

**Attentional Cueing Experiments on the Relation Between Scene Change Blindness and
Visual Memory Representations**

Jason Helbing

Johann Wolfgang Goethe-Universität Frankfurt am Main

Report for Empirisch-experimentelles Praktikum (PsyBSc8), Summer term 2017

Instructor: Dejan Draschkow, PhD

E-mail: jason.helbing@stud.uni-frankfurt.de

Contents

Abstract	3
1 Introduction	4
2 Method	8
2.1 Participants	8
2.2 Stimulus material	8
2.3 Apparatus	8
2.4 Procedure	9
2.5 Data analysis	11
3 Results	12
4 Discussion	15
5 References	20
Eigenständigkeitserklärung	24

Abstract

Research on visual memory representations has brought up to two seemingly opposite findings: Visual long-term memory can store a multitude of objects and scenes with a high degree of detail, yet people tend to miss rather significant changes to scenes when exposure is disrupted for just a brief moment. It has been suggested that the latter phenomenon, known as change blindness, may be the result of attentional deficits rather than an indicator of generally impoverished memory representations. To further our understanding of the relation between scene change blindness, attention, and visual memory representations, we conducted two experiments in which we simultaneously (Experiment 1) and sequentially (Experiment 2) cued a varying number of objects in images of naturalistic scenes before subjects had to indicate whether there had been a change after a one-second delay. While an observed decrease in accuracy as a function of the number of cues in both experiments indicates that limitations to the capacity of visual working memory can play a role in change detection, two other observations regarding attention and gaze stand out: First, simultaneously directing attention towards even a larger subset of objects containing the target led to above chance hit rates in change detection which supports the idea that attention plays a key role when we perceive changes. Second, hit rates were higher when people could scan the potential targets simultaneously compared to sequential exposure, indicating that natural gaze behavior boosts scene memory.

Keywords: change blindness, visual memory, memory fidelity, change detection, spatial cueing, scene perception, gaze direction, overt attention

Introduction

The importance of visual memory for human cognition cannot be overestimated: Without the ability to recollect objects and scenes previously perceived we could not think about our past experiences or reason about possible future events with the high degree of abstraction we are used to. Understanding the mental nature of visual memory representations has therefore been an objective of the cognitive sciences that spawned an abundance of research. In line with memory research in general, the distinction between long-term and working memory as major components of memory is applied when describing how we remember visual percepts: While long-term memory passively stores information permanently over longer periods of time, all content of working memory is available for active processing and decays quickly if it is not rehearsed (Atkinson & Shiffrin, 1968). Research on visual long-term memory (VLTm) and visual working memory (VWM) strives to explain how these systems perform with visual stimuli on both a quantitative and qualitative level: How much of what we see is stored for just a moment, which parts are long-lasting, how detailed are those representations, and how do we use them?

A classic finding about the capacity of VLTm is that its storage for distinct images is rather massive: Standing (1973) presented 10,000 pictures to participants for just a few seconds each and then tested their memory by making them decide which of two images they had seen before—a retrieval method known as two-alternative forced choice (2AFC). He found that people chose the correct image on 83% of trials and concluded that “the capacity of recognition memory for pictures is almost limitless” (Standing, 1973, p. 207). These early results, however, could not yet address the question of how detailed visual representations of the memorized images were: Remembering the very gist of a presented image without any visual detail would have been sufficient to differentiate it from the presented foil in the 2AFC test as these foils were not chosen in such a way as to resemble the images they were

presented with in any way. To estimate the degree of detail of object memory representations, Brady, Konkle, Alvarez, and Oliva (2008) had participants look at 2,000 images of isolated objects on a white background for 3 s each and then tested performance in a 2AFC task in which foils systematically differed in their similarity to the memorized objects: Some of them were entirely different objects (novel condition), others were different exemplars of the same object, and, lastly, some were the same object in a different state. They found performance to be remarkably high in all three conditions: The 93% accuracy rate in the novel condition can be considered a replication of Standing's (1973) earlier finding, whereas 88% and 87% in the exemplar and state conditions, respectively, clearly demonstrate that a considerable amount of object details can be stored in VLTm. A similar paradigm using varying amounts of scenes from the same categories in a set of approximately 3000 images for memorization demonstrated that this surprisingly high fidelity is not limited to object representations but also holds for scene memory: Even when presented with 64 scenes from the same category (e.g., bathrooms, offices, gyms, golf courses), participants picked the correct image on 76% of trials in a subsequent 2AFC test (Konkle, Brady, Alvarez, & Oliva, 2010). This high performance was observed even though scenes were not shown for a longer amount of time than isolated objects in the study by Brady et al. (2008).

