What you see is what you need

Jochen Triesch
Department of Cognitive Science, UC San Diego, La Jolla, CA, USA

Dana H. Ballard
Department of Computer Science, University of Rochester, Rochester, NY, USA

Mary M. Hayhoe
Center for Visual Science, University of Rochester, Rochester, NY, USA

Brian T. Sullivan
Center for Visual Science, University of Rochester, Rochester, NY, USA

We studied the role of attention and task demands for implicit change detection. Subjects engaged in an object sorting task performed in a virtual reality environment, where we changed the properties of an object while the subject was manipulating it. The task assures that subjects are looking at the changed object immediately before and after the change. Our results demonstrate that in this situation subjects’ ability to notice changes to the object strongly depends on momentary task demands. Surprisingly, frequent noticing is not guaranteed by task relevance of the changed object attribute per se, but the changed object attribute needs to be task relevant at exactly the right times. Also, the simplicity of the used objects indicates that change blindness occurs in situations where the visual short term memory load is minimal, suggesting a potential dissociation between short term memory limitations and change blindness. Finally, we found that changes may even go unnoticed if subjects are visually tracking the object at the moment of change. Our experiments suggest a highly purposive and task specific nature of human vision, where information extracted from the fixation point is used for certain computations only “just in time” when needed to solve the current goal.

Keywords: change blindness, inattentional blindness, eye movements, attention, visual cognition, virtual reality

Introduction

In recent years a number of studies have investigated a phenomenon that is now usually referred to as change blindness (Simons & Levin, 1997; Intraub 1997), which is closely related to so-called inattentional blindness (Mack & Rock, 1998; Simons, 2000b). In these experiments subjects display an often surprising inability to notice changes to the visual scene occurring during retinal transients, as produced by, e.g. saccades, eye blinks, movie cuts, or “mud splashes” (O’Regan, Rensink, & Clark, 1999). These experiments have questioned a number of assumptions about the nature of visual representations. Despite the recent surge of interest in this phenomenon, the underlying mechanisms are still controversial (Simons, 2000b). While it seems clear that limitations of visual short term memory are relevant for change blindness, it is less clear if such limitations are necessary or merely sufficient for change blindness.

In typical change blindness experiments the subjects are explicitly instructed to look for changes; these are explicit change detection tasks. It is unclear in how far results obtained in these experiments can be generalized to normal visually guided behavior where subjects do not expect any changes. To better understand this, we need to study change blindness in tasks where subjects are unaware that changes might happen. These are implicit change detection tasks. Ideally, these tasks are natural, closely reflecting the perceptual and computational demands present in real life behaviors (Shinoda, Hayhoe, & Shrivastava, 2001). This entails using natural 3-dimensional scenes of realistic extent, scale, and complexity instead of, say, simple 2-dimensional arrays of letters confined to a small region of the visual field. It may also be important to use self-paced, continuing tasks where the timing of visuo-motor operations is controlled by the subject rather than the experimenter. Simons & Levin (1998) have done pioneering work studying change blindness phenomena in the real world but the drawback of experimenting in the real world is that the stimulus cannot be controlled precisely and reproducibly and that it is more difficult to obtain behavioral measures like eye movement records than in a controlled laboratory environment. We feel that a good compromise is to use virtual reality technology. While rendering quasi-realistic natural scenes, it gives the experimenter perfect control over all details of the scene, and allows perfect reproduction of the visual stimulus. Although relatively new, we expect to see more research using virtual reality in the future (Pelz et.al., 1999; von der Heyde & Bülthoff, 2000).
Our main concern in this paper is the role of attention and task demands in determining subjects' ability to notice changes. Earlier work by Rensink and colleagues has used the notion of centers of interest (Rensink, O'Regan, & Clark, 1997; O'Regan et al., 1999). Subjects subjectively rated what image regions they perceived as most interesting. It was found that changes occurring at these centers of interest were noticed more easily than other changes in a standard flicker task. In the flicker paradigm, the display is switched back and forth between an image and a slightly changed copy of it with a briefly flashed blank screen masking the transition from original to copy. We feel that in dynamic ongoing tasks this notion of a center of interest is not powerful enough to accurately describe subjects' distribution of processing resources. In particular, we are interested in the more fine grained dynamic properties of attention during ongoing natural behaviors. Our hypothesis is that a crucial variable for subjects' abilities to notice changes is the exact timing of the point(s) in a task that a subject needs to extract a piece of task-relevant visual information. To explore this idea, we engaged subjects in different versions of an object sorting task, where the different versions of the task manipulated at what points in the task a changing object attribute would be relevant or irrelevant for the successful completion of the task.

