. Sawahata, M. Pomplun, and L.-F. Yu. Optimizing visual element placement via visual attention analysis. In *IEEE Conference on Virtual Reality and 3D User Interfaces (IEEE VR)*, pp. 464–473. IEEE, 2019. 2

[2] G. A. Alvarez and P. Cavanagh. The capacity of visual short-term memory is set both by visual information load and by number of objects. *Psychological Science*, 15(2):106–111, 2004. 2, 7, 8

[3] J. E. Attwood, C. Kennard, J. Harris, G. Humphreys, and C. A. Antoniades. A comparison of change blindness in real-world and on-screen viewing of museum artefacts. *Frontiers in Psychology*, 9:151, 2018. 2

[4] A. D. Baddeley and G. Hitch. Working memory. In *Psychology of Learning and Motivation*, vol. 8, pp. 47–89. Elsevier, 1974. 2

[5] P. M. Bays, R. F. Catalao, and M. Husain. The precision of visual working memory is set by allocation of a shared resource. *Journal of Vision*, 9(10):7–7, 2009. 1, 2

[6] K. Bergmann, A.-L. Schubert, D. Hagemann, and A. Schankin. Age-related differences in the p3 amplitude in change blindness. *Psychological Research*, 80(4):660–676, 2016. 1

[7] E. Bernal-Berdun, D. Martin, D. Gutierrez, and B. Masia. SST-Sal: A spherical spatio-temporal approach for saliency prediction in 360° videos. *Computers & Graphics*, 106:200–209, 2022. 2

[8] Bionumbers. Average duration of a single eye blink. https://bionumbers.hms.harvard.edu/bionumber.aspx?s=y&id=100706. Accessed: 2022-07-19. 3

[9] T. F. Brady, T. Konkle, and G. A. Alvarez. A review of visual memory capacity: Beyond individual items and toward structured representations. *Journal of Vision (JoV)*, 11(5):4–4, 2011. 1, 2

[10] J. Castillo Escamilla, J. J. Fernández Castro, S. Baliyan, J. J. Ortells-Pareja, J. J. Ortells Rodríguez, and J. M. Cimadevilla. Allocentric spatial memory performance in a virtual reality-based task is conditioned by visuospatial working memory capacity. *Brain Sciences*, 10(8):552, 2020. 1

[11] E. R. Chrastil and W. H. Warren. Active and passive contributions to spatial learning. *Psychonomic Bulletin & Review*, 19(1):1–23, 2012. 1

[12] N. Cowan. The magical number 4 in short-term memory: A reconsideration of mental storage capacity. *Behavioral and Brain Sciences*, 24(1):87–114, 2001. 2, 6, 7

[13] I. Fatt and B. A. Weissman. *Physiology of the eye: an introduction to the vegetative functions*. Butterworth-Heinemann, 2013. 3

[14] G. S. Halford, N. Cowan, and G. Andrews. Separating cognitive capacity from knowledge: A new hypothesis. *Trends in Cognitive Sciences*, 11(6):236–242, 2007. 2, 7, 8

[15] M. M. Hayhoe and C. A. Rothkopf. Vision in the natural world. *Wiley Interdisciplinary Reviews: Cognitive Science*, 2(2):158–166, 2011. 1

[16] A. Hollingworth. Scene and position specificity in visual memory for objects. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 32(1):58, 2006. 1, 2, 3, 6, 7, 8

[17] A. Hollingworth. Object-position binding in visual memory for natural scenes and object arrays. *Journal of Experimental Psychology: Human Perception and Performance*, 33(1):31, 2007. 1, 2, 3, 6, 7

[18] A. Hollingworth and J. M. Henderson. Accurate visual memory for previously attended objects in natural scenes. *Journal of Experimental Psychology: Human Perception and Performance*, 28(1):113, 2002. 2

[19] R. S. Kennedy, N. E. Lane, K. S. Berbaum, and M. G. Lilienthal. Simulator sickness questionnaire: An enhanced method for quantifying simulator

sickness. *The International Journal of Aviation Psychology*, 3(3):203–220, 1993. 4

[20] D. Kim, J. Kim, J.-e. Shin, B. Yoon, J. Lee, and W. Woo. Effects of virtual room size and objects on relative translation gain thresholds in redirected walking. *ArXiV (Preprint)*, 2022. 4

[21] E. Langbehn, F. Steinicke, M. Lappe, G. F. Welch, and G. Bruder. In the blink of an eye: leveraging blink-induced suppression for imperceptible position and orientation redirection in virtual reality. *ACM Trans. on Graphics*, 37(4):1–11, 2018. 2, 3

[22] C.-L. Li, M. P. Aivar, D. M. Kit, M. H. Tong, and M. M. Hayhoe. Memory and visual search in naturalistic 2d and 3d environments. *Journal of Vision*, 16(8):9–9, 2016. 1

[23] A. L. Lohse, C. K. Kjær, E. Hamulic, I. G. Lima, T. H. Jensen, L. E. Bruni, and N. C. Nilsson. Leveraging change blindness for haptic remapping in virtual environments. In *IEEE Workshop on Everyday Virtual Reality (WEVR)*, pp. 1–5, 2019. 2

[24] S. J. Luck and E. K. Vogel. The capacity of visual working memory for features and conjunctions. *Nature*, 390(6657):279–281, 1997. 2

