Vision Research 48 (2008) 2275–2283

Contents lists available at ScienceDirect

Vision Research

journal homepage: www.elsevier.com/locate/visres

The role of gist in scene recognition

Anthony Chad Sampanes, Philip Tseng, Bruce Bridgeman *

Department of Psychology, University of California, Santa Cruz, Social Sciences 2, Santa Cruz, CA 95064, USA

ARTICLE INFO

Article history:
Received 25 July 2007
Received in revised form 3 July 2008

Keywords:
Change blindness
Gist
Eye movements
Attention
Scene perception

ABSTRACT

Studies of change blindness suggest that we bring only a few attended features of a scene, plus a gist, from one visual fixation to the next. We examine the role of gist by substituting an original image with a second image in which a substitution of one object changes the gist, compared with a third image in which a substitution of that object does not change the gist. Small perceptual changes that affect gist were more rapidly detected than perceptual changes that do not affect gist. When the images were scrambled to remove meaning, this difference disappeared for seven of the nine sets, indicating that gist and not image features dominated the result. In a final experiment a natural image was masked with an 8×8 checker pattern, and progressively substituted by squares of a new natural image of the same gist. Spatial jitter prevented fixation on the same square for the sequence of 12 changes. Observers detected a change in an average of 2.1 out of 7 sequences, indicating strong change blindness for images of the same gist but completely different local features. We conclude that gist is automatically encoded, separately from specific features.

1. Introduction

Contrary to our intuitions, visual input is far from highly detailed and stable. Because only information falling on the fovea is detailed, only successive fixations can offer high-resolution information throughout a visual scene (Coltheart, 1999; Henderson & Hollingworth, 1998). A controversy has emerged, however, about how much information is carried over from one visual fixation to the next (Bridgeman & Mayer, 1983; Gilchrist & Harvey, 2000; Grimes, 1996; Henderson, 1997; Irwin, 1991, 2003; O'Regan, 1992; Pollatsek & Rayner, 1992) and to what degree visual memory is impoverished (Brockmole & Irwin, 2005; Irwin, 2003; McConkie & Currie, 1996).

Transsaccadic memory has been generalized to successive views interrupted by a brief blank, and the resulting memory has been tested with the phenomenon of change blindness (CB), a failure to detect striking changes in a scene that occur across brief visual disruptions. Without a disruption such changes are easily noticeable. Standard CB studies use a "flicker" paradigm; an image (A) alternates cyclically with an altered image (A′), with a brief white or gray screen (lasting at least 80 ms) interleaved between images. The dependent measure is response time to detect change accurately. Observers almost never detect the change on the first cycle and sometimes fail to detect it after 1 min of continuous flicker (Rensink, O'Regan, & Clark, 2000). The explanation is that the onset and offset of the images creates a global transient signal that delocalizes the local transient accompanying the change. Hence, attention is not attracted to the change location (Rensink, O'Regan, & Clark, 1997; Simons & Levin, 1997). Reports of detection may be delayed, however, so that the time of report is in a later sample than the sample during which detection takes place.

CB is a robust effect (reviewed by Rensink, 2002); changes can escape detection if they occur during a saccadic eye movement (Bridgeman, Hendry, & Stark, 1975; Grimes, 1996; Henderson & Hollingworth, 1999), eye blink (O'Regan, Deubel, Clark, & Rensink, 2000), or camera cut (Simons, 1996). CB is not limited to video displays. It can arise during real-world interactions, where a conversation partner is replaced by a different individual (Levin, Simons, Angelone, & Chabris, 2002; Simons & Levin, 1998).

CB can also take place when concomitant objects (e.g., "mudsplashes") are superimposed on a scene without covering the altering object (O'Regan, Rensink, & Clark, 1999). Neither is it limited to an abrupt transient (either by a blank screen, eye blink, or saccade). Changes can occur gradually (Simons, Franconeri, & Reimer, 2000), too slowly to attract attention. A faster, but still smooth change paradigm ramps a scene from normal contrast to zero contrast in 1 s and immediately ramps back to full contrast in another second (Turatto, Bettella, Umilta', & Bridgeman, 2003). The target feature changes during the instantaneous (10 ms) zero-contrast sample, and the original and altered images alternate. CB is just as strong as in the conventional flicker paradigm. Using a modified flicker paradigm, Hollingworth and Henderson (2004) rotated a scene

* Corresponding author.
E-mail addresses: acsampanes@yahoo.com (A.C. Sampanes), ptseng@ucsc.edu (P. Tseng), bruceb@ucsc.edu (B. Bridgeman).

0042-6989/$ - see front matter © 2008 Elsevier Ltd. All rights reserved.
doi:10.1016/j.visres.2008.07.011

1° after each target interruption and found that observers tended not to report a change that was obvious if a larger rotation was introduced in a single step, implying that visual short-term memory does not "look back" over successive samples. The scene was a rotation of the original, however, and not a completely new image of the same gist.

These findings imply that target visibility in CB is limited not by masking, but by focused attention (Enns & Di Lollo, 1997; Rensink, 2000b; Rensink, O'Regan, & Clark, 1997). While focused attention may be necessary for change detection, it certainly is not sufficient (Williams & Simons, 2000). Ballard, Hayhoe, and Pelz (1995) created a task where participants took blocks from a stockpile and duplicated the model in a workspace. During the task, undetected changes were made to the model. Because the goal of the task was to duplicate the model, it had surely been the subject of focused attention. Simons and colleagues have also been able to make undetected changes to objects of focused attention, when the central actor was replaced by a different actor during a movie cut (Levin & Simons, 1997), when a box replaced a soda bottle (which had been a central object), when a camera panned away during a movie (Simons, 1996), and when a conversation partner was replaced by a different person following a brief visual obstruction (Simons & Levin, 1998).

