Scene structure enhances change detection

D. Alexander Varakin a; Daniel T. Levin b

a Knox College, Galesburg, IL, USA
b Vanderbilt University, Nashville, TN, USA

First Published on: 13 November 2007
To link to this article: DOI: 10.1080/17470210701774176
URL: http://dx.doi.org/10.1080/17470210701774176
Short article

Scene structure enhances change detection

D. Alexander Varakin
Knox College, Galesburg, IL, USA

Daniel T. Levin
Vanderbilt University, Nashville, TN, USA

Theories of objects recognition, scene perception, and neural representation of scenes imply that jumbling a coherent scene should reduce change detection. However, evidence from the change detection literature questions whether jumbling affects change detection. The experiments reported here demonstrate that jumbling does, in fact, reduce change detection. In Experiments 1 and 2, change detection was better for normal scenes than for jumbled scenes. In Experiment 3, inversion failed to interfere with change detection, demonstrating that the disruption of surface and object continuity inherent to jumbling is responsible for reduced change detection. These findings provide a crucial commonality between change detection research and theories of scene perception and neural representation. We also discuss why previous research may have failed to find effects of jumbling.

Keywords: Change blindness; Change detection; Scene memory; Scene organization; Scene representation; Visual memory.

When people view a realistic scene, they often believe that they are aware of a large proportion of the visual features that make up the scene. However, observers often fail to detect changes that occur across views, a phenomenon called change blindness (CB; for review, see Rensink, 2002). Many demonstrations of CB utilize the flicker paradigm in which two versions of a scene are presented cyclically with a blank screen or flicker inserted in between each presentation (Rensink, O'Regan, & Clark, 1997). It can take a surprisingly long time for a change to be detected in these circumstances (over 50 s in some cases; Rensink et al., 1997). However, changes involving objects that are central to the meaning of a scene are detected faster than marginal interest changes (O'Regan, Rensink, & Clark, 1999; Rensink et al., 1997). This central interest effect demonstrates that change detection performance is affected by the semantic relationship between the changing object and the scene's structure (see also Hollingworth & Henderson, 2000). The current experiments address whether organized scene structure, by itself, can facilitate change detection performance.

Before describing this research, we would like to define what we mean by scene structure. As used here, scene structure refers to the perceptual...
features that support an understanding of the relative location of objects (and other features) and overall spatial structure of the depicted scene. The most important parts of this structure are the (usually) immobile surfaces that support objects, confine locomotion, and provide cues for navigation (were one actually in the depicted scene). This definition closely follows that given by Henderson and Hollingworth (1999). A critical point to emphasize is that their definition includes reference to the semantic coherence of the scene, whereas ours does not. Although semantic coherence is important for scene perception, we do not include it in our definition of scene structure primarily because scene-responsive brain regions (e.g., the parahippocampal place area, PPA) respond minimally to semantic qualities of the scene such as familiarity (Epstein, 2005). Thus, we focus primarily on how the perceptual structure of a scene might affect change detection.

Various lines of research are consistent with the idea that scene structure affects how a scene is perceived and represented and might therefore be expected to affect change detection. For example, people can identify objects more accurately in coherently structured scenes than in jumbled scenes (Biederman, 1972), and the abovementioned PPA seems dedicated to representing the local geometry of structured scenes in both perceptual and visual imagery tasks (Epstein & Kanwisher, 1998; O’Craven & Kanwisher, 2000). More recent research has demonstrated that observers take longer to determine spatial relationships of locations that straddle local discontinuities in scene structure than those of locations that do not straddle such discontinuities (Sanocki, Michelet, Sellers, & Reynolds, 2006). All together, this research implies that scene structure should also affect change detection, because detecting changes relies in part on each of these processes.