[25] L.-Q. Ma, K. Xu, T.-T. Wong, B.-Y. Jiang, and S.-M. Hu. Change blindness images. *IEEE Trans. on Visualization and Computer Graphics*, 19(11):1808–1819, 2013. 1, 2, 3, 6, 9

[26] S. Malpica, A. Serrano, D. Gutierrez, and B. Masia. Auditory stimuli degrade visual performance in virtual reality. *Scientific Reports*, 10(1):1–9, 2020. 9

[27] D. Martin, S. Malpica, D. Gutierrez, B. Masia, and A. Serrano. Multimodality in VR: A survey. *ACM Computing Surveys (CSUR)*, 54(10s):1–36, 2022. 9

[28] D. Martin, A. Serrano, A. W. Bergman, G. Wetzstein, and B. Masia. ScanGAN360: A Generative Model of Realistic Scanpaths for 360º Images. *IEEE Trans. on Visualization and Computer Graphics*, 28(5):2003–2013, 2022. 2

[29] V. Mazza, M. Turatto, and C. Umilta. Foreground–background segmentation and attention: A change blindness study. *Psychological Research*, 69(3):201–210, 2005. 2

[30] R. A. Moses. Adler's physiology of the eye. *Clinical Application*, 1975. 3

[31] Y. Pertzov, M. Y. Dong, M.-C. Peich, and M. Husain. Forgetting what was where: The fragility of object-location binding. *PLoS One*, 7(10):e48214, 2012. 1, 2

[32] R. A. Rensink, J. K. O'regan, and J. J. Clark. To see or not to see: The need for attention to perceive changes in scenes. *Psychological Science*, 8(5):368–373, 1997. 2

[33] A. Serrano, D. Martin, D. Gutierrez, K. Myszkowski, and B. Masia. Imperceptible manipulation of lateral camera motion for improved virtual reality applications. *ACM Trans. on Graphics (Proc. SIGGRAPH Asia)*, 39(6), 2020. 2

[34] A. Serrano, V. Sitzmann, J. Ruiz-Borau, G. Wetzstein, D. Gutierrez, and B. Masia. Movie Editing and Cognitive Event Segmentation in Virtual Reality Video. *ACM Trans. on Graphics (Proc. SIGGRAPH)*, 36(4), 2017. 2

[35] D. J. Simons and D. T. Levin. Change blindness. *Trends in Cognitive Sciences*, 1(7):261–267, 1997. 1, 2

[36] D. J. Simons and R. A. Rensink. Change blindness: Past, present, and future. *Trends in Cognitive Sciences*, 9(1):16–20, 2005. 1, 2

[37] V. Sitzmann, A. Serrano, A. Pavel, M. Agrawala, D. Gutierrez, B. Masia, and G. Wetzstein. Saliency in VR: How do people explore virtual environments? *IEEE Trans. on Visualization and Computer Graphics*, 36(4), 2018. 4

[38] D. Smith and T. Schenk. Reflexive attention attenuates change blindness (but only briefly). *Perception & Psychophysics*, 70(3):489–495, 2008. 1

[39] F. Steinicke, G. Bruder, K. Hinrichs, and P. Willemsen. Change blindness phenomena for virtual reality display systems. *IEEE Trans. on Visualization and Computer Graphics*, 17(9):1223–1233, 2011. 2

[40] E. A. Suma, S. Clark, D. Krum, S. Finkelstein, M. Bolas, and Z. Warte. Leveraging change blindness for redirection in virtual environments. In *IEEE Conference on Virtual Reality and 3D User Interfaces (IEEE VR)*, pp. 159–166. IEEE, 2011. 2, 9

[41] Q. Sun, A. Patney, L.-Y. Wei, O. Shapira, J. Lu, P. Asente, S. Zhu, M. McGuire, D. Luebke, and A. Kaufman. Towards virtual reality infinite walking: dynamic saccadic redirection. *ACM Trans. on Graphics*, 37(4):1–13, 2018. 2

[42] Q. Sun, L.-Y. Wei, and A. Kaufman. Mapping virtual and physical reality. *ACM Trans. on Graphics*, 35(4):1–12, 2016. 2

[43] H. Taschenbuch Verlag Schiffman. Sensation and perception: An integrated approach, 2001. 3

[44] M. Turatto, A. Angrilli, V. Mazza, C. Umilta, and J. Driver. Looking without seeing the background change: Electrophysiological correlates of change detection versus change blindness. *Cognition*, 84(1):B1–B10, 2002. 2

[45] L. G. Ungerleider, S. M. Courtney, and J. V. Haxby. A neural system for human visual working memory. *Proceedings of the National Academy of Sciences*, 95(3):883–890, 1998. 1

[46] M. Vasser, M. Kängsepp, and J. Aru. Change blindness in 3d virtual reality. *ArXiV (Preprint)*, 2015. 2, 3, 4, 6

[47] J. O. Wobbrock, L. Findlater, D. Gergle, and J. J. Higgins. The aligned rank transform for nonparametric factorial analyses using only anova procedures. In *Proc. of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '11, p. 143–146, 2011. 4, 5

[48] Y. Xu, Z. Zhang, and S. Gao. Spherical dnns and their applications in 360° images and videos. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 2021. 2