At first glance the CB results, indicating volatile visual representations with limited capacity, seem to contradict earlier results demonstrating robust and extensive memory for images (Intraub, 1981; Potter & Levy, 1969; Shepard, 1967; Standing, 1973; Standing, Conezio, & Haber, 1970). Though the degree of memory for visual detail is controversial, the gist of an image is clearly maintained. Gist is the general visual category of a scene, generally describable in a few words. It is unclear whether participants in early image memory studies were using memory for visual details, memory for gist (conceptual information specific to the image), or verbal encoding to discriminate old from new images. Discriminations typically involved images and distracters that were very different from one another in visual detail and gist; verbal encoding was almost always uncontrolled. Results in these early studies could be due to gist encoding or verbal encoding.

Detailed memory of visual properties is actually quite poor (Mandler & Ritchey, 1977; Pezdek et al., 1988). Participants are unable to detect when a photograph has been mirror-reversed (Standing et al., 1970). Detail memory for a US penny is also poor (Nickerson & Adams, 1979).

There is evidence that gist (including information about schemas and local and global features) exhibits automatized encoding (Friedman, 1979). It can be determined within 120 ms of presentation (Biederman, 1981; Intraub, 1981; Potter & Levy, 1969). Feature recognition is easier when given external structure; for example, object identification is speeded in a coherent scene (Biederman, 1972). When global properties are stressed, schematic drawings are identified faster than detailed images (Ryan & Schwartz, 1956) and global features are recognized better than figural features (Palmer, 1977).

Schemas seem to play a role in identifying gist (Biederman, Mezzanotte, & Rabinowitz, 1982) and subsequent scene memory (Brewer & Treyans, 1981). For example, people often remember details from a scene that were never presented, for instance in "boundary extension" (Brewer & Treyans, 1981; Intraub, 1999; Intraub, Bender, & Mangels, 1992; Intraub & Richardson, 1989): when the top of an image of a car is cropped (cropping the top of the car), participants remember seeing the entire car. Schematic, higher order information may inhibit encoding of lower order information necessary for change detection (Johnston & Hawley, 1994; Pezdek et al., 1988).

CB might be affected by the meaning of an object change. When changed objects are semantically inconsistent (informative), detection latency is shorter than when they are consistent (non-informative) (Biederman,1972; Hollingworth & Henderson, 2000). For example, the appearance of a fire hydrant in a living room was detected more quickly than the appearance of a chair. These results do not conclusively demonstrate that a change in gist is more readily detected than a change that does not change gist, however. Even if an object has meaning relevant to gist, changing that object does not necessarily change the gist (the meaning of the whole image). There must be some measure of gist itself, not just a variable that contributes to gist.

CB and image memory findings lead to the conclusion that gist plays a major role in representation and recognition of visual scenes. The present consensus is that very little is carried over from one visual fixation to the next, perhaps only 3–4 prominent objects that been attended previously (Cowan, 2000; Irwin & Zelinsky, 2002; Luck & Vogel 1997; Vogel, Woodman, & Luck, 2001), some information about spatial configuration (Biederman, 1981; Rensink, 2000a; Sancocki & Epstein, 1997; Simons, 1996), and the gist (general subject) (Becker & Pashler, 2002; Irwin & Andrews, 1996; Simons & Levin, 1997; Wolfe, 1998, 1999). The first two aspects of the consensus have been well studied, but the third is less well understood. If gist is indeed as critical as this work suggests, the visual system might simply assume that a scene is unchanged as long as its gist remains constant. Consistent gist, then, might mask changes.

In order to separate gist change from geometric change in images, we compare detection of changes in image pairs that differ in gist with changes in image pairs that maintain a constant gist.

2. Experiment 1

The first step in evaluating the detectability of gist vs. non-gist changes is to generate sets of images of constant gist, with alternate forms of differing gist but similar geometric similarity. In order to manipulate gist it is necessary to ascertain the degree of change in gist that a given image modification offers. In this preparatory experiment we exploit a procedure introduced by van Montfort (2007) to detect gist change. It consists of three steps: (1) obtain gist descriptions of the images; (2) have a naïve rater classify each description with the image(s) it fits with; and (3) analyze the fit patterns. The results of this procedure should allow us to change gist between the experimental images and the respective original images, while maintaining gist in a comparison set of control images.

We use this procedure rather than a pixel-by-pixel change measure because almost all pixels remain unchanged in both the constant gist and the changed gist conditions. All pixels are identical except for the few representing a changed object. The important changes are in the objects, not in the pixel-by-pixel alterations of the images, because people perceive objects, not pixels.

2.1. Method

2.1.1. Participants

Fifty-seven undergraduates (20 men and 37 women, mean age 20.7 years) from the University of California, Santa Cruz volunteered to participate in the experiment to fulfill a course requirement.

2.1.2. Stimuli

Each image originated from a natural color photograph. We used Adobe Photoshop to change each image (A), pairing it with either an image of different gist (A′) or a control image of the same gist (A′′). Images of different gist (A′) were the same as the original image (A) except that one item had been substituted to change the meaning of the scene. For example, a log immediately in the path of a man kayaking down a river (A) was changed into another kayak (A′). Control images of the same gist (A″) were the same as the original image (A) except that one item had been changed that did not change gist. For example, the log in the kayaking image (A) was changed to a rock. Substitutions in both the gist change image (A′) and the control image (A″) were made to the same object.

2.1.3. Design and procedure

The original image, the control image (no gist change), and the experimental image (the image that was labeled as a gist change) were presented to a "generator", who gave a verbal description of what he/she thought the image was about. Participants were told to write the gist (essence or meaning) of the scene in no more than 4–5 words. Each generator viewed only the original image, the control image or the experimental image for a given set of images. In other words, any one generator viewed only one of the three images for each set of images.

The descriptions of the original image and the experimental image, or original image and the control image, were given to a different person, the "rater". The rater had no knowledge of which of the two images was presented to the generator. The rater read each description and determined whether the description was appropriate for each of the two images. There were, therefore, four possible categories: the description was (1) *not* appropriate for image 1 or image 2 (fit-pattern 0,0); (2) appropriate for image 1, but *not* for image 2 (fit-pattern 1,0); (3) *not* appropriate for image 1, but was for image 2 (fit-pattern 0,1); and (4) appropriate for both images (fit-pattern 1,1). This procedure was repeated pairing the original image and the experimental image for nine sets.