However, disrupting scene structure does not seem to have very strong effects on change detection, at least in the flicker paradigm. In experiments by Yokosawa and Mitsumatsu (2003), participants searched for changes in scenes that were divided into 24-grid sections. The experimenters manipulated scene structure in two ways: (a) by removing some of the grid sections and (b) by jumbling the grid sections (i.e., moving grid sections out of their original positions). In an initial experiment directly comparing normal and jumbled scenes, jumbling had no effect at all on reaction time to detect changes. In a second experiment, Yokosawa and Mitsumatsu tested for an interaction between the proportion of grid squares present (so, some sections contained scene information, and some were blank) and scene configuration (e.g., normal vs. jumbled). Removal of grid sections had a large effect; observers generally took longer to find changes when there were more visible grid sections. However, under most conditions, scene configuration did not interact with this RT effect. The only exception was that RTs increased between 70%-complete scenes and 100%-complete scenes for jumbled scenes, but not for normal scenes (although no interaction between scene configuration and percentage of completion was reported). Yokosawa and Mitsumatsu took this as evidence that normal scenes afforded a perceptual efficiency that allowed subjects to avoid a search-time cost for the more feature-filled complete scenes as compared to the incomplete (and therefore less feature-filled) scenes. Because this single indirect finding of a configuration effect was isolated among many failures to find a configuration effect in more direct comparisons, the authors concluded that “the disruption of a scene by jumbling and elimination hardly impaired change detection” (p. 47). The finding that jumbling a scene affects change detection so little is surprising, given the presumed importance of scene structure in specifying a scene’s spatial layout, meaning, and neural representation.

The current experiments expand on the results of Yokosawa and Mitsumatsu (2003) by testing whether scene structure facilitates change detection using a one-shot change paradigm (cf. Rensink, 2002) instead of the repeated-flicker paradigm used in the previous report. In one-shot studies observers are exposed to the change only once, thus forcing observers to encode and compare as much information as they can in...
single viewing of the pre- and postchange scene. In contrast, the flicker paradigm does not place such heavy demands on visual attention or short-term memory, because observers have the option of serially encoding and comparing small sections of a scene. If jumbling a scene affects the ability to extract a large amount of information from a scene, but does not affect the speed with which observers can serially deploy attention to individual objects or locations, then the effects of jumbling may be more apparent in a one-shot paradigm then the flicker paradigm.

EXPERIMENT 1

In the first experiment, we compared observers’ ability to detect changes to normal scenes and jumbled scenes whose organization had been disrupted by rearranging their parts. Jumbled scenes contain much of the same information as did their normal counterparts; however, their status as a coherent scene is disrupted because of the jumbling.

Method

Participants

A total of 24 members of the Vanderbilt University community participated for cash compensation.

Stimuli

The stimuli consisted of original and changed versions of 60 digital photographs of various indoor and outdoor scenes (close-up and distant shots) and their jumbled counterparts (see Figure 1 for an example). Changed and jumbled scenes were

Figure 1. Example scenes from Experiment 1 (shown in colour online).
created using commercially available software (Adobe Photoshop). To make the changed versions of each scene, the original scenes were sized to be 600 pixels wide and 450 pixels in height and were set at a resolution of 72 pixels per inch. This allowed a six-section grid (3 × 2) with equally sized sections to be created. Each scene contained whole objects in at least four grid sections (following Biederman, 1972). For each original scene, two kinds of changes were made inside one of the grid sections: (a) deletions/additions of a whole objects or (b) changes to the object’s properties (e.g., colour, shape, etc.). Note that objects to which property changes were made were the same objects as those that were deleted/added. To create jumbled versions, the six grid sections were randomly rearranged with the two restrictions: (a) jumbled scenes could not contain two horizontally or vertically adjacent grid sections that were also adjacent in the normal scene, and (b) across the entire set of scenes, changes appeared in each section an equal number of times (the same random jumble was used for each version of a given scene). After being manipulated, all stimuli were resized to 640 pixels wide and 480 pixels in height at a resolution of 28.346 pixels per cm (72 pixels per inch).

