

Effects of scene inversion on change detection of targets matched for visual salience

Todd A. Kelley Department of Psychology, Vanderbilt University, Nashville, TN, USA

Marvin M. Chun Department of Psychology, Vanderbilt University, Nashville, TN, USA 

Kao-Ping Chua Department of Psychology, Vanderbilt University, Nashville, TN, USA

This work examines how context may influence the detection of changes in flickering scenes. Each scene contained two changes that were matched for low-level visual salience. One of the changes was of high interest to the meaning of the scene, and the other was of lower interest. High-interest changes were more readily detected. To further examine the effects of contextual significance, we inverted the scene orientation to disrupt top-down effects of global context while controlling for contributions of visual salience. In other studies, inverting scene orientation has had inconsistent effects on detection of high-interest changes. However, this experiment demonstrated that inverting scene orientation significantly reduced the advantage for high-interest changes in comparison to lower-interest changes. Thus, scene context influences the deployment of attention and change-detection performance, and this top-down influence may be disrupted by scene inversion.

Keywords: change detection, visual attention, visual perception, context, scene processing

Introduction

The visual world is highly complex, and visual experience appears to be rich and detailed. However, at any given moment, we have detailed access to only the small, attended subset of a scene. Because the attended portion is not selected at random, a fundamental question is how the visual system deploys attention. Primarily, there are many bottom-up physical factors that draw attention, such as color, size, proximity, and brightness (Bravo & Nakayama, 1992; Treisman & Gelade, 1980; Wolfe, 1994). Of greater interest here, though, are those factors that are the result of background knowledge of the object and the world in which the object typically occurs. One example of such top-down influence is visual context.

Considerable evidence shows that our expectations and knowledge of a scene influence how we perceive objects associated with that scene. Identification of objects is impaired when the given object is incongruent with the context of a paired scene (Biederman, Mezzanotte, & Rabinowitz, 1982; Boyce, Pollatsek, & Rayner, 1989; Palmer, 1975; but see Hollingworth & Henderson, 1998). For example, when asked to specify whether a given object appeared at a probed position in a scene, subjects did not perform as well if the object violated the context of the scene, such as a couch floating in the sky and a fire hydrant sitting on top of a mailbox. Clearly, the context of a scene influences how embedded objects are perceived.

In addition to these studies, a paradigm called contextual cueing further reveals how contextual information facilitates visual search for targets embedded

in complex arrays (Chun, 2000; Chun & Jiang, 1998; Chun & Phelps, 1999). Targets were detected more quickly when the surrounding contexts were predictive of target location or shape. Observers learned which contexts were predictive through implicit learning of repeated displays.

The effects of context on scene processing can also be studied using the change blindness paradigm. Change blindness refers to the difficulty in detecting alterations in scenes, revealing that subjects do not have ready access to certain events within scenes (Simons & Levin, 1997). When provided with written verbal cues to guide attention, subjects improved at detecting changes (Rensink, O'Regan, & Clark, 1997). This indicates that attention is crucial for noticing changes. Using change blindness tasks, Rensink et al. showed that the context of a scene might also direct attention independently of any outside cues. Subjects were presented with scenes¹ in which a change occurred in an object of central (high) interest to the scene (e.g., a helicopter seen from the cockpit of another aircraft) and others where a change occurred to objects of marginal (low) interest (e.g., a railing located behind two people eating lunch). The images were presented using the flicker paradigm. This method cycles the standard and altered scenes with a blank scene in between, creating the impression that the scene is flickering on and off. The flicker creates the global visual transient needed to distract attention away from the local transient occurring at the location of change. Subjects performing this task displayed a center

of interest effect, locating the central changes in less than half the time required for the marginal changes.

Shore and Klein (2000) extended Rensink and colleagues' (1997) findings in a series of experiments that used upside-down (inverted) scenes to disrupt the effects of global context. Using the same set of stimuli from Rensink et al., Shore and Klein presented pairs of printed images side by side and measured the time required to identify a difference between the two images. For upright images, a significant reaction time (RT) advantage was shown for detecting central changes versus marginal changes. However, when the image pairs were presented upside down to weaken the influence of global context, the difference between central and marginal change detection was reduced dramatically. Thus, Shore and Klein extended the findings of Rensink et al. using a different type of change-detection task and a novel manipulation to disrupt the effects of scene context.