2.2. Results

Fit-pattern distributions for the descriptions of all sets of the original image and the experimental image (i.e., gist change) were analyzed using the Fisher–Freeman–Halton exact test (Freeman & Halton, 1951). All 9 sets were significantly different from one another (5 sets $p < .05$, 3 sets $p < .01$, and 1 set $< .001$) in gist.

The fit-pattern distributions for the descriptions of the 9 sets of the original image and the control image were also analyzed using the Fisher–Freeman–Halton exact test. Seven sets of images were not statistically different in gist from one another ($p > .1$), one set of images was marginally significant ($>.1$ $p > .05$), and one set was statistically significantly distinct in gist ($p < .05$).

2.3. Discussion

All images that were thought to differ in gist, did in fact differ in gist. The results for the control images support the claim that these images did not change the gist, with two exceptions. One image, which was thought to leave gist unchanged, actually did change gist a bit (marginally significant). In this image a man in a BBQ scene manipulates some meat with a pair of metal tongs. The tongs change to a metal spatula. It seems that the reason for the marginally significant results is because some participants describing the two images keyed in on the particular cooking utensil being used (i.e., the tongs or the spatula), making specific mention of one or the other and therefore excluding the possibility that a rater could assign the description dually to both images.

The other exception which significantly changed the gist was a scene of two surfers surfing a wave with two onlooking surfers sitting on their boards. The control image changed the surfers on the wave to two boogey boarders. Again, the reason for the change in gist appears to be the specific mention of objects in the scene—what the actors on the wave were riding (in Santa Cruz, a surf town, surfing and boogie boarding are quite distinct). Furthermore, the boards the onlookers are sitting on are somewhat ambiguous. Depending on the context, they could be sitting on surf boards or boogie boards. In the majority of cases, then, our goals for gist change or preservation were met, and described differences did not greatly alter the meanings of the images.

A matter that should be further studied is how to properly define gist. In the present experiment gist was defined as what participants could write in no more than 4–5 words. It was thought that giving little instruction left participants unbiased and best able to define gist according to their personal interpretations.

3. Experiment 2

Most change blindness studies have compared two scenes of the same gist, usually two identical scenes with only a detail changed. When two scenes are different in gist, so typically are the features of the scene. Zelinsky (2003) has argued for a continuum between perceptual changes and category changes, though the two can be distinguished by the amount of change in meaning of a scene. For instance, an ocean scene will usually share more common features with another ocean scene than with a forest scene. Is detection of the change difficult because the changed detail is not perceived in one or another of the samples, or because a constant gist suppresses perception of differences? The goal of Experiment 2 is to answer this question. If constant gist suppresses detection of changes, then a small perceptual (featural) change that strongly affects gist should be more easily detected than a perceptually equivalent change that does not affect gist. However, if gist is not critical to eliciting change blindness, then a small featural change should be difficult to detect whether it changes the gist of the scene or not.

3.1. Method

3.1.1. Participants

Eighty undergraduates (23 men and 57 women, mean age 19.8 years) from the University of California, Santa Cruz volunteered to participate in the experiment to fulfill a course requirement.

3.1.2. Apparatus

Images were displayed on a color CRT monitor 23° high × 32° wide and 1280 × 1024 pixels in resolution scanning at 75 Hz.

3.1.3. Stimuli

The images displayed were those evaluated in the preparatory Experiment 1.

3.1.4. Design

CB was examined with a new design, a 2-image alternation paradigm. Presentations consisted of cycles between two interleaved sets of images. In control trials, one set of images (A) alternated with a second set of unrelated images (B) to create an ABAB sequence (Fig. 1). None of the images changed during the trial. Experimental trials alternated between an original image (A) and a changed image (A′). Trial sequence was determined in two steps. First, either a control or a change trial was selected at random, with both equally probable. Second, for change trials the change occurred randomly in either the first set (ABA′B) or the second set (ABAB′). Then the 4-image sequence was repeated. Each image was displayed for 500 ms with no mask and no blank interval. (The blank interval used in most change blindness paradigms could be eliminated because each image set served as an interruption of the other image set).

Using the flicker paradigm, Rensink, O'Regan, and Clark (2000) have shown that luminance of the blank does not affect the size of the CB effect. This result seems to support a new-object

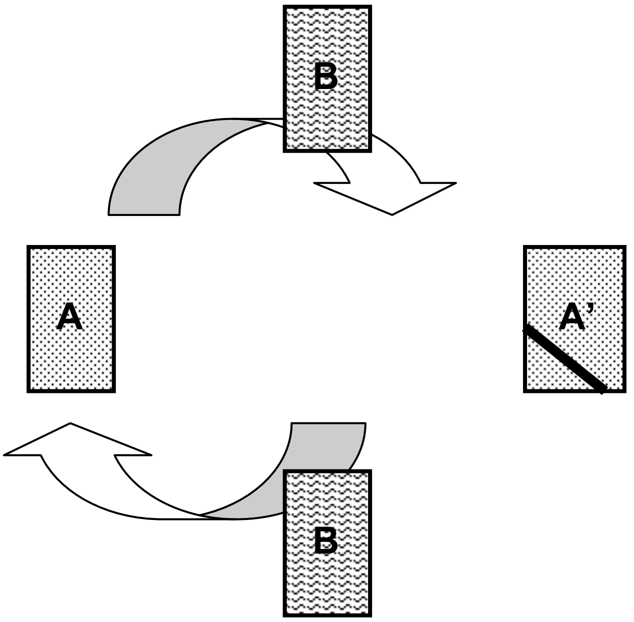

Fig. 1. The 2-image alternation paradigm. Illustrated is an ABA′B experiments sequence. In control sequences, A and A′ are identical.

hypothesis, indicating that the appearance of new elements in the scene and not the visual disruption caused by transient signals affects CB. In our design, this meant that we could use a second image of slightly different brightness to mask the first, and so on through the sequence.

3.1.5. Procedure

Participants sat with their eyes 60 cm from the monitor at eye height, with one finger placed on a key. They were instructed to press the key immediately if they noticed a change in the displayed image, but they were not told anything about the nature of the changes. Participants were free to direct their eye fixations without restriction.