Apparatus

Scenes were presented against a white background at the centre of 15-inch monitors set at a resolution of 1,024 × 768 and at 16-bit colour depth (thousands of colours). Each scene subtended approximately 21.3 × 16.1 degrees visual angle at a viewing distance of 60 cm. E-Mac computers controlled the stimulus presentation and recorded responses.

Design and procedure

The primary factors of interest were scene configuration (normal or jumbled) and change type (addition, deletion, or property change). Scene configuration was blocked and counterbalanced across participants. Each block consisted of 30 trials (one trial per scene) and contained an equal number of addition, deletion, and property changes, presented in a random order. Across participants, each scene appeared in each condition and served as both pre- and postchange an equal number of times. Thus, each script consisted of 60 total trials, with 10 trials for each of the six conditions.

For each trial in both conditions, a prechange scene was presented for 2,000 ms, followed by a 100-ms blank screen and then the postchange scene. The postchange scene was visible for 1,000 ms before a question appeared at the bottom of the screen asking participants to press “Y” if they detected a change and to press “N” if they did not. Since there was a chance on every trial, if a participant pressed “Y”, another prompt appeared asking the participant to use the mouse cursor to click at the location of the change. If “N” was pressed, the next trial was initiated. There was a 500-ms intertrial interval. Participants were not given feedback as to the accuracy of their localizations.

Results and discussion

A trial was considered a hit (i.e., a claim of change detection with accurate localization) if the participant responded “yes” to the change question and clicked within a rectangle with an area that was 30.6% larger than the smallest rectangle that could surround the changed region (for this and all subsequent experiments).

A 2 (scene configuration) × 3 (change type) repeated measures analysis of variance (ANOVA) was used to analyse the data. Significance was evaluated using the lower bound degrees of freedom adjustment. Generalized eta squared ($\eta^2_G$) was used to measure effect size, because these estimates are reported to be comparable across a variety of research designs (Olejnik & Algina, 2003). There was a main effect of scene configuration, $F(1, 23) = 7.510, M^2SE = 0.017, \rho = .012, \eta^2_G = .06$. More changes were detected in normal scenes ($M = 26%$) than in jumbled scenes ($M = 20%$; see Figure 2). There was also a main effect of change type, $F(1, 23) = 5.808, M^2SE = 0.041, \rho = .024, \eta^2_G = .11$. Post hoc tests (LSD) revealed that property changes
were detected less often than additions ($M = 25\%$) or deletions ($M = 27\%$), $t(23) > 2.63$, $ps < .02$, and that additions and deletions did not differ, $t(23) < 1$. The interaction between scene configuration and change type was not significant, $F(1, 23) = 0.52$, $MSE = 0.03$, $\eta^2 = .008$.

These results clearly demonstrate that jumbling a scene reduces change detection, and they also replicate the finding that changes involving the global layout of a scene are detected better than changes that do not (see also Aginsky & Tarr, 2000; Simons, 1996).

**EXPERIMENT 2**

In Experiment 2 we tested the hypothesis that the jumbling effect is caused by terminators. This is necessary because it is possible that the large number of unnatural terminations inherent to jumbling cause a high level of artificial perceptual noise that draws attention from the changing objects. Therefore, in the present scenes, an occluding black border divided the jumbled sections from each other. This produces the appearance of a scene that is being viewed through a window. These “windowed” jumbled scenes were compared with windowed normal scenes (see Figure 3), which had the same number of terminators and a similar perceptual grid as jumbled scenes. Normal scenes without windows were also included. If the grid and simple presence of terminators causes the jumbling effect, then there should be no difference in change detection between the normal windowed scenes and the jumbled windowed scenes, and both should show significantly less change detection than normal scenes. In contrast, if the impact of jumbling on change detection goes beyond interference from terminators, then change detection should be lower in window jumbled scenes than in window normal scenes. We chose not to include a

Figure 2. Percentage change detection by scene configuration from Experiments 1, 2, and 3 (E1, E2, and E3, respectively). Error bars represent the standard error of the mean.
nonwindowed jumbled condition because of our finite set of scenes and because we were more interested in testing the effect of windowing in normal scenes than in jumbled scenes.

