Memory & Cognition
2003, 31 (6), 930-940

Testing a conceptual locus for the inconsistent object change detection advantage in real-world scenes

ANDREW HOLLINGWORTH
Yale University, New Haven, Connecticut

and

JOHN M. HENDERSON
Michigan State University, East Lansing, Michigan

Changes to objects that are inconsistent with the scene in which they appear are detected more accurately than changes to consistent objects. In three experiments, we tested whether this inconsistent object advantage derives from the differential retention of conceptual codes generated from a brief view of a real-world scene in accordance with a conceptual short-term memory (CSTM) hypothesis. A scene was presented for 250 msec, followed by a brief mask and a test scene in which a target object was either changed or not changed. In Experiment 1, changes that altered conceptual content (object deletion) were contrasted with visual changes (left–right orientation changes). In Experiment 2, the duration of the mask was manipulated to vary the amount of time available for conceptual consolidation of the initial scene. In Experiment 3, the type of mask was manipulated: Either a meaningless pattern mask or a meaningful, and thus conceptually disruptive, scene was shown. The inconsistent object advantage was obtained in each experiment, yet in none was it modulated in the direction predicted by the CSTM hypothesis. Instead, the inconsistent object advantage is likely to be caused by contextual influence on memory for visual object representations.

To what extent does the scene in which an object appears influence perception of and memory for that object? In the perceptual domain, it was long thought that scene context directly supports the identification of local objects that are consistent with (i.e., likely to be found in) the scene, either by facilitating the construction of perceptual descriptions of consistent objects (e.g., Biederman, Mezzanotte, & Rabinowitz, 1982; Boyce, Pollatsek, & Rayner, 1989) or by lowering the criterion for determining a match between the long-term memory representations of consistent object types and current perceptual information (e.g., Friedman, 1979; Palmer, 1975). The strongest evidence for contextual facilitation of object perception came from experiments by Biederman et al. (1982) in which participants were asked to detect the presence of an object in a briefly presented scene. Biederman et al. found greater detection sensitivity when the target object was consistent with the scene (such as a fire hydrant in a street scene) than when it was inconsistent (such as a fire hydrant in a diner scene). However, Hollingworth and Henderson (1998, 1999; for a review, see Henderson & Hollingworth, 1999) demonstrated that the apparent advantage for consistent object detection does not result from perceptual facilitation of consistent objects but, rather, from guessing strategies based on knowledge of typical scene content. When response bias was eliminated from the Biederman et al. paradigm, no facilitating effect of consistent scene context was observed. In fact, we observed an advantage for the detection of inconsistent objects, an *inconsistent object advantage*.

Empirical work in the study of scene context effects on long-term object memory has progressed similarly. In an early study by Brewer and Treyens (1981), participants were seated in a mock office with a number of consistent and some inconsistent objects for 35 sec. On a subsequent verbal free-recall test, probability of recall was positively correlated with independent ratings of object consistency with the office scene. Brewer and Treyens concluded that general knowledge of the composition of offices (i.e., an office schema) supported the retention of object information consistent with expectations. However, further work demonstrated that this advantage was likely due to guessing based on knowledge of typical scene content rather than on facilitated retention of consistent objects. Pedzek,

---

This research was supported by grants from the U.S. Army Research Office (DAAH04-94-G-0404) and the National Science Foundation (BCS-0094433). The opinions expressed in this article are those of the authors and do not necessarily represent the views of the Department of the Army or any other governmental organization. Reference to or citations of trade or corporate names do not constitute explicit or implied endorsement of those entities or their products by the author or by the Department of the Army. Correspondence concerning this article should be addressed to A. Hollingworth, Department of Psychology, University of Iowa, 11 Seashore Hall E, Iowa City, IA 52242, or to J. Henderson, Cognitive Science Program, Michigan State University, East Lansing, MI 48824-1117 (e-mail: andrew-hollingworth@uiowa.edu or john@eyelab.msu.edu).