Another corpus of research in the domain of visual cognition has studied our performance in detecting changes to all sorts of stimuli, which is something, it turns out, we are stunningly bad at: a phenomenon known as change blindness (Simons & Levin, 1997). One prominent way of illustrating this is the so-called flicker paradigm, developed by Rensink, O'Regan, and Clark (1997): An image and a modified version of it are presented alternately for a short moment with a blank of usually about 80–150 ms between the two images until the observer detects the difference between them. This usually requires several switches between the images, even for rather significant alterations; for changes to details it

can take a couple of dozen switches (Rensink et al., 1997). This difficulty in detecting changes, however, is not limited to static images: Continuity errors in video material (i.e., changes to scenes in motion pictures between cuts) are also easily missed (Levin & Simons, 1997) and even swapping the interlocutor in a real-world conversation goes unnoticed way more often than intuition would predict (Simons & Levin, 1998). Change blindness has naturally been taken as evidence of low-fidelity visual memory representations (Simons & Levin, 1997; Wolfe, 1998)

The outlined findings undeniably seem contrary to one another: How can we have a memory system capable of storing details of a massive number of stimuli for very long durations and at the same time miss changes to scenes after just a fraction of a second of not being directly exposed to them? Previous research dealing with this question suggests that classic change blindness tasks like the flicker paradigm using very short display times between the unchanged and manipulated image might not be solid support of the idea that we cannot store detailed representations of scenes in visual memory: They might rather just indicate that serially attending to details in scenes is essential for their encoding into these representations (for a review discussing the legitimacy of conclusions drawn from change blindness experiments, see Simons & Rensink, 2005). In fact, Hollingworth and Henderson (2002) reported remarkable improvements in change detection performance when they used eye tracking to ensure that the target object had been fixated before the change occurred. Also, Brady, Konkle, Oliva, and Alvarez (2009) conducted a change detection experiment with arrays of six real-world objects from their previous memory study (Brady et al., 2008) in which they altered the display duration and found that, given sufficient time to encode the objects in the first place, many more changes, even such ones which just affected details of objects (i.e., state changes), were detected.

The aim of our experiments was to test change detection performance for naturalistic scenes under varying VWM load while controlling for the location of the change being paid attention to before the change is to occur. This enabled us to assess how strongly memory limitations directly influence the ability to notice changes to real-world scenes: The possibility that the change went unnoticed because it had not been fixated before was eliminated—or at least markedly less likely—as red frames were put around either four, seven, or ten objects and participants were informed that changes could only occur in these cued areas. In Experiment 1, all cues were visible throughout the entire study phase enabling subjects to freely choose a gaze pattern for scanning them. This resembled natural viewing behavior with the restriction that participants knew which objects were relevant. To explore the effects of even stronger gaze direction, we conducted Experiment 2 in which we extended control over how the cues had to be attended by having them appear sequentially across the entire study phase. This meant, participants were less free in choosing where to look since they had to follow the given order of cues as they would have otherwise likely missed the appearance of new cues, at least in the conditions with seven and ten frames. The cue numbers were chosen for laying within (i.e., four), at the boundary of (seven), and above (ten) what we would expect to fit into VWM storage based on previous research (Luck & Vogel, 1997).

In accordance with common sense and an intuitive understanding of cognitive load, we expected to find a decrease in change detection performance with an increasing number of cues. We were, however, less certain about the exact dimension of this drop: Neither an extremely shallow nor a boundless decline seemed plausible as we did not expect change detection to be either independent or solely conditional on whether the target object was attended to before the change. Our experiments should therefore be characterized as rather exploratory investigations.

Method

Participants

Subjects were recruited at Goethe University Frankfurt am Main, on social media, and via word-of-mouth referrals. The sample included 18 participants in Experiment 1 ($M_{\text{age}} = 24.7$, age range: 18–46, 15 female and 3 male) and 17 participants in Experiment 2 ($M_{\text{age}} = 24.6$, age range: 19–48, 12 female and 5 male). All subjects volunteered, three were acquaintances of the experimenters (two of them in Experiment 1); the rest were psychology students compensated with participation credit. Subjects passed eyesight examinations for normal or corrected-to-normal visual acuity (at least 20/20 vision) and had normal color vision as assessed by the Ishihara test.

Stimulus material

Images were taken from the Change Blindness Database maintained by the Visual Attention Lab at Harvard University: Eighty-four scenes from the main CB database (Sareen, Ehinger, & Wolfe, 2016) and 36 scenes from the additional CB images (Ehinger, Allen, & Wolfe, 2016), both accessible at <http://search.bwh.harvard.edu/new/CBDatabase.html>, were used. All changes were object deletions, that is, an object from the unchanged version (hereinafter referred to as the target object) is no longer present in the changed one. For example, a soap dispenser in a bathroom or a loaf of bread in a kitchen may disappear. Eighty-six of the images (71.7%) displayed indoor scenes (predominantly living and bedrooms, kitchens and bathrooms); the rest were outdoor scenes (e.g., gardens, parking lots, waterfronts).

Apparatus

Both experiments were conducted in unlit rooms in front of 19-in. screens with a refresh rate of 60 Hz and a resolution of 1680×1050 px. Images were displayed in the center of the screen at a resolution of 1024×768 px against a black background. Image display and

response measurement was controlled by OpenSesame (version 3.0.7; Mathôt, Schreij, & Theeuwes, 2012) on computers running the Windows 7 operating system.