The flicker paradigm and the simultaneous paradigm are just two examples of a wider variety of methods for studying change blindness (Grimes, 1996; Henderson & Hollingworth, 1999; McConkie & Currie, 1996; Rensink et al., 1997; O'Regan, Rensink, & Clark, 1999; Shore & Klein, 2000). Ideally, the particular method for testing change blindness should not affect whether context effects will be observed or not. However, Shore and Klein revealed a difference in results between the flicker paradigm and the simultaneous paradigm. For upright images presented in the flicker paradigm, they replicated an advantage for detection of central changes. But for inverted images, they failed to show a corresponding reduction in the advantage for central changes, suggesting that search was not guided by the contextual meaning of scenes in flicker tasks. This null finding by Shore and Klein suggests that flickering images may be processed differently than simultaneous images. More specifically, subjects may rely more heavily on detection of low-level visual transients in the flicker paradigm, decreasing the reliance on scene meaning (context) to guide orienting. Low-level transients do not exist in the simultaneous paradigm, so attention is guided by scene meaning and endogenous orienting mechanisms. According to this hypothesis, the advantage for central changes in the flicker paradigm may have been due to differences in low-level visual properties between the scenes that contained central changes versus the scenes that contained peripheral changes.

Thus, the goal of our study is to further test whether scene inversion affects context-guided change detection in the flicker task. To address potential imbalances in the low-level discriminability of changes tested in prior studies, we employed a new set of images that were more explicitly controlled. Rensink et al. (1997) equated brightness, color, and size between central changes and marginal changes. However, central and marginal changes still occurred across different images, which may have contributed some variance. To minimize such variance, our study attempts to equate the salience of central and marginal changes within

the same images. In other words, each image contained two competing changes: one change was central to the context of the scene and the other change was marginal. Subjects were instructed to report whatever change they detected first. We expected that for two objects so matched, the one with greater significance given the context of the scene would be noticed more often.

To further isolate the effects of context and to show that the competing changes were matched in visual salience, we also employed a scene inversion manipulation (Shore & Klein, 2000). In one half of our trials, scenes were presented upright, and in the other half, they were presented upside down. Inverting the orientation of the scenes should hinder their recognition and the effects of context (Intraub, 1984; Klein, 1982; Rock, 1974). Thus, in our task, if contextual significance influences the frequency with which one item is attended over another, then inverting the scene should reduce or eliminate that difference. We anticipated that our efforts to minimize differences in visual salience within scenes would allow us to detect the scene inversion effects that were not observed in Shore and Klein's flicker experiment.

Methods

Participants

Fifteen subjects were used in the pilot phase, and a different group of 34 subjects participated in the main experiment. All were college students, aged 18 through 22 years, with normal or corrected-to-normal vision. Subjects were recruited to take part in the experiment in exchange for course credit. Informed consent was obtained after the nature of the experiment had been explained. The research followed the tenets of the World Medical Association Declaration of Helsinki, and the procedures were approved by the Vanderbilt University Institutional Review Board. Informed consent was obtained from the subjects after explanation of the nature and possible consequences of the study.

Stimuli and apparatus

The experiment was programmed and executed using MATLAB 5.2.1, using the Psychophysics Toolbox extensions (Brainard, 1997; Pelli, 1997). Displays were presented at a resolution of 640 pixels \times 480 pixels on a 15-in. iMac monitor.² Scenes measured between 170 and 450 pixels in height and 268 and 587 pixels in width. Subjects sat at a distance of 18 to 24 in. from the monitor.

Twenty-one images were generated with two changes in each. Changes were differentiated by the experimenters as having either high contextual significance (hi) or low contextual significance (lo). The changes were generated by modifying some detail of the scene, either by changing



Figure 1. Two examples of scenes containing competing changes. The first row illustrates a disappearance change (the clown's spot and the pig's face). The second row illustrates an object color change (the ladder and the satellite arm).

its color or by removing it from the scene entirely, using Adobe Photoshop 5.0.2 software. The same manipulation was used for both changes within a scene; for example, both target objects could disappear in a scene. If the color change manipulation was used, objects were matched for color before and after the change (see Figure 1 for an example). The items that were altered ranged in height and width from 10 pixels to 50 pixels. Within scenes, the competing modifications were matched as carefully as possible with respect to size, color, eccentricity from the center, and background contrast.