Whetstone, Reynolds, Askari, and Dougherty (1989) replicated the Brewer and Treyens study but under conditions that eliminated contextually guided guessing. With this corrected design, they observed a robust advantage for the recall of inconsistent objects over consistent objects. In addition, Pedzek et al. employed a token change detection manipulation to test memory for specific visual feature information. A subset of objects in the test room was replaced by different objects from the same category (e.g., a phone was replaced by a different example of a phone). Again, change detection sensitivity was significantly higher for inconsistent objects than for consistent objects. Similar inconsistent object advantages have been observed in long-term memory paradigms using depicted, as opposed to actual, scenes (Friedman, 1979; Goodman, 1980; Hock, Romanski, Galie, & Williams, 1978). Thus, both studies of perceptual detection and studies of long-term object memory demonstrate that inconsistent objects are preferentially represented from natural scenes.

Perhaps there is a single explanation that accounts for the inconsistent object advantage both in perceptual detection studies and in long-term memory studies. In the object detection studies of Hollingworth and Henderson (1998, 1999), the inconsistent object advantage was observed only when task performance depended on memory for the scene. In one version of the object detection paradigm, a target object label was displayed, followed by a briefly presented scene (250 msec or less) and a pattern mask. Participants could detect an object and prepare a response without holding scene information in memory, and no detection difference for consistent versus inconsistent objects was observed. In a second version of the object detection paradigm, the target label (or target alternatives in a forced-choice design) was presented after the scene. This method required participants to maintain object information in memory, because the target was not specified until after the scene was removed. It was in this version of the paradigm that we observed an inconsistent object advantage. Thus, a plausible explanation for our inconsistent object detection advantage, in keeping with evidence from long-term memory studies, is that scene context influenced memory for objects in the initial scene. If this is the case, given the timing of the object detection studies, contextual influences must have been operating on short-term object memory.

To test more directly whether scene context influences short-term memory for object information in scenes, Hollingworth and Henderson (2000) employed a change detection paradigm in which a target object either was mirror reversed or remained the same across two images of a scene. The initial scene image was presented for 250 msec, followed by an 80-msec pattern mask, followed by a test scene. To detect a change, participants must have retained in memory visually specific object information from the first scene. In accordance with the object detection studies, changes to inconsistent objects were detected more accurately than changes to consistent objects. Further evidence of an inconsistent object change detection advantage comes from Hollingworth, Williams, and Henderson (2001), in which a target object was changed during the saccadic eye movement away from that object after it had been fixated for the first time. Token changes were detected more accurately for inconsistent than for consistent target objects. In summary, the inconsistent object advantage is a robust phenomenon that has been observed across a variety of experimental paradigms. Although knowledge of typical scene composition does not appear to interact with the initial perceptual identification of an object, it does appear to influence memory for the contents of a scene, in both long-term and short-term memory paradigms.

In the present study, we investigated the type of memory representation responsible for producing the inconsistent object change detection advantage. We contrasted two hypotheses. The first, a *visual short-term memory (VSTM) hypothesis*, states that knowledge of typical scene composition influences short-term memory for visual object representations. The second, a *conceptual short-term memory (CSTM) hypothesis*,[1] states that knowledge of typical scene composition influences short-term memory for conceptual representations. The VSTM hypothesis must be considered the default explanation, because the inconsistent object advantage is found in paradigms requiring visual judgments, such as token or orientation change detection. However, conceptual-level explanations have been successful in accounting for phenomena that appear to be perceptual at first blush, such as the attentional blink (Chun & Potter, 1995). The inconsistent object advantage reflects the influence of generalized knowledge of typical scene composition on some form of brief memory, and thus it is an attractive candidate for a conceptual-level explanation.