Procedure

Each trial began with a fixation dot (1 s), followed by the unchanged image, that is, the scene still containing the target object. It was first displayed without any cues for 250 ms in both experiments. We included this preview to ensure that the initial processing of the scene was natural and not to any extent obstructed by the presence of the cueing frames that would be visible later: Substantial effects of very early scene processing on guidance of eye-movements in subsequent search tasks were demonstrated (Castelhano & Henderson, 2007; Võ & Henderson, 2010) and we had to assume that such modulation by the initial percept might occur in our paradigm as well. In Experiment 1, this preview was followed by a set of either four, seven, or ten red frames (stroke width of three points) simultaneously appearing around individual objects or groups of objects clearly forming a unit (e.g., a pile of books or a salt shaker and a pepper mill next to each other). In Experiment 2, these cues appeared sequentially, that is, one after another until four, seven, or ten frames were visible. The set of objects cued always contained the target. The overall length of this study phase was 7 s in both experiments with cueing frames appearing at equal time intervals in Experiment 2, that is, one additional frame every 1.75, 1, and 0.7 s for four, seven, and ten cues, respectively. The cueing frames then disappeared in both experiments while the unchanged scene remained visible for another second to avoid afterimages of the very salient red frames during the subsequent delay, which consisted of 1 s of gray screen. After this blank, either the changed or once again the unchanged image appeared, and participants had to indicate whether there had been a change to the scene by pressing buttons on a keyboard (Y and N keys for “yes” and “no”; therefore hereinafter referred to as the YN task). Response times of this choice were measured. If subjects stated they had noticed a change, they additionally had

to specify its location with the click of a mouse. Before the start of the next trial, a 1 s alert told participants to get ready and reminded them which keys to use for their response. A sample image pair and how it would appear under the different cueing conditions is illustrated in Figure 1, as well as the procedures of individual trials of both experiments.

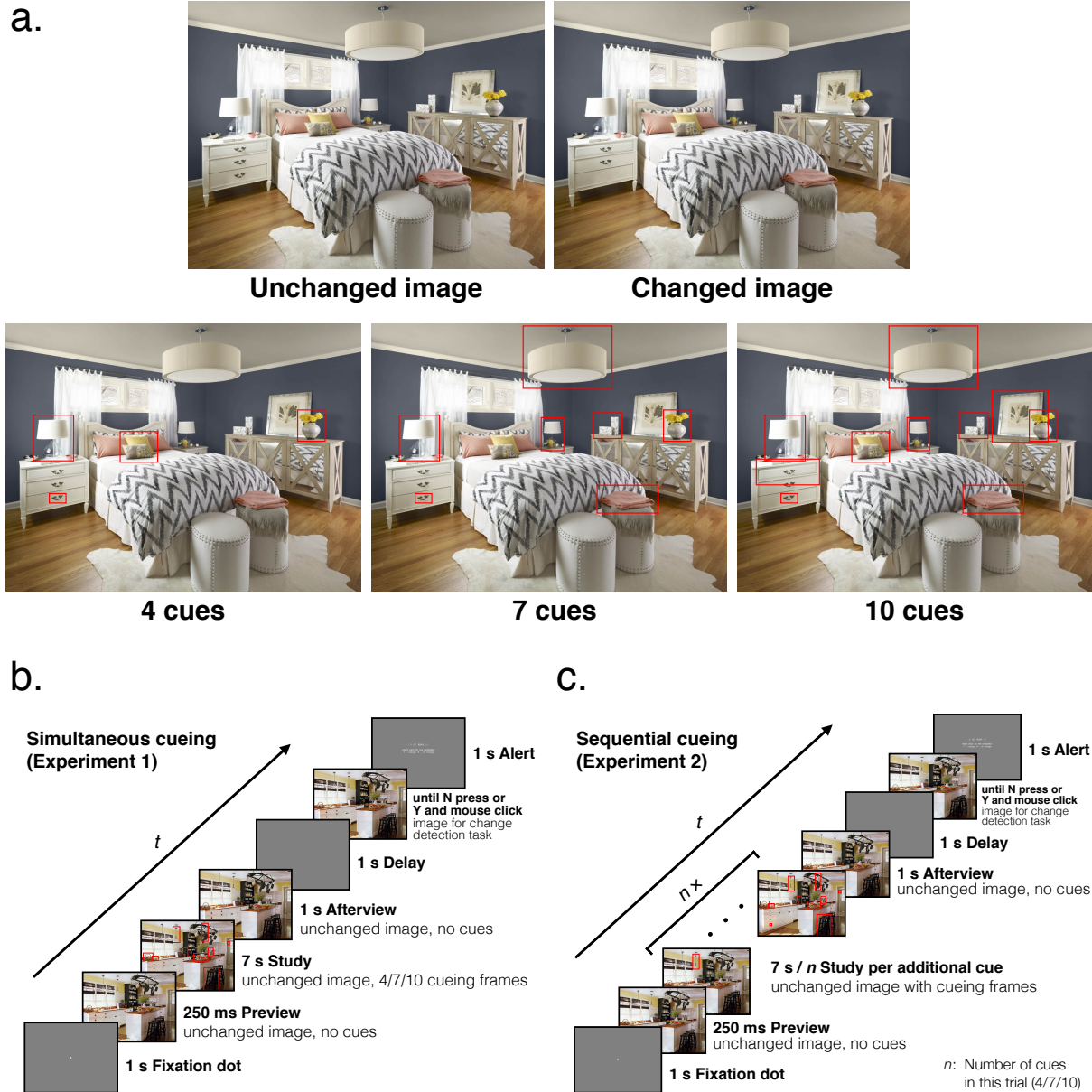


Figure 1. Sample scene depicted in all configurations of the experiment (a) and trial procedures of Experiment 1 (b) and 2 (c). Note that the relative size of the image on the screen was enlarged for display purposes in the procedures' diagrams.