In half of the trials an image was alternated from A to A′, or B to B′; in the other half, each image was present throughout with the temporal parameters given above. Key presses in these trials were scored as false alarms.

3.2. Results

Independent *t*-tests with Bonferonni's correction for multiple tests revealed that, in all nine sets of images, small perceptual changes that affected gist were more rapidly detected than perceptual changes that did not affect gist (see Table 1), though the size of the effect varied from set to set.

Table 1
Detection of featural vs. gist change detection (CD)

Set	Feature CD		Gist CD		*T* value (*df* = 78)	*P* value
	M (s)	*SD* (s)	M (s)	*SD* (s)		
Set 1	34.84	22.85	15.85	12.63	4.60	<.001
Set 2	17.80	18.11	7.15	3.69	3.64	<.001
Set 3	12.88	12.78	7.03	7.15	2.52	<.05
Set 4	8.89	9.36	4.87	2.37	2.63	<.01
Set 5	20.77	22.48	7.70	8.22	3.45	<.001
Set 6	26.84	26.79	5.83	5.85	4.84	<.001
Set 7	12.05	9.79	5.47	6.04	3.63	<.001
Set 8	37.57	21.82	20.59	17.61	3.83	<.001
Set 9	10.41	13.49	5.06	3.14	2.44	<.05

3.3. Discussion

The results indicate that gist has an important role in the encoding of visual information. A change in the gist of a scene will be detected more quickly than an equivalent perceptual change that does not change the gist. The results of Experiment 1 suggest that the differences in the detection rates were not due to confounding perceptual differences, but rather to changes in the meaning of the images.

Here we evaluated perceptual differences between an original image and a control image (no gist change), or the original image and an experimental image (gist change). The conclusion was that gist change is critical in change detection, while alteration of image features is not. The purpose of Experiment 3 is to provide a behavioral measure to test this claim.

4. Experiment 3

In Experiment 2 a gist change image or a control change was coupled with the perceptual changes that occurred in each image. In Experiment 3, by eliminating meaning from both images while maintaining the perceptual changes we attempt to decouple gist from perceptual change. If perceptual changes were matched between the two images when gist is eliminated, one would expect the latency of change detection to be equivalent between the two types of image changes. If one image had perceptual qualities that could more readily elicit change detection then one would expect that image to have more rapid change detection even when gist is eliminated.

4.1. Method

4.1.1. Participants

Fifty undergraduates (15 men and 35 women, mean age 18.5 years) from the University of California, Santa Cruz volunteered to participate in the experiment to fulfill a course requirement.

4.1.2. Stimuli

The control images and experimental images from Experiment 1 were used. Each image was divided into 32 pieces of equal size (a 4×8 grid), and the pieces randomly arranged to form a new image.

4.1.3. Design and procedure

A standard flicker paradigm was used, with a 70 ms ISI; the new image alternated repeatedly with the original image until detection. For each image set, half the subjects viewed the original paired with the control image and half the subjects viewed the original paired with the experimental image. Detection times were compared between subjects for the control images and experimental images within the same set (from Experiment 1).

4.2. Results

Independent *t*-tests comparing the change detection time of the control image and the gist change image from Experiment 1 revealed that change detection times of seven of the nine images were not significantly different from one another or were different from one another in a direction that did not favor detection of gist. (i.e., detection times were significantly longer for gist change images, as in sets 5 and 8 and marginally for set 7). Detection times for one set of images (set 2) were significantly longer for the control image (see Table 2).

Table 2
Detection of featural vs. gist changes from Experiment 3 when gist is eliminated

Set	CD of control pics		CD of experimental pics		*T* value (*df* = 48)	*P* value
	M (s)	*SD* (s)	M (s)	*SD* (s)		
Set 1	7.81	11.28	14.76	17.11	1.70	>.05
Set 2[a]	23.6	20.98	7.21	3.76	3.84	<.001
Set 3	6.71	6.78	4.29	1.49	1.74	>.05
Set 4	3.67	1.98	3.12	1.54	1.03	>.05
Set 5	3.62	2.40	5.92	4.08	2.42	<.05
Set 6	3.12	1.49	3.71	2.72	.96	>.05
Set 7	2.69	.72	3.93	3.04	1.98	>.05
Set 8	2.30	1.00	7.56	8.61	3.04	<.01
Set 9	3.8	2.05	3.62	1.99	.32	>.05

[a] Statistically significant favoring detection of gist more quickly.

4.3. Discussion

The equivalent detection times or detection times that did not favor gist detection in seven of the nine sets of images support the gist assertions from Experiment 1. That is, for most image groups the results from Experiment 3 support the notion that detection of change in Experiment 2 was due to differences in gist rather than any perceptual or featural qualities.

In one set of images detection was significantly faster for the experimental image than the control image. It is likely that the difference in detection time was due to a luminance difference between the images. The change in the experimental image was from a dark gray object to a bright white object. For one further set of images detection time was marginally longer for the control image. This could be due to a slight difference in the orientation of the head of one of the central characters. This orientation change occurred only in the experimental image and not in the control image.

5. Experiment 4

In most previous work on CB, including our main Experiment 2, the stimulus alternates between the original and an altered stimulus. If change really is not detected, though, the altered stimulus can be used as a new "original" pattern, and another transformation can be added. The process can be repeated indefinitely, resulting in a modification of the flicker paradigm, a "progressive transformation paradigm" (Fig. 2). Hidden image substitution (HIS) would be the result of using this paradigm, where gist is preserved while the image changes progressively without the change being detected. This is a stronger test of the effect of visual short-term memory on recognition than the flicker or mud-splashes designs, because all of the details of the original image are eventually replaced.