According to Potter (1976, 1993, 1999), CSTM constitutes the brief retention of conceptual codes in the service of producing structured representations of complex stimuli, such as visual scenes. In this view, CSTM comprises three principal operations: (1) the brief retention of conceptual codes derived from object and scene identification, (2) the retrieval of structured conceptual information from long-term memory, such as a scene schema specifying typical objects found in a scene of that type (see Friedman, 1979; Mandler & Johnson, 1976), and (3) the integration of current conceptual codes within this structured representation. Potter has provided evidence that conceptual codes are activated very quickly from visual scenes, yet are highly prone to interference from subsequent conceptual information unless there is enough time to structure these codes and stabilize them—a process termed *conceptual consolidation* (Potter, 1976).

Under the CSTM hypothesis, the presentation of the initial scene leads to the identification of the scene and some of the constituent objects. Conceptual codes generated from object and scene identification are maintained briefly, and the appropriate scene schema is retrieved from memory. During the process of integrating conceptual codes derived from visual identification within the schema, conceptual codes that are inconsistent with scene context (i.e., absent from the schema) are preferentially retained, perhaps because the conceptual codes for consistent objects are more likely to be replaced by default values in the schema (Friedman, 1979). Thus, conceptual information from identification of inconsistent targets would more often be available for comparison with the test scene, leading to the inconsistent object advantage. One obstacle for this hypothesis is to provide an explanation for consistency effects on the detection of changes to the visual form of target objects (such as left–right orientation change). However, one need only assume that successful conceptual retention provides an effective retrieval cue for visual representations encoded from the initial scene. Differences in the availability of visual representations would result from differences in conceptual retention.

In the present study, a single change paradigm was employed, in which an initial scene was presented for 250 msec, followed by a mask, followed by a test scene (as in Hollingworth & Henderson, 2000, Experiments 2 and 3). The CSTM hypothesis generates three predictions that were tested in three experiments. First, if the inconsistent object advantage derives from differential retention of conceptual codes, scene changes that alter conceptual information (such as target deletion) should produce a larger inconsistent object advantage than scene changes that do not significantly alter conceptual information (such as left–right orientation change). This prediction was tested in Experiment 1. Second, because the consolidation of conceptual codes depends critically on stimulus duration (Potter, 1976), a larger inconsistent object advantage should be observed when more time is available for consolidation, a prediction tested in Experiment 2. Finally, because conceptual codes generated from a brief view of a scene are highly susceptible to interference from subsequent conceptual information (Potter, 1976), the inconsistent object advantage should be reduced or eliminated if a conceptually meaningful distractor scene (a *conceptual mask*) is presented between the initial and test scenes. This prediction was tested in Experiment 3.

EXPERIMENT 1

In Experiment 1, we compared detection of two types of change to target objects within line drawings of scenes: deletion and left–right orientation change. Deletions alter the semantic content of the scene and could be detected by the retention of a conceptual identity code for the target object. Left–right orientation changes do not significantly alter the semantic content of the scene and could not be detected solely by the retention of an identity code for the target. The CSTM hypothesis predicts that the inconsistent object advantage should be larger in the deletion condition than in the orientation change condition. Deletion could be detected directly by the retention of a conceptual identity code and, thus, should produce a large inconsistent object advantage if that advantage derives from differential conceptual retention. In contrast, orientation change detection performance—a visual task—could show an effect of differential conceptual retention only indirectly, and therefore should produce a less robust inconsistent object advantage under the CSTM hypothesis. If a reverse pattern of results (i.e., a larger inconsistent object advantage for orientation changes than for deletions) was obtained, this would provide evidence against the CSTM hypothesis and in favor of the VSTM hypothesis.

Method

Participants. Twenty-four Michigan State University undergraduate students participated. All the participants reported normal or corrected-to-normal vision, were naive with respect to the purposes of the research, and received partial credit in introductory psychology for their participation.

Apparatus. The stimuli were displayed on a NEC XE15 (Multisync) monitor operating at 100 Hz. The participants used a button box to start each trial and to record their "same"/"changed" decisions. Stimulus presentation and response collection were controlled by a 486-66 microcomputer. The microcomputer was also interfaced with the button box.