There were 120 trials in each experiment using the same set of distinct scenes and targets in both experiments and for each participant. The trials were equally divided into the

three cueing conditions, resulting in 40 trials per condition, in half of which the target object actually disappeared (change-present trials), whereas on the other half, there was no change (change-absent trials). The scenes' assignment to these different conditions was counterbalanced across participants. The locations of the target objects were furthermore controlled for equal distribution across the left and right half of the images. In some images, cueing frames for objects that were connected or close to one another overlapped; it was, however, ensured that the frames did not obscure the objects in a way that would hide visual information central to the object's identification and memorization.

Participants in both experiments were instructed to carry out the YN and mouse tasks as described above. They were informed that changes to the scenes could only occur within the areas previously cued. Participants did not receive feedback on any of their responses, neither during nor after the experiments.

Data Analysis

Mixed-design ANOVAs were used to analyze the effects of the number of cues (repeated measures; levels four, seven, and ten) and of cueing technique (between-subjects), that is, Experiment 1 (simultaneous) versus Experiment 2 (sequential), on overall accuracy in the YN task and on hit rate, that is, accuracy on change-present trials. This distinction was motivated by an observed effect of change presence when including it as an additional factor in the former ANOVA, thus questioning whether the overall accuracy is an appropriate dependent variable to assess change detection performance in this context. Specific conditions were compared with pairwise t tests afterwards, using alpha levels adjusted with the Holm–Bonferroni method (Holm, 1979). One sample t tests were used to compare hit change detection rates against chance level.

The analysis of our data was carried out in the R programming language (version 3.4.1; R Core Team, 2017) using RStudio (version 1.0.143; RStudio Team, 2015). The ez

package was used for ANOVAs (Lawrence, 2016). Plots were created with the ggplot2 package (Wickham, 2009) and additional functions from Chang (n.d.), that were used to aggregate means with their respective standard errors for graphic display.

Results

Analyses focus on the participants' performance in the YN task. When using overall accuracy as the dependent variable, a two-way mixed-design ANOVA revealed a significant effect of the number of cues, $F(2, 66) = 105.34, p < .001, \eta_p^2 = .53$. There was no significant effect of experiment and no interaction. On average, subjects in Experiment 1 responded correctly on 85.4% ($SE = 1.4$), 71.3% ($SE = 1.6$), and 65.3% ($SE = 2.0$), in Experiment 2 on 85.9% ($SE = 1.1$), 68.8% ($SE = 1.4$), and 61.8% ($SE = 1.8$) of trials with four, seven, and ten cues, respectively (see Figure 2a). Paired t tests comparing YN task accuracy in the three cue number conditions in Experiment 1, using Holm–Bonferroni corrected .05 alpha levels, found significant differences between the four and seven cue means, as well as the four and ten cue means, and no significant difference when comparing seven and ten cues. In Experiment 2 all of these differences were significant. Test statistics together with effect sizes of these comparisons can be found in the upper half of Table 1. Note that in each experiment a single Holm–Bonferroni correction process was used for six comparisons, that is, all three possible pairs of cueing conditions with means of overall accuracy as well as hit rates which will be presented later in this section. This was to account for the dependence of these two measures.

Taking change presence into account as an additional factor (three-way mixed-design ANOVA), revealed that it has a significant effect on accuracy, $F(1, 33) = 29.61, p < .001, \eta_p^2 = .27$. This distinction between experiment and cue condition as well as hit and correct rejection rate (change presence) is depicted in Figure 2b. There were, furthermore, significant interactions between change presence and the experiment, $F(1, 33) = 4.16, p = .0495, \eta_p^2 = .05$, as well as change presence and cue number, $F(2, 66) = 20.25, p < .001, \eta_p^2 = .10$. Both of

these interactions are apparent in Figure 2b: First, hit rates in Experiment 1 are notably above those of Experiment 2 whereas correct rejection rates of Experiment 1 are slightly below those of Experiment 2, and second, there is a distinct overall drop in hit rates compared to relatively stable correct rejection rates as a function of cue number. Unsurprisingly, the effect of cue number was also highly significant, $F(2, 66) = 105.94, p < .001, \eta_p^2 = .32$, whereas there was no significant main effect of experiment. No further interactions reached significance.