**Methods**

**Virtual Reality Setup**

Experiments are performed using the virtual reality setup shown in Figure 1. The system's backbone is an SGI ONYX-2 workstation rendering stereo image pairs at a frame rate of 60Hz. The images are displayed using head mounted goggles. We use a V8 virtual reality helmet from Virtual Research with dual 640 by 480 pixels. The helmet is equipped with a magnetic head tracking device that measures the head's position and orientation with respect to a fixed laboratory reference frame (Polhemus Fastrak). The magnetic tracker operates at 120Hz with a 4ms internal latency. This information is passed on to the graphics engine to determine the viewpoint(s) from which to render the virtual scene with a 1-2 frame latency. Integrated into the helmet is a video based eye tracker (bright pupil type, model 501 from Applied Science Laboratories) with 1 degree accuracy operating at 60Hz. Force feedback from physical interaction with objects in the environment is given with two haptic stimulation devices that allow subjects to grasp objects between thumb and index finger of one hand while experiencing realistic forces. To this end we use two Phantom-3 devices from SensAble Technologies in opposition — one for the index finger and one for the thumb (Figure 1). The usable work-space volume is about 40cm by 40cm by 40cm. The haptic force feedback is provided to the subject at a rate of 1 kHz. Subjects get visual feedback about their thumb and index finger position in the form of small spheres displayed in the virtual world (compare Figure 1, one sphere is on the face of the central brick, the other is hidden behind the brick). This visual feedback is provided with a typical delay below 17ms that originates from rendering the scene at 60Hz.

**Task Environment**

Subjects sort bricks of two different heights (but same width and depth) located in a pick-up area onto two conveyor belts according to different rules. The dimensions of the short and tall bricks are 6cm by 6cm by 8cm and 6cm by 6cm by 10cm, respectively. For the typical viewing distance the two different heights correspond to about 7.6 and 9.5 degrees of visual angle. Subjects can easily categorize a single brick as being short or tall without seeing it next to bricks of the other category (see Figure 1).

The “atomic behavioral unit” of the experiment is a single pick and place action consisting of pick up, carry over, and put down. A block consists of five pick and place actions, after which five new bricks appear. A session...
comprises 20 blocks for a total of 100 pick and place actions. In all change blindness experiments, the changes have to be masked by some retinal transient. Previous work has used flashed blank screens, saccades, eye blinks, temporary occlusions, and more. For our experiment, we found that subjects who are told to do the task quickly show a quite reliable pattern of saccadic eye movements that we exploit for masking the size changes. We found that subjects typically fixate the brick they intend to move during pick up. Once they have lifted it off the ground, they make a saccade to the conveyor belt area and fixate there for guiding the brick onto one of the conveyor belts. Since this pattern is very reliable, we can simply change the brick’s size when it is mid way between the pick up area and the conveyor belts. This ensures that the change occurs during or shortly after the saccade most of the time. In 10 percent of the pick and place actions, the height of the brick changes while the subject moves it from the pick up area to the conveyor belts. Since subjects are instructed to grasp the bricks with their fingers touching the front and back side of the brick (see Figure 1), the pure height change does not give subjects any haptic feedback about the change.

Subjects are not told that these changes can happen but are instructed to report any suspicious events they notice since the software is still under development. If a subject reports a size change, we instruct the subject to also report future occurrences. While subjects are performing the task we record their hand movements with the haptic feedback devices and their eye-movements using the eye tracker positioned inside the head mounted display. We also record a video of the subjects’ view inside the helmet with superimposed cross-hairs marking their moment-to-moment gaze direction.