Method

Participants
A total of 36 members of the Vanderbilt University community participated for course credit.

Stimuli
Stimuli consisted of 54 of the 60 scenes that were used in Experiment 1. The 6 scenes for which change detection was worst in Experiment 1 were excluded. In addition, frames were added to normal scenes (see Figure 3). The frames were similar to the frames used with the jumbled scenes, except that frames used with normal scenes were drawn over the pictures, and so they occluded some parts of the image. This was done so that the continuity of portions of the images that crossed grid sections was not disrupted.

Design and procedure
For this and subsequent experiments, the design and procedure were identical to those of Experiment 1 except as noted. The factors of interest in Experiment 2 were scene configuration (normal, windowed normal, windowed jumbled) and change type (addition, deletion, or property change). As in Experiment 1, scene configuration was blocked and counterbalanced across participants. There were 18 trials per scene configuration block, each with 6 trials per change type presented in a random order.

Results and discussion
A 3 (scene configuration) × 3 (change type) repeated measures ANOVA was used to analyse
hits. As in Experiment 1, there were main effects of scene configuration, $F(1, 35) = 7.610$, $MSE = 0.054$, $p < .01$, $\eta^2_G = .05$, and change type, $F(1, 35) = 18.34$, $MSE = 0.074$, $p < .001$, $\eta^2_G = .15$, and no significant interaction, $F(1, 35) = 0.718$, $MSE = 0.093$, $\eta^2_G = .008$. More changes were detected in normal scenes ($M = 29$%) and windowed normal scenes ($M = 26$%) than in windowed jumbled scenes ($M = 21$%), $t(36) > 2.68$, $ps < .05$, but there was no difference between normal and windowed normal scenes, $t(36) = 1.13$, $p = .252$; see Figure 2. Also replicating Experiment 1, detection of additions ($M = 29$%) was equivalent to detection of deletions ($M = 31$%), $t(36) < 1$, while both additions and deletions were detected more often than property changes ($M = 16$%), $t(36) > 4.37$, $ps < .001$.

The results of Experiment 2 suggest that the number of terminators per se does not account for the jumbling effect. In the windowed normal scenes, change detection was more accurate than that in jumbled scenes, even though both had the same number of terminators. Thus, these results are consistent with the hypothesis that scene structure facilitates change detection.

**EXPERIMENT 3**

If the disruption of scene structure is what caused the jumbling effect in Experiments 1 and 2, then scene inversion should not reduce change detection relative to normal scenes, as scene inversion preserves the properties that define scene structure as we have defined it (see also Kelley, Chun, & Chua, 2003; Shore & Klein, 2000). For example, in inverted scenes objects still occlude each other in mostly natural ways, and foreground and background objects and surfaces obey the same grouping principles as in normal scenes.

**Method**

**Participants**

A total of 25 members of the Vanderbilt University community participated for course credit.

**Stimuli**

Stimuli consisted of the 60 scenes from Experiment 1.

**Design**

The factors of interest in Experiment 3 were scene orientation (normal or inverted) and change type (addition, deletion, or property change). Scene orientation was blocked and counterbalanced across participants. Each block contained 30 trials, with equal numbers of the three change types presented in a random order.

**Results**

A 2 (scene orientation) $\times$ 3 (change type) repeated measures ANOVA was used to analyse verified hits. The finding of most interest is that change detection was equivalent for normal ($M = 33$%) and inverted scenes ($M = 34$%); $F(1, 24) < 1$, $MSE = 0.021$, $\eta^2_G < .001$; see Figure 2. As in previous experiments, there was a main effect of change type, $F(1, 24) = 14.289$, $MSE = 0.063$, $p = .001$, $\eta^2_G = .226$, and no significant interaction, $F(1, 24) = 1$, $MSE = 0.045$, $\eta^2_G = .014$. Property changes ($M = 22$%) were detected less often than deletions ($M = 38$%) and additions ($M = 39$%), $t(24) > 4.4$, $ps < .001$, but deletions and additions did not differ, $t(24) < 1$. Change detection was not affected by inversion, suggesting that the jumbling effect in Experiments 1 and 2 was caused by disruption of perceptual scene structure.