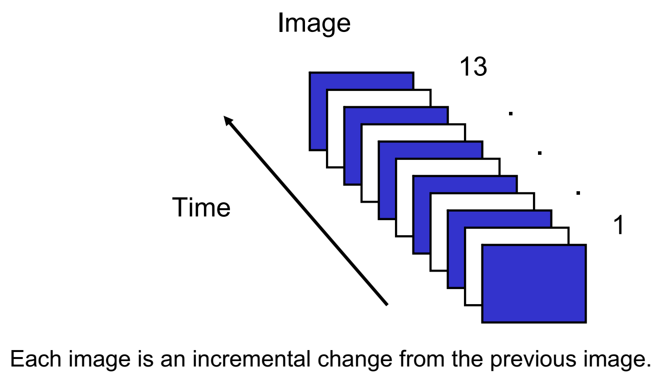

Fig. 2. The progressive transformation paradigm. Interleaved unshaded blank screens are displayed for 70 ms.

One other demonstration of change blindness uses a progressive transformation paradigm, developed independently at the Exploratorium in San Francisco, CA by Richard Brown and Ted Koterbas. In this demonstration a highly geometric urban street scene is transformed stepwise into another urban scene by replacing panels or objects one by one in meaningful units, with brief blanks between images. For instance, in one change a shop sign is changed into a different sign, but the location of the sign remains the same in the new image. Thus, the original and transformed scenes remain highly correlated and in spatial register. In our procedure the correlation is not necessary, as a natural unretouched photograph of one scene is exchanged with a natural photograph of a completely different scene of the same gist.

We adapt the technique of Blackmore, Brelstaff, Nelson, and Troscianko (1995), who used CB in a novel way to test the composite image model of the nature of transsaccadic memory and visual representation. They displayed images of natural scenes on a computer (image A: a full glass of milk on breakfast table) and programmed the image to "jump" from the center to a random corner, thus inducing a saccade in viewers. During the jump, a change was made to the image (image A′: an empty glass on breakfast table) simultaneously. They reasoned that if a detailed visual representation was indeed stored across saccades, observers should be able to rely on visual short-term memory to compare the current image with the stored image. This was contrary to what they found, however, as they observed a striking 55% detection rate from their viewers.

5.1. Method

5.1.1. Participants

Thirty undergraduate students from the University of California, Santa Cruz volunteered to participate in the experiment to fulfill course requirement. All had normal or corrected-to-normal vision.

5.1.2. Apparatus

Color photographs of natural scenes were displayed on a CRT monitor scanning at 85 Hz, 24° high $\times$ 32° wide in size and 1024 $\times$ 768 pixels in resolution.

5.1.3. Stimuli

Twenty-one photographs were used, 7 of which were control images that did not have an alteration. The remaining 14 were 7 pairs of images, each pair containing two distinct images that shared the same gist (picture A and picture A′). For instance, a picture of a car (picture A) was paired with a picture of a different car (picture A′) in a different location.

5.1.4. Procedure

Observers sat with their eyes approximately 60 cm from the computer monitor and placed one finger on a response button. They were told that images would "jump" in random directions and sometimes "morph" in small increments into another picture of similar theme. Their task was to press the response button in a timely manner if they saw a change.

The control and experimental trials shared the same time course and design except for the progressive transformation. Observers viewed the initial picture for 1 s. After a 70 ms blank screen, the second image (with a superimposed 8 $\times$ 8 checkerboard of blank squares) appeared at another place on the screen for 250 ms in an unpredictable fashion. This sequence was repeated with the same image regions always obscured by the white checkers until the 13th step was complete (Fig. 3). The control pictures stayed the same throughout all 13 jumps. The experimental

Fig. 3. Unmasked (left) and masked (right) images. In all trials observers saw the unmasked image for 1 s, followed by the masked image alternated with 70 ms blanks. In control trials the image remained the same for 13 iterations, while in experimental trials it was progressively transformed into another image by substituting squares of the new image.

images had cumulative substitutions of 2 or 3 checkered squares with squares from the new image during each step, in an unpredictable sequence. Order of presentation of the sequences was randomized for each observer. We measured the number of substitutions accomplished before the observer pressed the response key, and whether the key was pressed at all. Pressing the key ended the display sequence.

A key press during an experimental picture display was recorded as a hit, whereas key press during a control picture display was recorded as a false alarm. Display duration was reduced from the Experiment 3 duration to prevent observers from consistently tracking a single square and monitoring it for changes, an algorithm that would always result in change detection.

5.2. Results

5.2.1. Failure of detection

Averaging across images and subjects we found a 30% detection rate (62 hits out of the 210 total possible hits). Fig. 4 shows the three least-detected sets of images.

On average, each observer made 2.1 out of 7 possible hits, a strikingly low hit rate. This rate is significantly below the expected hit rate of 7 in a *t*-test, $t(29) = -16.95$, $p < .001$.

We also conducted signal detection analysis on each individual participant. For eleven of the 30 participants, *d*-primes were either at zero or below. In other words, a little more than one-third of our observers had no discrimination, or slightly worse due to being unlucky.

5.2.2. Delay of detection

The composite image model, along with other studies (Jonides, Irwin, & Yantis, 1982), suggests that detailed visual representations accumulate across saccadic eye movements. The model that they suggested has been rejected because of stimulus persistence artifacts (Bridgeman & Mayer, 1983). The images used were artificial geometric arrays, however, much less detailed than natural scenes. To test this hypothesis with natural images, we looked at the distribution of our observers' responses across trials. If detail accumulates across eye movements, then memory of the first template image should be fairly intact. Furthermore, we should observe a strong positive linear relationship between hit rate and number of iterations because as the trials went on, more and more changes were made to the current image, thus making it more and more different from the unmasked original and reducing the difficulty of detecting a change. However, if Blackmore et al.'s (1995) suggestion is correct, that only a few attended perceptual details and a gist survive a saccade, then the accumulation of visual representation would be much weaker across the trials, and we should observe a flatter, if at all positive, slope between hit rate and iteration.

Fig. 5 shows the number of substitutions required for detection, summed across all images. After the first few iterations the function is flat, contradicting the composite image model. The gradual early rise in detection rate may reflect observers' reaction time in pressing their button, as the presentation rate was faster in this experiment than in the above experiments.