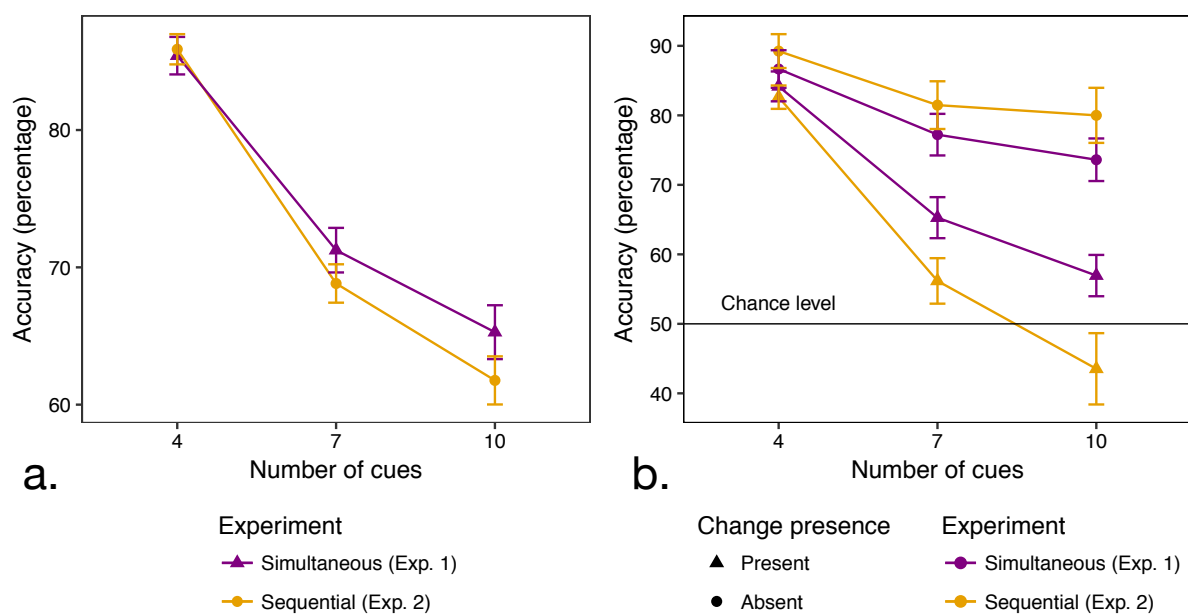


Figure 2. Overall accuracy in the YN task for each cue number (a) and hit and correct rejection rates (i.e., accuracy separated for change-present and change-absent trials) for each cue number (b).

Looking at target-present trial data only and using accuracy as the dependent variable, a two-way mixed-design ANOVA revealed significant effects of both the number of cues, $F(2, 66) = 99.71, p < .001, \eta_p^2 = .53$, and the experiment, $F(1, 33) = 4.86, p = .03, \eta_p^2 = .09$, yet no significant interaction between these two, $F(2, 66) = 3.13, p = .05, \eta_p^2 = .03$. Participants in Experiment 1 detected actual changes in 84.2% ($SE = 2.1$), 65.3% ($SE = 3.0$), and 56.9% ($SE = 3.0$) of cases with four, seven, and ten cues, respectively. In Experiment 2 those rates were at 82.6% ($SE = 1.7$), 56.2% ($SE = 3.3$), and 43.5% ($SE = 5.1$). When

comparing these means using paired t test and applying the above introduced Holm–Bonferroni correction, we found hit rates significantly differing between four and seven as well as four and ten cues in both Experiment, whereas the difference between seven and ten cues was only significant in Experiment 2 (see lower half of Table 1).

Table 1

Comparisons of means of YN task overall accuracy and hit rates (i.e., target-present trial accuracy) for all possible pairs of cue numbers

		Simultaneous Cueing (Exp. 1)					Sequential Cueing (Exp. 2)				
		$t(17)$	p	α	\tilde{p}	d	$t(16)$	p	α	\tilde{p}	d
Overall accuracy	4 vs. 7	7.65	<.001	.01	.003	1.80	11.58	<.001	.01	<.001	2.81
	7 vs. 10	2.20	.04	.03	.08	0.52	2.91	.01	.05	.01	0.71
	4 vs. 10	8.32	<.001	.01	.001	1.96	11.49	<.001	.01	<.001	2.79
Hit rate	4 vs. 7	7.16	<.001	.02	.005	1.69	9.69	<.001	.02	<.001	2.35
	7 vs. 10	2.29	.04	.03	.08	0.54	3.07	.01	.03	.02	0.75
	4 vs. 10	9.15	<.001	.01	<.001	2.16	9.69	<.001	.01	<.001	2.35

Note. \tilde{p} = Holm–Bonferroni corrected p values. α = Holm–Bonferroni corrected .05 alpha levels.

Using Welch's t test to compare hit rates between experiments revealed significant differences for seven, $t(30.86) = 2.09$, $p = .04$, $d = 0.71$, and ten cues, $t(27.03) = 2.30$, $p = .03$, $d = 0.79$; not for four cues, $t(32.73) = 0.47$, $p = .64$, $d = 0.16$.

One sample t tests showed that hit rates significantly differed from chance (i.e., 50%) in all conditions in Experiment 1, but only for four cues in Experiment 2 (see Table 2).