**GENERAL DISCUSSION**

The experiments reported here have repeatedly demonstrated that scene structure affects change detection. In Experiments 1 and 2, jumbling a scene reduced change detection, and in Experiment 2, adding frames to normal scenes to equate the number of terminators between normal and jumbled scenes did not eliminate the jumbling effect. Experiment 3 demonstrated that the normal-scene advantage is probably due to the organizing perceptual-features characteristic
of normal scenes, because inversion did not reduce change detection. Furthermore, in all experiments, jumbling had similar effects for property changes and for additions and deletions, in that detection of property changes was always worse than detection of additions and deletions (for more results relevant to how change type affects change detection see Aginsky & Tarr, 2000; Simons, 1996).

Why, then, did jumbling of scene structure reduce change detection? Our experiments suggest that the key proximal cause was the disruption of the surfaces and layout that constitute scene structure as we have defined it. There are two general mechanisms by which such jumbling may have reduced overall change detection. First, jumbling may effectively increase the amount of information that a scene contains, thus reducing the probability that the changing object will be represented. This idea is similar to the idea that visual short-term memory (VSTM) has a fixed capacity that is determined by the number of objects, regardless of the complexity of each individual object (e.g., Vogel, Woodman, & Luck, 2001, but see Alvarez & Cavanagh, 2004; Olson & Jiang, 2002). For example, in many of our scenes, large objects straddled grid boundaries and appeared as object parts in jumbled scenes (e.g., the white board in Figure 1). These divided objects may have required only one slot in VSTM in normal scenes and two (or more) slots for jumbled scenes. The second mechanism is that jumbling may have disrupted observers’ ability to create a large chunk of visual information that could have been compared with the postchange scene with sufficient certainty about object properties and their locations to confidently isolate a specific change. The basic difference between these two mechanisms is that the first attributes the jumbling effect to the increased number of perceived objects in jumbled scenes without appealing to changes in the overall amount or quality of information that is represented. In contrast, the second mechanism is that either the amount or the quality of information that observers can represent increases due to the organizational features characteristic of normal scenes. It is worth noting that these hypotheses are not mutually exclusive and that both could have contributed to the jumbling effect observed here. However, the current experiments cannot determine the extent to which either of these mechanisms contributed to the jumbling effect.

Why did these experiments repeatedly find an effect of jumbling whereas Yokusawa and Mitsumatsu’s (2003) experiments did not? In the Introduction we provided reasons why their choice of task (the flicker paradigm) may not be ideally suited for uncovering effects of jumbling on change detection. Additionally, Yokusawa and Mitsumatsu may have failed to find the jumbling effect because the effect is not very large (as indexed by \( \eta^2_g = .06 \) in Experiment 1 and .05 in Experiment 2). To illustrate, if we treat the data from our Experiment 1 as a within-groups \( t \) test (ignoring change type for purposes of illustration), we find a significant mean difference of 6% (\( SD_{\text{diff}} = 11\% \); \( t(23) = 2.74, p = .01 \), between normal and jumbled scenes. Observed power in this situation is .75. Assuming a mean difference of 6% (\( SD_{\text{diff}} = 11\% \)), power is .57 at a sample size of 17 (the sample size in Experiment 1 from Yokusawa and Mitsumatsu), and power is only .27 at a sample size of 9 (the sample size in Experiment 3 from Yokusawa and Mitsumatsu).

In conclusion, the experiments reported here demonstrated that scene structure facilitates change detection. This finding confirms that scene structure constrains our ability to represent and compare information from one view to the next and provides a crucial link between the change detection and research on how scene structure affects object recognition, neural representations, and scene perception.

Original manuscript received 16 August 2007
Accepted revision received 4 October 2007
First published online 13 November 2007

**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.