Change-Detection Procedure

The task was to quickly detect a change between two cycling images. Images were displayed on-screen in the following cycle: the unchanged image was presented for 240 ms, followed by an 80 ms blank phase, then the altered image was presented for 240 ms, followed by another 80 ms blank. This sequence made the target

objects appear to change back and forth between their standard and altered states. This cycle was repeated until the subject responded or until the trial timed out (120 s). Subjects responded by first pressing the space bar to indicate that a change had been noticed. This stopped the cycle and brought the unchanged (standard) image onto the screen. To ensure accurate responses, the subject then indicated what aspect of the scene had been altered by using the computer mouse to click a 50 pixel x 50 pixel transparent outline box cursor over the changing portion of the image. Reaction time was measured using the time lapsed between the trial onset and the subject's first (key-press) response.

Pilot phase procedure

The pilot phase served to measure the selection rate for high contextual significance (hi) and low contextual significance (lo) changes (as determined by the experimenters) within scenes to be used in the

experiment. Prior to initiation of the experiment, subjects were presented with written instructions on the screen and were also given a verbal description of the task. The experimenter then observed the subject in the completion of two practice trials, after which the actual task proceeded. For each trial, the subject was directed to respond as soon as a change was noticed and then to box in the observed change with the computer mouse. The data were analyzed to see if a preference existed for selecting one type of change (hi or lo) more often than the other. For each image, we redefined the hi change to be the one that the majority of subjects detected (the other item was defined as the lo change). This performance-dependent measure allowed us to objectively define which of the two changes was more salient. Because we attempted to equate the visual salience of the two changes, we are assuming that the detectability of each change reflects its contextual relevance to the overall scene. However, to the extent that our visual manipulations were not perfect, low-level feature differences may have contributed as well. Nevertheless, our scene inversion manipulation will rule out any confounds due to differences in low-level visual salience.

Experimental phase procedure

In the experimental phase, two of the images from the pilot phase were excluded because subjects showed poor accuracy (53% and 67%) in correctly identifying either of the two possible changes. One image was excluded because the difference in the rate of selection between the two changes was deemed to be too small (53% vs. 47%). The remaining images were divided into two groups of nine each. For each subject, one group of images was randomly selected to be shown upright; the other was shown inverted. The images used in each orientation condition were counterbalanced across subjects.

Subjects were instructed as in the pilot phase with the additional warning that some of the images would appear upside down, and that these images were to be treated as normal trials. The procedure was otherwise identical to the pilot phase.

The two types of images (upright and inverted) were presented to each subject in a randomly intermixed manner. The upright orientation condition served as a control condition and should replicate the preference for detecting hi changes, as measured in the pilot phase. The inverted orientation condition was predicted to disrupt global context information, reducing the preference for detecting hi changes.

Results

The primary result is the difference between the rate of selection for the hi context items in the upright versus inverted orientations. During the experimental phase of this study, subjects selected the hi context item in 81% of

all displays presented in their upright orientation. For the inverted orientation, the preference dropped to 69% [$t(33) = 2.936, p = .006$]. The significant difference in the hi selection rate suggests that contextual information directed attention even when visual features were equated. Mean response times for detecting changes were slower for the inverted condition than the upright condition ($M = 8.8$ s vs. 8.0 s), but the increase was not significant.

Discussion and Conclusions

These results extend earlier work by [Rensink et al. \(1997\)](#) and [Shore and Klein \(2000\)](#) to confirm that changes in a scene are more easily noticed when those changes involve objects relevant to the scene's context. To isolate the effects of context while controlling for contributions of low-level visual salience, we inverted the images in one half of the trials, and demonstrated that this reduced the advantage for changes that were of high contextual relevance. In other words, scene inversion reduced the effects of global context on change detection.

It is worthwhile to compare our findings with those of [Shore and Klein's \(2000\)](#) flicker-task experiment that revealed no effect of scene orientation. Their null finding stands in contrast to their first experiment, which demonstrated a robust effect of scene orientation when the images were presented side by side. They attributed the different results to potential differences in how flickering images and simultaneously presented images may be processed. Specifically, Shore and Klein proposed that the flicker paradigm inhibits processing of contextual information contained in the scenes because subjects may utilize a search strategy that focuses on exogenous detection of local visual transients. In contrast, low-level visual transients are not present in the simultaneous paradigm, so change detection is guided by endogenous attention. According to this view, the apparent presence of context effects in the flicker paradigm may have been due to potential imbalances in the relative visibility of central versus marginal changes in the stimuli set used by [Rensink et al. \(1997\)](#). It is important to note that Shore and Klein did not suggest that scene context should never affect change detection in flicker tasks.