Table 2

Comparisons of change detection rates and .5 (chance) for different cue numbers

		Simultaneous Cueing (Exp. 1)			Sequential Cueing (Exp. 2)		
		$t(17)$	p	d	$t(16)$	p	d
4	13.85	<.001		3.26	14.93	<.001	3.62
7	5.67	<.001		1.36	1.81	.09	0.44
10	2.25	.04		0.53	-1.31	.21	0.32

Discussion

We conducted two change detection experiments with naturalistic scenes as stimuli in which we directed our participants' attention towards a varying number of objects before assessing their performance in detecting whether one of these had disappeared. Attentional cueing was achieved with salient frames around objects that appeared simultaneously in the first and sequentially in the second experiment. Unsurprisingly, we found highest change detection rates, which were above 80%, when cueing just four objects with hardly any difference between the experiments, suggesting that this amount of visual information could comfortably fit into VWM storage under the given circumstances. Increasing the number of cues to seven led to a clear performance drop in both experiments indicating difficulties to retain all information presented. Furthermore, at this point, a gap between the change detection rates of the two experiments became visible: Simultaneous cueing in Experiment 1 outperformed sequential cueing in Experiment 2, where those rates did not significantly differ from chance in this condition anymore. Between seven and ten cues, another, yet more moderate decline in change detection rates could be observed that only reached significance in Experiment 2 with average rates now below—yet not significantly differing from—chance level, whereas rates in Experiment 1 proved somewhat more stable, being still significantly above both chance and the rates of Experiment 2.

What does this tell us about the relation between change blindness and the fidelity of memory representations mentioned earlier? The only definite conclusion our results allow is that briefly attending to an object in a scene will not eradicate the possibility that a change to this object is missed after a brief visual disruption: Given a sufficiently rich scene and limited viewing time, even changes to objects that were attended can go unnoticed, most likely due to limitations of VWM. For a more precise assessment of how strongly directing attention towards a larger number of objects (seven and ten cues) might enhance change detection

performance, a baseline measure for the stimuli and display times used in our experiment would be required; that is, we would need to know how well changes are detected if the scenes are presented without any cues for the same amount of time they were in our experiments. To the best of our knowledge, no such data has been gathered by previous research for comparable stimulus material. This limits further inferences concerning the role of attention to a rather speculative level: With respect to Experiment 1, we would argue that (a) the rather moderate and statistically insignificant drop in change detection rates observed between seven and ten cues together with (b) the significant majority of changes (i.e., 57%) still detected when ten cues were displayed, indicate that directing attention towards a considerably large subset of objects had an advantageous effect on change detection performance for these objects. As participants were not aware that exactly half of trials were change-present, we also see no reason to assume that change detection rates could not theoretically have fallen well below chance level. Taking both this and the (later described) response bias towards reporting there was no change into account, it does not seem too far-fetched to consider change detection rates in Experiment 1 for seven and ten cues rather high and therefore argue that attending to even a larger number of different objects beneficially influenced change detection performance. This highlights the importance of taking attentional deficits as a possible cause of change blindness into account when using change detection paradigms to make inferences about the fidelity of memory representations.

Change detection rates in Experiment 1 also showed a pattern similar to those found by Luck and Vogel (1997) when they conducted change detection experiments to assess VWM capacity with a varying number of colored squares on a grey background as stimuli and shorter display times of just a tenth of a second. This indicates that similar mechanisms of encoding and retrieval were triggered in our experiment and supports the prominent idea that VWM capacity is to some extent limited by a certain amount of items rather than

features as objects cued in our experiment were way more complex than single-colored shapes (e.g., Luck & Vogel, 1997; for a review see Brady, Konkle, & Alvarez, 2011).

The observed superiority of simultaneous to sequential cueing is not trivial: One may at first believe that when provided with all cues at the same time, subjects could not use the limited study phase to scan objects in an efficient way as they would get stuck on more salient objects or misestimate the time limit and either scan objects in an overly thorough manner or refixate too hastily, thus not being able to store enough relevant information. The fact that we not only could not find evidence for such shortcomings but observed the opposite, that is, simultaneous cueing being superior to its sequential counterpart, indicates that the creation of visual representations of scenes is a more complex process than just serially accessorizing some sort of mental canvas with objects that we are attending to. The set of manipulations chosen in our experiments, however, does not enable us to give a definite characterization of the nature of this process as several aspects might have contributed to the higher detection rates in Experiment 1, three of which we consider plausible enough to outline them in the following. First, only simultaneous presentation of cues allowed participants to freely choose the order in which to look at the cues which might have caused them to apply more effective chunking techniques. For example, they might have scanned cued objects in close proximity at once and remembered them as a unit (e.g., a collection of different objects on a table). This perhaps reduced the number of memory slots occupied with distinct object information. When cues appeared sequentially in a random order in Experiment 2, this chunking process was probably applied less as it was not likely that cues for objects next to each other appeared right after another, probably resulting in rather counterintuitive saccades. We, however, highlight that this is just speculation as the participants' eye-movements were not recorded. Second, it is likely that participants in the first experiments refixated on previously attended objects and were thus able to rehearse