The Three Conditions

For each pick and place action the subject has to make two decisions: which brick to pick up and where to put it down. In order to systematically address the role of attention and task demands for noticing changes we gave subjects three different instructions for sorting the bricks that altered for which of the two decisions the size of the bricks was relevant:

1. “Pick up the bricks in front to back order and place them on the closer conveyor belt.” In this case size is irrelevant for both decisions.
2. “Pick up the tall bricks first and put them on the closer conveyor belt. Then, pick up the small bricks and also put them on the closer conveyor belt.” For this condition size matters for only the first decision (which to pick up).
3. “Pick up the tall bricks first and put them on the closer conveyor belt. Then, pick up the small bricks and put them on the distant conveyor belt.” Here, brick size is relevant for both decisions.

Example movies of subjects performing the three different tasks are shown in the Appendix together with a movie of a virtual play back of a section of an experiment. We also considered a fourth condition, where subjects were asked to pick up bricks in front to back order and put the tall ones on the closer conveyor belt and the short ones on the far conveyor belt. The rationale of this being that brick size would be relevant only during put down of a brick. Preliminary results in this condition were identical to condition three, however. We believe that the reason for this is that in this fourth condition the brick size is already relevant during the pick up of the brick because the subsequent arm movement has to be targeted towards the proper conveyor belt. On these grounds we decided to abandon the fourth condition.

Fifty-nine subjects participated in the experiment – 17, 22, and 20 in conditions 1, 2, and 3, respectively. Subjects were students at the University of Rochester with normal or corrected to normal vision. They received monetary reimbursement for their participation. Satisfactory eye tracking was obtained for 44 out of the 59 subjects (75%). Subjects were naive to the purpose of the experiment. In addition to the trial by trial reporting subjects filled in a questionnaire at the end of the experiment that explicitly asked whether they had noticed any bricks changing size and if so, how often they noticed.

Results

Noticing of Changes

Regarding the reporting of noticed changes, it turned out that some subjects did not report size changes right away but nevertheless claimed to have noticed some when asked at the end of the experiment, sometimes stating that they deemed the changes to be irrelevant. This was observed most frequently in the first condition. The results are depicted in Figure 2. The questionnaire responses are consistently higher, and the difference between the two reports is significant for the first and second condition (t-test, \( p=0.01 \) and \( p=0.04 \), respectively). There are obvious difficulties in getting subjects to report changes without telling them about them. Neither of the two measures we collect may be equal to the true probability of detection. But critical for the current experiment are the differences in noticing between the three conditions. These show the same trends irrespective of which measure of subjects’ noticing is considered.

Subjects noticed very few changes in condition 1, a few more in condition 2, but they noticed many changes in condition 3. The differences are significant in all cases (pair-wise t-tests, verbal reports: C1-C2: \( p=0.007 \), C1-C3: \( p<0.001 \), C2-C3: \( p=0.001 \); questionnaire reports: C1-C2: \( p=0.022 \), C1-C3: \( p<0.001 \), C2-C3: \( p=0.026 \)).
The data in Figure 2 do not show how the frequency of noticing varies between subjects in the same task condition. To illuminate this, we computed histograms where subjects were sorted into bins depending on what percentage of changes they spontaneously reported. The data are shown in Figure 3. Clearly, noticing of changes for an individual subject is not “all-or-nothing” but typically individual subjects will spontaneously report a varying fraction of the changes. We also computed the percentage of subjects who did not spontaneously report any brick changes at all. These are 88%, 45%, and 5% for the three groups, respectively. The strongly increased ability to notice changes in the third condition is also reflected in the number of unnoticed changes that occurred before the first change was noticed by a subject. In conditions one and two, the average number of unnoticed changes prior to the first noticed change was 7.2 and 6.5, respectively, while it was only 1.0 for condition 3. Note that this analysis uses the following definition: if the subject did not notice any change, we defined the number of unnoticed changes prior to noticing the first change as the total number of changes occurring. Hence, our figures for conditions one and two must be regarded as lower bounds to the true number of changes that initially go unnoticed.