5.3. Discussion

Our progressive transformation paradigm provides a stronger test of the lack of detailed image representation than conventional CB paradigms, because a continuously changing image, becoming less and less like the original, should become progressively easier to detect with successive change iterations. The flicker (Rensink, O'Regan, & Clark, 1997) and "mud-splashes" (O'Regan et al., 1999) paradigms, however, fail to differentiate whether detection occurs because an observer happens to attend the right detail at the right moment or because a gradually improving internal representation has achieved enough detail to enable detection. The nearly flat rates of change detection that we find over iterations of progressive change suggest that visual representation does not build up over time. This finding is consistent with the findings of Mitroff and Simons (2002), who also found that change detection performance does not accumulate across successive iterations. Hollingworth and Henderson (2004) came to the same conclusion when comparing detection rates from a gradual-change view with a sudden-change view, although their data show a climb of detection rates over 81 trials, which implies an accumulation of internal representation. One possible explana-

Fig. 4. The three image pairs that resulted in the greatest failures of change detection in Experiment 4.

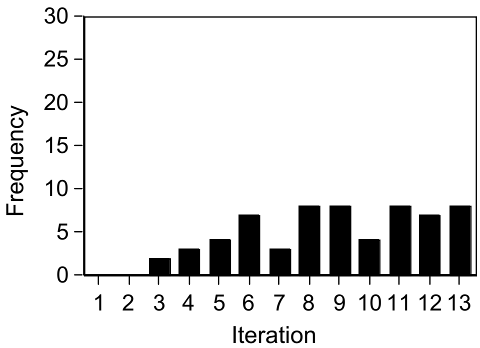

Fig. 5. Rates of detection of image change as a function of cumulative number of changes. The detection rate increases slowly for the first few iterations because the iterations are short, so that a delay in responding spreads over subsequent iterations.

tion to account for the difference in our data is that the accumulation of information is very slow, so that our 13-step transformation is too short to demonstrate their climbing effect. Inspection of real-world scenes generally conforms to the shorter rather than the longer sample time.

The eye can saccade to a displaced image only after it has changed and jumped to a new location. Saccadic suppression begins before the eye begins to move, but ends with the arrival of the eye at the new fixation point (Bridgeman et al., 1975). Therefore, saccadic suppression takes place after the change, so that our observed low detection rate confirms that attention rather than saccadic suppression is the main factor behind change blindness as previously suggested (Rensink, O'Regan, & Clark, 1997; Turatto, Bettella, Umilta', & Bridgeman, 2003). The relatively flat detection rate in Fig. 5 indicates that observers do not use information from more than 1 or 2 previous samples to aid in change detection.

6. General discussion

Since CB occurs whenever detecting a change requires more specific information than is held in visual short-term memory (Henderson, 1997), we can conclude from the present experiments that gist is coded separately from the specifics of the scene from which it is extracted. Details must, of course, be used to build up a gist, but apparently this gist once established becomes independent of scene details, since changing all of the details often fails to trigger change detection, and changing a few details hardly ever triggers detection. The specific lines, edges and features coded in early vision apparently are used to encode a gist, and then are lost. Object substitutions that change gist are more easily detected than substitutions that leave gist intact.

Experiment 2 demonstrates that small perceptual changes are more readily detected when they change the meaning of a scene, indicating that gist is an important factor in the encoding and representation of a scene. These results substantiate a long-held but previously unsubstantiated belief that a change to the gist of a scene is more likely to be detected than a change of similar magnitude that does not affect gist.

Experiment 1 supports our assumption that gist was changed in the experimental images used in Experiment 2. All of the experimental images differed significantly from their respective original images. Though two of the control images are problematic, the remaining images are adequate to substantiate the study's overall conclusions. van Montfort's (2007) method, used in Experiment 1, is particularly useful and innovative in helping to determine whether the gist of two images differs significantly.

Experiment 3 supports the proposition that feature changes were largely equivalent in detectability (notwithstanding gist) in Experiment 1. Only two of the nine images were at all problematic. In each case, the resolution seems to be relatively simple (a change in luminance in one image and a slight change in orientation in the other). Seven of the nine images were equivalent in their perceptual change or differed in a way that would be contrary to (not favor) the results of Experiment 1. In short, gist was very important in determining which changes would be detected, while degree of feature similarity was not systematically related to detectability.

We have found it difficult to manipulate the gist between two natural images (A and A′) and create an equivalent perceptual change in a control image (A″) that does not affect gist. Most typically, when the gist of a scene changes the features of that scene change also. In other words, things that are different in meaning in the natural world tend also to be different in appearance (i.e., their features or perceptual qualities); extensive change in features becomes a reliable signal for a change in gist.

The progressive transformation design used here in Experiment 4 extends the traditional flicker paradigm, which uses alternating images that are identical except for one change, to progressive substitution of one natural image for another, resulting in HIS. Thus, the pixels of one image are substituted with uncorrelated pixels of another image. Another finding of change blindness in complete image change without a change in gist is the substitution of a natural black-and-white image with its negative, at the same time changing a target object in the image (Turatto, Bettella, Umilta', & Bridgeman 2003). This effect can be identified as hidden entire reversal substitution (HERS), showing that changing the polarity of edges in a scene can induce change blindness without a blank flicker interval. The result complements the finding that large changes in low-level details without a change in the meaning or gist of a scene often go undetected.

As noted above, a central issue that must be further studied is how to appropriately define gist, and the level at which it should be defined. Should it be defined differently depending on the context of the task? For instance, are there some questions that might be best answered defining gist with low-level perceptual components? How important are individual differences in how people define the gist of a scene? Finally, questions remain about the means of change detection. Is gist the primary criterion for change detection? Do some people first use spatial arrangement or spatial relationships in an attempt to detect change? More generally, what is the primary strategy(s) for change detection?