information about objects in the scene leading to a richer representation. The successive nature of cue appearance in Experiment 2 limited the amount of time participants had to do this as they had to focus on information of the newly incoming cues. Third, it is possible that subjects in Experiment 1, due to their more natural perception of the scene, used its structure (i.e., additional information from outside the cued areas) and meaningful contextual links between the cued objects more effectively than subjects in Experiment 2 and therefore stored a stronger representation of the scene as a collective, rather than a fragile list of individual objects. The importance of spatial context and relations between individual objects for storage of information in VWM has been demonstrated for abstract shapes (Jiang, Olson, & Chun, 2000) and might therefore also apply to scenes, which are even more rule-governed and rich in familiar context. To investigate whether the scene's structure has an impact on change detection performance for objects in visually rich environments like the ones used in our experiments, future research could—in a similar fashion as Josephs, Draschko, Wolfe, and Võ (2016) did for memory in visual search—study the effect of a removal of scene-relevant information and how gradually adding it affects change detection accuracy.

The fact that differences between the experiments are hardly visible when only overall accuracy of the YN task is considered, might stem from a response bias towards indicating that there was no change, resulting in disproportionately inflated correct rejection rates, especially in Experiment 2: As a matter of fact, participants did on average press “N” on 55% and 61% of trials in Experiments 1 and 2, respectively. Correct rejections, however, cannot be expected to capture change detection performance as they just indicate that the subject did not notice a change, not, however, that they were aware that every specific object was still present and in place. We also cannot rule out that the mere fact that there was a change localization task (mouse click to the area of change) influenced behavior in the YN

task: Participants might have pressed “N” even if they felt something had changed because they apprehended failure localizing the difference between the images.

Returning to the question of whether “massive memory” findings can stand the test of change detection experiments, so to speak, we would lastly like to mention the possibility that deficits in detecting changes may occur while memory representations are highly detailed at the same time. Though this may seem like the most paradox explanation as both tasks rely on the same information, research by Varakin and Levin (2006) using arrays of real-world object drawings as stimuli, indicates that recognition memory for object details can be high while changes to these very same details still go unnoticed. They therefore conclude that “visual memory is, in some sense, richly detailed, but at the same time, acts poorly” (p. 73). Future research could investigate whether this also holds for change detection in naturalistic scenes like the ones used in our experiments.

In summary, our findings support the idea that both attentional deficits and limitations of memory contribute to the phenomenon of change blindness. Differences we observed between simultaneous and sequential attentional direction indicate that embedding object percepts into the mental representation of a scene is a more than just taking in one object after another in a snapshot-like fashion: It is rather a more complex process that might likely also rely on contextual relations of the different objects. On a broader level, it seems like the idea of characterizing memory systems in terms of their overall capacity and the fidelity of their representations alone, may not be sufficient to appropriately capture their essence. Instead, a meaningful description of a memory system and its limitations should remain linked to the task that was used to capture it. Otherwise estimates of the boundaries of visual memory fidelity gained from different paradigms like alternative forced choice recognition and change detection seem to lack compatibility.

References

- Atkinson, R. C., & Shiffrin, R. M. (1968). Human memory: A proposed system and its control processes. In K. W. Spence & J. T. Spence (Eds.), *The Psychology of Learning and Motivation* (Vol. 2, pp. 89–195). New York, NY: Academic Press.
[https://doi.org/10.1016/S0079-7421\(08\)60422-3](https://doi.org/10.1016/S0079-7421(08)60422-3)
- Brady, T. F., Konkle, T., & Alvarez, G. A. (2011). A review of visual memory capacity: Beyond individual items and toward structured representations. *Journal of Vision*, 11(5), 4. <https://doi.org/10.1167/11.5.4>
- Brady, T. F., Konkle, T., Alvarez, G. A., & Oliva, A. (2008). Visual long-term memory has a massive storage capacity for object details. *Proceedings of the National Academy of Sciences*, 105(38), 14325–14329. <https://doi.org/10.1073/pnas.0803390105>
- Brady, T. F., Konkle, T., Oliva, A., & Alvarez, G. A. (2009). Detecting changes in real-world objects. *Communicative & Integrative Biology*, 2(1), 1–3.
<https://doi.org/10.1073/pnas.0803390105>. www.landesbioscience.com
- Castelhano, M. S., & Henderson, J. M. (2007). Initial scene representations facilitate eye movement guidance in visual search. *Journal of Experimental Psychology: Human Perception and Performance*, 33(4), 753–763. <https://doi.org/10.1037/0096-1523.33.4.753>
- Chang, W. (n.d.). Plotting means and error bars (ggplot2). Retrieved July 8, 2017, from [http://www.cookbook-r.com/Graphs/Plotting_means_and_error_bars_\(ggplot2\)/#helper-functions](http://www.cookbook-r.com/Graphs/Plotting_means_and_error_bars_(ggplot2)/#helper-functions)
- Ehinger, K. A., Allen, K., & Wolfe, J. M. (2016). Change blindness for cast shadows in natural scenes: Even informative shadow changes are missed. *Attention, Perception, & Psychophysics*, 78(4), 978–987. <https://doi.org/10.3758/s13414-015-1054-7>
- Hollingworth, A., & Henderson, J. M. (2002). Accurate visual memory for previously