It is an interesting question in how far the first noticed change sensitizes a subject to noticing subsequent changes. In the extreme case, it may be that once a subject notices a change the subject will be sensitized enough to detect all subsequent changes. However, our data do not support this. Even if a subject notices a change, the subject may miss a number of subsequent changes. On average, we found that the number of missed changes subsequent to the first detected one is 5.7 for group one (n=3), 2.8 for group two (n=9) and 1.5 for group three (n=11).

Gaze Analysis

We studied subjects’ gaze direction at the time of the object change in order to verify our assumption that changes would normally occur during saccades. Also we wanted to find out whether the big differences in noticing of changes in the three conditions may be caused by subjects using their gaze differently. The results are shown in Figure 4.
We distinguished seven classes for the eyes' activity at the moment of the size change. The classification is based on a frame by frame analysis of the videotaped records of the eye tracker's estimate of gaze position and the videotaped image of the eye tracker's camera monitoring the subject's left eye. Combining the two, we could estimate saccade beginnings and saccade ends to a temporal precision of one video frame. The seven classes are:

- **tracking**: the eyes are tracking the brick,
- **saccade onset**: the eyes are just starting to move from the pick up place or the brick to the put-down region ($\pm 1$ video frame),
- **saccade**: the change happens during a saccade from pick-up to put-down region,
- **after saccade**: the eyes are just arriving in the put-down region ($\pm 1$ video frame),
- **blink**: the change happens during an eye blink that does not happen in conjunction with a saccade,
- **elsewhere**: the eyes are fixating or making smooth pursuit movements in a different region of the workspace,
- **other**: everything else including track losses.

The distribution of gaze activity looks very similar in the three conditions. In particular, the ratio of trials where subjects are tracking the brick during the change, and where we expected change detection to be most likely, is about equal in all three conditions.

Pair-wise $\chi^2$ tests showed only insignificant differences between the conditions from which we conclude that the different ratios of reported changes are not an effect of subjects using their gaze differently at the moment of the change. To our surprise, we found that tracking the object is not sufficient for detecting the size change. There are instances when the size change happens while the subject is looking directly at the brick but the subject neither reports noticing the change spontaneously nor during questioning at the end of the experiment. This is illustrated in Figure 5.

While the analysis of gaze activity at the moment of the object change did not reveal any differences between the three conditions, we also tested whether the overall pattern of fixations employed by subjects would reveal different patterns indicating different gaze strategies in the three conditions. Also, we wanted to see whether we could find differences of eye movement patterns in the presence of unnoticed object changes that would suggest an "implicit noticing" of the changes. To answer these questions we used the video records of a number of subjects' gaze activities during the experiment and coded them for locations and durations of fixations. The analysis was performed on 9 subjects of group one, 8 subjects of group two and 11 subjects of group three. Regarding overall differences in gaze strategies in the

![Figure 5](http://jov.arvojournals.org/pdfaccess.ashx?url=/data/journals/jov/933043/)
different conditions it is most interesting to look at the fixations during the put-down phase of the pick-and-place actions, since they occur after potential size changes. We defined as the start of this put-down phase the time when the eyes cross the mid-plane of the work space from left to right before the brick is put down. The end of the put-down phase is reached when the eyes cross the mid-plane of the work space from right to left after the brick has been dropped on the conveyor belt. We tested whether the time during this put-down interval that subjects spent fixating the brick was different between the three conditions. Since we are interested in differences due to different processing strategies that subjects may be using in the three conditions rather than differences occurring because different numbers of changes were noticed, it is useful to compare the trials where no change occurred. The summed fixation durations that subjects spent looking at the brick during putting it down are plotted in Figure 6.