By using competing objects within the same scene that were matched for color, size, eccentricity, and luminance, it was possible for us to minimize the effects of bottom-up visual salience between central and marginal changes. Using such controlled stimuli, we not only demonstrated a robust context effect, we further established that context effects are significantly disrupted by scene inversion. Thus, we have extended earlier work to confirm that scene meaning does influence change-detection performance of targets in flicker tasks. More broadly, our study further highlights the general

importance of top-down contextual information in viewing of natural scenes.

Acknowledgments

The research was supported by National Science Foundation Grant BCS-0096178. We thank Jenny Lee for assistance in running subjects. Commercial relationships: None.

Footnotes

1. Before the experiment, the scenes were shown to a group of naïve volunteers who were asked to give verbal descriptions of the scenes. These descriptions were the basis for determining items that were of central interest and those that were of marginal interest.
2. Three subjects were tested on Macintosh G4 machines on 16-in. screens. They were tested at the same screen resolution, resulting in somewhat larger images. No difference was observed in the performance of these subjects.

References

- Biederman, I., Mezzanotte, R. J., & Rabinowitz, J. C. (1982). Scene perception: Detecting and judging objects undergoing relational violations. *Cognitive Psychology*, *14*, 143-177. [PubMed]
- Boyce, S. J., Pollatsek, A., & Rayner, K. (1989). Effect of background information on object identification. *Journal of Experimental Psychology: Human Perception & Performance*, *15*, 556-566. [PubMed]
- Brainard, D. H. (1997). The Psychophysics Toolbox. *Spatial Vision*, *10*, 433-436. [PubMed]
- Bravo, M. J., & Nakayama, K. (1992). The role of attention in different visual-search tasks. *Perception & Psychophysics*, *51*, 465-472. [PubMed]
- Chun, M. M. (2000). Contextual cueing of visual attention. *Trends in Cognitive Science*, *4*, 170-178.
- Chun, M. M., & Jiang, Y. (1998). Contextual cueing: Implicit learning and memory of visual context guides spatial attention. *Cognitive Psychology*, *36*, 28-71. [PubMed]
- Chun, M. M., & Phelps, E. A. (1999). Memory deficits for implicit contextual information in amnesic subjects with hippocampal damage. *Nature Neuroscience*, *2*, 844-847. [PubMed]
- Grimes, J. (1996). On the failure to detect changes in scenes across saccades. In A. A. Kathleen (Ed.), *Perception: Vancouver studies in cognitive science* (Vol. 5, pp. 89-110). New York: Oxford University Press.
- Henderson, J. M., & Hollingworth, A. (1999). The role of fixation position in detecting scene changes across saccades. *Psychological Science*, *10*, 438-443.
- Hollingworth, A., & Henderson, J. M. (1998). Does consistent scene context facilitate object perception? *Journal of Experimental Psychology: General*, *127*, 398-415. [PubMed]
- Intraub, H. (1984). Conceptual masking: The effects of subsequent visual events on memory for pictures. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *10*, 115-125. [PubMed]
- Klein, R. (1982). Patterns of perceived similarity cannot be generalized from long to short exposure durations and vice versa. *Perception & Psychophysics*, *32*, 15-18. [PubMed]
- McConkie, G. W., & Currie, C. B. (1996). Visual stability across saccades while viewing complex pictures. *Journal of Experimental Psychology: Human Perception & Performance*, *22*, 563-581. [PubMed]
- O'Regan, J. K., Rensink, R. A., & Clark, J. J. (1999). Change-blindness as a result of 'mudsplashes.' *Nature*, *398*, 34. [PubMed]
- Palmer, S. E. (1975). The effects of contextual scenes on the identification of objects. *Memory and Cognition*, *3*, 519-526.
- Pelli, Denis G. (1997). The VideoToolbox software for visual psychophysics: Transforming numbers into movies. *Spatial Vision*, *10*, 437-442. [PubMed]
- 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.
- Rock, I. (1974). The perception of disoriented figures. *Scientific American*, *230*, 78-85. [PubMed]
- Shore, D. I., & Klein, R. M. (2000). The effects of scene inversion on change blindness. *Journal of General Psychology*, *127*, 27-43. [PubMed]
- Simons, D., & Levin, D. (1997). Change blindness. *Trends in Cognitive Science*, *1*, 261-267.
- Treisman, A. M., & Gelade, G. (1980). A feature-integration theory of attention. *Cognitive Psychology*, *12*, 97-136. [PubMed]
- Wolfe, J. M. (1994). Guided Search 2.0: A revised model of guided search. *Psychonomic Bulletin & Review*, *1*, 202-238.