- attended objects in natural scenes. *Journal of Experimental Psychology: Human Perception and Performance*, 28(1), 113–136. <https://doi.org/10.1037//0096-1523.28.1.113>
- Holm, S. (1979). A simple sequentially rejective multiple test procedure. *Scandinavian Journal of Statistics*, 6(2), 65–70. <https://doi.org/10.2307/4615733>
- Jiang, Y., Olson, I. R., & Chun, M. M. (2000). Organization of visual short-term memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 26(3), 683–702. <https://doi.org/10.1037/0278-7393.26.3.683>
- Josephs, E. L., Draschkow, D., Wolfe, J. M., & Võ, M. L.-H. (2016). Gist in time: Scene semantics and structure enhance recall of searched objects. *Acta Psychologica*, 169, 100–108. <https://doi.org/10.1016/j.actpsy.2016.05.013>
- Konkle, T., Brady, T. F., Alvarez, G. A., & Oliva, A. (2010). Scene memory is more detailed than you think: The role of categories in visual long-term memory. *Psychological Science*, 21(11), 1551–1556. <https://doi.org/10.1177/0956797610385359>
- Lawrence, M. A. (2016). ez: Easy analysis and visualization of factorial experiments. Retrieved from <https://cran.r-project.org/package=ez>
- Levin, D. T., & Simons, D. J. (1997). Failure to detect changes to attended objects in motion pictures. *Psychonomic Bulletin & Review*, 4(4), 501–506. <https://doi.org/10.3758/BF03214339>
- Luck, S. J., & Vogel, E. K. (1997). The capacity of visual working memory for features and conjunctions. *Nature*, 393(1996), 1996–1998.
- Mathôt, S., Schreij, D., & Theeuwes, J. (2012). OpenSesame: An open-source, graphical experiment builder for the social sciences. *Behavior Research Methods*, 44(2), 314–324. <https://doi.org/10.3758/s13428-011-0168-7>
- R Core Team. (2017). R: A language and environment for statistical computing. Vienna,

Austria. Retrieved from <https://www.r-project.org/>

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

<https://doi.org/10.1111/j.1467-9280.1997.tb00427.x>

RStudio Team. (2015). RStudio: Integrated development environment for R. Boston, MA.

Retrieved from <http://www.rstudio.com/>

Sareen, P., Ehinger, K. A., & Wolfe, J. M. (2016). CB Database: A change blindness database for objects in natural indoor scenes. *Behavior Research Methods*, 48(4), 1343–1348. <https://doi.org/10.3758/s13428-015-0640-x>

Simons, D. J., & Levin, D. T. (1997). Change blindness. *Trends in Cognitive Sciences*, 1(7), 261–267.

Simons, D. J., & Levin, D. T. (1998). Failure to detect changes to people during a real-world interaction. *Psychonomic Bulletin & Review*, 5(4), 644–649.

<https://doi.org/10.3758/BF03208840>

Simons, D. J., & Rensink, R. A. (2005). Change blindness: Past, present, and future. *Trends in Cognitive Sciences*, 9(1). <https://doi.org/10.1016/j.tics.2004.11.006>

Standing, L. (1973). Learning 10,000 pictures. *Quarterly Journal of Experimental Psychology*, 25(2), 207–222. <https://doi.org/10.1080/14640747308400340>

Varakin, D. A., & Levin, D. T. (2006). Change blindness and visual memory: Visual representations get rich and act poor. *British Journal of Psychology*, 97(1), 51–77. <https://doi.org/10.1348/000712605X68906>

Võ, M. L.-H., & Henderson, J. M. (2010). The time course of initial scene processing for eye movement guidance in natural scene search. *Journal of Vision*, 10(3), 14.1-13. <https://doi.org/10.1167/10.3.14>

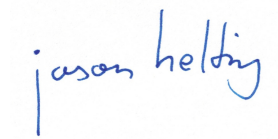
Wickham, H. (2009). *ggplot2: Elegant graphics for data analysis*. Springer-Verlag New

York. Retrieved from <http://ggplot2.org>

Wolfe, J. M. (1998). Visual memory: What do you know about what you saw? *Current Biology*, 8(9), R303-4. [https://doi.org/10.1016/S0960-9822\(98\)70192-7](https://doi.org/10.1016/S0960-9822(98)70192-7)

Eigenständigkeitserklärung

Ich versichere, dass ich die vorliegende schriftliche Arbeit selbstständig verfasst und keine anderen als die angegebenen Quellen und Hilfsmittel benutzt habe. Die Stellen, die anderen Werken im Wortlaut oder dem Sinn nach entnommen sind, sind durch Quellenangaben im Text deutlich gemacht. Ich habe die Arbeit in keinem anderen Studiengang, auch nicht auszugsweise, als Studien- oder Prüfungsleistung verwendet.



Frankfurt am Main, 19. Juli 2017

Jason Helbing