The reported times are very similar in the three conditions and indeed we could not find any significant differences. The data for trials where changes did occur but were not noticed is very similar to these data. There are no significant differences between the three conditions or between no-change trials and unnoticed-change trials for the same condition. Thus, unnoticed changes do not appear to be accompanied by prolonged overall fixation durations after the change. This result should be contrasted to a previously reported study, where prolonged fixation durations had been observed for a blocks copying task (Hayhoe, Bensinger, & Ballard, 1998). When a change was noticed by the subject the total time fixating the brick during put down was increased by roughly a third of a second averaged over all conditions (after removal of one outlier with fixation time exceeding 2 sec.) and this difference was significant for all conditions. We also performed the same analysis considering the cumulative times that subjects spent during the entire put-down action (not just the time they fixated the brick). This led to similar conclusions.

**Discussion**

We studied whether and how task demands can affect observers’ ability to notice changes in a natural task. The influence of movement and task was previously studied in a change blindness paradigm by Wallis and Bulthoff (2000) who compared change detection for active drivers versus passive passengers of a virtual car. They found that detection of changes away from the line of motion was impaired only for active drivers. Unfortunately, in this experiment the gaze direction of subjects was not measured. Thus, the experiment could not answer whether or not differences in noticing changes were just due to different use of gaze. The importance of gaze for the noticing of changes was investigated by Henderson and Hollingworth (1999). They found that fixation position and saccade direction play an important role in determining whether changes will be noticed. In particular, they found that the disappearance of an object was noticed more easily when it occurred during a saccade towards the object rather than away from the object.

We wanted to systematically vary the relevance that the changed object attribute had at different stages of the task. To this end, subjects faced different experimental conditions, where the changed attribute (size of a brick) was relevant at different stages of the task (pick up and put down of the brick). We used a virtual reality setup in order to be able to a) provide a naturalistic environment, b) control the visual scene precisely and reproducibly, and c) measure a number of behavioral variables including eye and hand movements. We found that subjects’ ability to notice changes was strongly affected by when exactly the changed object attribute was task relevant. Surprisingly, in conditions two and three the brick sizes were task relevant but the results in these conditions were strikingly different. Subjects noticed many changes only in condition three where the brick size was relevant before and after the change. We confirmed that this effect is not due to a different use of gaze in the three conditions. The distribution of gaze activity at the moment of change and the patterns of fixations during put-down of a brick are very similar in all conditions. Thus, the effect is likely due to a difference in central processing. Interestingly, some changes went unnoticed even if the subject was tracking the brick with his/her eyes. A similar finding has recently been reported by O’Regan, Deubel, Clark, & Rensink (2000). They studied change blindness in a flicker paradigm with changes made during eye blinks. They found that some changes may go unnoticed even if the
subject is looking directly at the changed location (within 1 degree). A possible interpretation of these results is that much less information is computed automatically by the visual system than was previously thought. Most information may be computed “on demand” by engaging specialized functional routines at just the right times (Ullman, 1984; Ballard, Hayhoe, & Pelz, 1995; Hayhoe, 2000; Roelfsema, Lamme, & Spekreijse, 2000).

The differences in frequency of change detection between the three conditions are striking. While we would like to suggest an interpretation of this data where information during fixation is extracted largely on demand, it is instructive to consider alternatives. For example, one might argue that the differences in size change detection between the three conditions are due to an increase in general "perceptual arousal", driven by "task complexity" (however this notion would be formalized). This position would argue that the visual system is more attentive to changes in any stimulus feature, regardless of the task relevance of that feature, simply due to the enhanced general attention required by the subject when performing a task with a more complex set of rules. While we cannot rule out such a possibility on the basis of our data, the question is amenable to experimental analysis in the following way. The “task complexity” hypothesis would predict an increase in change detection for any increase in task complexity, regardless of whether or not the complexity was related to the objects’ features. In contrast, we predict a higher frequency of change detection only if the increased task complexity is related to information selectively extracted during fixation. Future work should address this issue.