While any one experiment in this paper is subject to some ambiguity of interpretation, it is difficult to find a thesis that better matches all of the data than a modified impoverished-representation theory, enhanced with the gist thesis. However, it is now evident that conscious report of detection can underestimate the integrity of visual representation, and should not be used as a conclusive evidence for impoverished-representation theories (Fernandez-Duque & Thornton, 2003; Hollingworth, 2006; Hollingworth, Williams, & Henderson, 2001; also see Simons & Rensink, 2005, for a review). It is therefore possible that a somewhat detailed representation is stored in a channel or format that cannot be accessed by consciousness. One notable exception comes from Hollingworth and Henderson (2002), where observers detected token, type, and orientation changes above chance level. We have not observed the same accuracy as the Hollingworth and Henderson study. In our final experiment the entire scene is changed, however, while in theirs one object is changed or rotated while most of the scene remains unchanged. Thus, attention might be driven to the change in their study. Our scenes, then, correspond to type or token changes in their nomenclature for everything in the scene. Together, our results suggest that even an entire scene change can go undetected if gist remains the same, while small changes that alter gist are more easily detected. We therefore conclude that gist is a powerful component of scene recognition, sometimes overshadowing details of the image itself.

References

Ballard, D. H., Hayhoe, M. M., & Pelz, J. B. (1995). Memory representations in natural tasks. *Journal of Cognitive Neuroscience, 7*, 66–80.

Becker, M. W., & Pashler, H. (2002). Volatile visual representations: Failing to detect changes in recently processed information. *Psychonomic Bulletin & Review, 9*, 744–750.

Biederman, I. (1972). Perceiving real-world scenes. *Science, 177*, 77–79.

Biederman, I. (1981). On the semantics of a glance at a scene. In M. Kubovy & J. R. Pomerantz (Eds.), *Perceptual organization* (pp. 213–252). Hillsdale, NJ: Lawrence Erlbaum Associates Inc..

Biederman, I., Mezzanotte, R., & Rabinowitz, J. (1982). Scene perception: Detecting and judging objects undergoing relational violations. *Cognitive Psychology, 14*, 143–177.

Blackmore, S. J., Brelstaff, G., Nelson, K., & Troscianko, T. (1995). Is the richness of our visual world an illusion? Transsaccadic memory for complex scenes. *Perception, 24*, 1075–1081.

Brewer, W. F., & Treyans, J. C. (1981). The role of schemata in memory for places. *Cognitive Psychology, 13*, 207–230.

Bridgeman, B., Hendry, D., & Stark, L. (1975). Failure to detect displacement of the visual world during saccadic eye movements. *Vision Research, 15*, 719–722.

Bridgeman, B., & Mayer, M. (1983). Failure to integrate visual information from successive fixations. *Bulletin of the Psychonomic Society, 21*, 285–286.

Brockmole, J., & Irwin, D. E. (2005). Eye movements and the integration of visual memory and visual perception. *Perception & Psychophysics, 67*, 495–512.

Coltheart, V. (1999). Introduction: Perceiving and remembering brief visual stimuli. In V. Coltheart (Ed.), *Fleeting memories* (pp. 1–12). London, England: MIT.

Cowan, N. (2000). The magical number 4 in short-term memory: A reconsideration of mental storage capacity. *Behavioral and Brain Sciences, 24*, 87–185.

Enns, J. T., & Di Lollo, V. (1997). Object substitution: A new form of masking in unattended visual locations. *Psychological Science, 8*, 135–139.

Fernandez-Duque, D., & Thornton, I. M. (2003). Explicit mechanisms do not account for implicit localization and identification of change: An empirical reply to Mitroff et al.. *Journal of Experimental Psychology, 29*, 846–858.

Freeman, G. H., & Halton, J. H. (1951). Note on an exact treatment of contingency goodness-of-fit and other problems of significance. *Biometrika, 38*, 141–149.

Friedman, A. (1979). Framing images: The role of knowledge in automatized encoding and memory for gist. *Journal of Experimental Psychology: General, 108*, 316–355.

Gilchrist, I. D., & Harvey, M. (2000). Refixation frequency and memory mechanisms in visual search. *Current Biology, 10*, 1209–1212.

Grimes, J. (1996). On the failure to detect changes in scenes across saccades. In K. Atkins (Ed.), *Perception. Vancouver studies in cognitive science* (pp. 89–110). New York: Oxford University. Vol. 5.

Henderson, J. M. (1997). Transsaccadic memory and integration during real-world object perception. *Psychological Science, 8*, 51–55.

Henderson, J. M., & Hollingworth, A. (1998). Eye movements during scene viewing: An overview. In G. Underwood (Ed.), *Eye guidance in reading and scene perception* (pp. 269–283). Oxford, England: Elsevier.

Henderson, J. M., & Hollingworth, A. (1999). The role of fixation position in detecting scene changes across saccades. *Psychological Science, 10*, 438–443.

Hollingworth, A. (2006). Visual memory for natural scenes: Evidence from change detection and visual search. *Visual Cognition, 14*, 781–807.

Hollingworth, A., & Henderson, J. M. (2000). Semantic informativeness mediates the detection of changes in natural scenes. *Visual Cognition, 7*, 213–235.

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*, 113–136.

Hollingworth, A., & Henderson, J. M. (2004). Sustained change blindness to incremental scene rotation: A dissociation between explicit change detection and visual memory. *Perception & Psychophysics, 66*, 800–807.

Hollingworth, A., Williams, C. C., & Henderson, J. M. (2001). To see and remember: Visually specific information is retained in memory from previously attended objects in natural scenes. *Psychonomics Bulletin & Review, 8*, 761–768.

Intraub, H. (1981). Rapid conceptual identification of sequentially presented images. *Journal of Experimental Psychology: Human Perception and Performance, 7*, 604–610.

Intraub, H. (1999). Understanding and remembering briefly glimpsed images: Implications for visual scanning and memory. In V. Coltheart (Ed.), *Fleeting memories* (pp. 47–70). London, England: MIT.

Intraub, H., Bender, R. S., & Mangels, J. A. (1992). Looking at images but remembering scenes. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 18*, 180–191.

Intraub, H., & Richardson, M. (1989). Wide-angle memories of close-up scenes. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 15*, 179–187.

Irwin, D. E. (1991). Integrating information across saccadic eye movements. *Cognitive Psychology, 23*, 420–456.

Irwin, D. E. (2003). Eye movements and visual cognitive suppression. In D. E. Irwin & B. H. Ross (Eds.), *Cognitive vision: The psychology of learning and motivation* (pp. 265–293). San Diego, CA, USA: Academic Press.

Irwin, D. E., & Andrews, R. (1996). Integration and accumulation of information across saccadic eye movements. In T. Inui & J. L. McClelland (Eds.), *Attention and performance. XVI: Information integration in perception and communication* (pp. 125–155). Cambridge, MA: MIT.

Irwin, D. E., & Zelinsky, G. J. (2002). Eye movements and scene perception: Memory for things observed. *Perception and Psychophysics, 64*, 882–895.

Johnston, W. A., & Hawley, K. J. (1994). Perceptual inhibition of expected inputs: The key that opens closed minds. *Psychonomic Bulletin and Review, 1*, 56–72.

Jonides, J., Irwin, D. E., & Yantis, S. (1982). Integrating visual information from successive fixations. *Science, 215*, 192–194.

Levin, D. T., & Simons, D. J. (1997). Failure to detect changes to attended objects in motion images. *Psychonomic Bulletin and Review, 4*, 501–506.

Levin, D. T., Simons, D. J., Angelone, B. L., & Chabris, C. F. (2002). Memory for centrally attended changing objects in an incidental real-world change detection paradigm. *British Journal of Psychology, 93*, 289–302.

Luck, S. J., & Vogel, E. K. (1997). The capacity of visual working memory for features and conjunctions. *Nature, 390*, 279–281.

Mandler, J., & Ritchey, G. H. (1977). Long-term memory for pictures. *Journal of Experimental Psychology: Human Learning and Memory, 3*, 386–396.

McConkie, G. W., & Currie, C. B. (1996). Visual stability across saccades while viewing complex images. *Journal of Experimental Psychology: Human Perception and Performance, 22*, 563–581.

Mitroff, S. R., & Simons, D. J. (2002). Changes are not localized before they are explicitly detected. *Visual Cognition, 9*, 937–968.

Nickerson, R. S., & Adams, J. J. (1979). Long-term memory for a common object. *Cognitive Psychology, 11*, 287–307.

O'Regan, J. K. (1992). Solving the "real" mysteries of visual perception: The world as an outside memory. *Canadian Journal of Psychology, 46*, 461–488.

O'Regan, J. K., Deubel, H., Clark, J. J., & Rensink, R. A. (2000). Image changes during blinks: Looking without seeing and seeing without looking. *Visual Cognition, 7*, 191–211.

O'Regan, J. K., Rensink, R. A., & Clark, J. J. (1999). Change-blindness as a result of mudsplashes. *Nature, 398*, 34.

Palmer, S. E. (1977). Hierarchical structure in perceptual representation. *Cognitive Psychology, 9*, 441–475.

Pezdek, K., Maki, R., Valencia-Laver, D., Whetstone, T., Stoeckert, J., & Dougherty, R. (1988). Image memory: Recognizing added and deleted details. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 14*, 468–476.

Pollatsek, A., & Rayner, K. (1992). What is integrated across fixations? In K. Rayner (Ed.), *Eye movements and visual cognition: Scene perception and reading* (pp. 166–191). New York: Springer-Verlag.

Potter, M. C., & Levy, E. I. (1969). Recognition memory for a rapid sequence of images. *Journal of Experimental Psychology, 81*, 10–15.

Rensink, R. A. (2000a). Visual search for change: A probe into the nature of attentional processing. *Visual Cognition, 7*, 345–376.

Rensink, R. A. (2000b). Seeing, sensing, and scrutinizing. *Vision Research, 40*, 1469–1487.

Rensink, R. A. (2002). Change detection. *Annual Review of Psychology, 53*, 245–277.

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.

Rensink, R. A., O'Regan, J. K., & Clark, J. J. (2000). On the failure to detect changes in scenes across brief interruptions. *Visual Cognition, 7*, 127–145.

Ryan, T., & Schwartz, C. (1956). Speed of perception as a function of mode of representation. *American Journal of Psychology, 69*, 60–69.

Sancocki, T., & Epstein, W. (1997). Priming spatial layout of scenes. *Psychological Science, 8*, 374–378.

Shepard, R. N. (1967). Recognition memory for words, sentences, and images. *Journal of Verbal Learning and Verbal Behavior, 6*, 156–163.

Simons, D. J. (1996). In sight, out of mind: When object representations fail. *Psychological Science, 7*, 301–305.

Simons, D. J., Franconeri, S. L., & Reimer, R. L. (2000). Change blindness in the absence of a visual disruption. *Perception, 29*, 1143–1154.

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

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

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

Standing, L. (1973). Learning 10,000 images. *Quarterly Journal of Experimental Psychology, 25*, 207–222.

Standing, L., Conezio, J., & Haber, R. N. (1970). Perception and memory for images: Single-trial learning of 2500 visual stimuli. *Psychonomic Science, 19*, 73–74.

Turatto, M., Bettella, S., Umilta', C., & Bridgeman, B. (2003). Perceptual conditions necessary to induce change blindness. *Visual Cognition, 10*, 233–255.

van Montfort, X. (2007). What makes a difference: A method to determine whether a change in an image affects the perceived gist. *Behavior Research Methods, 39*, 399–406.

Vogel, E. K., Woodman, G. F., & Luck, S. J. (2001). Storage of features, conjunctions, and objects in visual working memory. *Journal of Experimental Psychology: Human Perception and Performance, 27*, 92–114.

Williams, P., & Simons, D. J. (2000). Detecting changes in novel, complex three-dimensional objects. *Visual Cognition, 7*, 297–322.

Wolfe, J. M. (1998). Visual memory: What do you know about what you saw? *Current Biology, 8*, 303–304.

Wolfe, J. M. (1999). Inattention amnesia. In V. Coltheart (Ed.), *Fleeting memories* (pp. 71–94). Cambridge, MA: MIT.

Zelinsky, G. (2003). Detecting changes between real-world objects using spatiochromatic filters. *Psychonomic Bulletin & Review, 3*, 533–555.