Explaining Change Blindness

Previous attempts at explaining change blindness effects have usually considered limitations of visual short term memory as the underlying cause (Irwin, 1996; Irwin & Gordon, 1998). While limitations of visual short term memory clearly set an upper limit on the ability to notice changes, in our experiment, however, visual short term memory requirements are arguably minimal. The only parameters of a brick’s appearance are its size and its (irrelevant) color and subjects attend to the brick directly before and after the change. Our experiment supports the intriguing possibility that the failure to notice changes in change blindness experiments may not always be due to the limited capacity of visual short term memory but rather a failure to engage it despite attending to the object. This idea is consistent with earlier findings suggesting that humans seem to structure tasks so as to minimize short term memory requirements (Ballard, Hayhoe, & Pelz, 1995; Hayhoe et al., 1998).

Recently, Simons proposed five hypothetical “causes of change blindness” (Simons, 2000b). Reviewing the experimental evidence he finds support for each of them.

2. First Impression: The old representation persists, the new one is ignored.
3. Feature combination: The representation after the change has elements of the object's appearance before and after the change.
4. Nothing is stored: No representation of the object is maintained at all.
5. Nothing is compared: Representations of the object before and after the change co-exist without being compared.

We feel that rather than being independent causes of change blindness the first four of these represent intimately related consequences of the highly purposive and task specific nature of visual operations. Our hypothesis is that in every day tasks only a very limited amount of visual information is “computed” at each fixation — just enough to solve the current sensorimotor micro task. Under this hypothesis, Simon’s causes of change blindness are merely different effects of the highly purposive and task specific nature of visual processing. It appears that the visual system extracts certain information from the visual scene only “just in time” when needed to solve the current goal. This interpretation raises a set of new questions: just how is it that people select their moment to moment goals in everyday tasks? What tactics do they use to negotiate multiple simultaneous goals? What are the neural correlates of these dynamic changes in processing? It is interesting to note in this context that our results show graded differences between the three conditions rather than an “all-or-nothing” result. It is not that subjects never notice any changes in condition one or that subjects notice all changes in condition three. Also, even if subjects in condition three have already noticed a couple of changes they may still miss subsequent ones. This suggests that changes in subjects' processing strategies between the three conditions may also be of a gradual nature.

Appendix

For illustration, we include movies of the actual task in Movie 1. Shown are movies of the virtual workspace scene displayed inside the head mounted display. The superimposed cross hairs indicate the subject’s momentary direction of gaze. The superimposed image in the upper left hand corner is the image of the subject’s left eye as seen by the eye tracking device mounted inside the helmet. The infrared illumination of the eye tracker makes the pupil appear bright. Cross hairs and eye image are superimposed only for the purpose of analysis and are not visible to the subject during the actual task.
The detailed data record we gather during an experiment allows us to construct a virtual playback of the entire experiment from an arbitrary point of view. This is illustrated in the movie in Movie 2.

Movie 1. Example movies of subjects performing the task in conditions one, two, and three, respectively. The movies are slowed down by a factor of two for easier viewing. Left: Condition 1 with unnoticed change of the third brick being moved. Middle: Condition 2 with unnoticed change of the fourth brick being moved. Right: Condition 3 with noticed change of the third brick being moved. Higher resolution movies may be obtained from the authors.

Movie 2. Virtual playback of the experiment.

Acknowledgments

This work was supported in part by NIH/PHS research grant P41 RR09283 and by grant EY05729. We would like to thank Jelena Jovancevic and Diane Kucharczyk for coding of eye movements and Paul Ilardi, Keith Parkins, and Peter Skirko for programming support. Commercial relationships: none.

Footnotes

1. The exact wording was: "We just updated the software for the experiment today and haven't had a chance to test it thoroughly. The virtual objects should behave as if they were real objects. If anything unusual happens at any point during the experiment, can you please stop and tell me immediately so we can look at what's going wrong?"

2. The distribution of frequency of noticing across different subjects is not strongly bimodal, so the application of a t-test appears justified. Compare Figure 3.

3. According to the questionnaire reports these numbers are 53%, 27%, and 5% for the three groups, respectively.

4. On the other hand, in some situations one might predict a decrease of noticing ability with increased task complexity due to the additional attentional load interfering with the processing resources needed for change detection.

References


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, 368-373.