Stimuli. The stimuli were the same as those used in Hollingworth and Henderson (2000). Line drawings of 20 scenes and 20 target objects were generated from photographs of natural scenes. Fourteen scenes were modified from those used by van Diepen and De Graef (1994), and the other 6 scenes were generated from photographs taken in the East Lansing, Michigan area. In both cases, the main contours of the scenes were traced using computer graphics software to create gray-scale line drawings. The images generated from the two sources were not distinguishable. Target objects were created independently of the scene stimuli by digital tracing of scanned images. One object was chosen as consistent with each scene, the 20 scenes were paired, and the inconsistent target conditions were created by swapping objects across paired scenes, thus controlling for visual differences between scene items. Paired target objects appeared in the same position in each scene, which did not coincide with the experimenter-determined initial fixation position. An earlier norming study using a similar set of stimuli demonstrated that consistent objects were indeed considered likely to appear in the scene and inconsistent objects were considered unlikely to appear in the scene (Henderson, Weeks, & Hollingworth, 1999). Figure 1 presents a sample scene and illustrates the consistency manipulation.

For each scene in both the consistent and inconsistent conditions, target objects were manipulated in one of two ways to create the object-change conditions. For the deletion condition, the target object was removed from the scene. For the orientation change condition, the target object was mirror reflected about its vertical axis. Finally, in the *same* condition, the target object was not manipulated. The scenes subtended a visual angle of 23° (width) × 15° (height) at a viewing distance of 64 cm. Target objects subtended an average of 2.75°, measuring the largest extent of each object. All images were displayed as gray-scale contours on a white background at a resolution of 800 × 600 pixels by 16 levels of gray. The pattern mask that was presented between the initial scene and the test scene consisted of overlapping line segments, curves, and angles. The mask was slightly larger than the scenes, and the scenes were completely obliterated when presented simultaneously with the mask.

Procedure. The participants were tested individually. The experimenter explained that the task was to determine whether any object changed between successive presentations of a scene. The experimenter also described the nature of the possible changes. The participants were then seated in front of the computer monitor, with one hand resting on each button of the button box. Viewing distance was maintained by a forehead rest.

During each trial, the participants saw a fixation cross and a prompt instructing them to press a pacing button to begin the trial. Once a participant pressed the button, the fixation cross remained on the

Figure 1. An example of the type of scene used and the target object consistency manipulation. The top scene contains a consistent target object (mixer), whereas the bottom scene contains an inconsistent target object (chicken). This kitchen scene was paired with a farmyard scene in which the chicken was consistent and the mixer inconsistent.

screen for an additional 500 msec. The initial scene was then presented for 250 msec, followed by the pattern mask for 400 msec, followed by the test scene. The test scene remained visible until response. The participants were instructed to press a button labeled "yes" if they detected a change and "no" if no change was evident. Figure 2 illustrates the sequence of events in a trial. The duration of the pattern mask appearing between initial and test scenes was set at 400 msec to allow sufficient time for conceptual consolidation (Potter, 1976).

A practice block of 16 trials was initially presented (two scene items appeared in each of the eight conditions). The two scenes used in the practice block were not used in the experimental trials. The participants then completed an experimental session of 160 trials produced by a within-participants factorial combination of 2 consistency conditions × 4 change conditions (left–right orientation change, deletion, and two levels of the *same* condition) × 20 scenes. Two levels of the *same* condition were included to equate the number of trials on which a change occurred with the number of trials on which a change did not occur, and were combined for the purpose of statistical analysis. The trials were presented in a random order generated independently for each participant. The entire session lasted approximately 40 min.

Results

Two measures of change detection performance, A' and percentage correct, were analyzed. A' is a nonparametric signal detection measure with a functional range of .5 (chance) to 1.0 (perfect sensitivity) (Grier, 1971). For each participant, A' was calculated using the mean hit rate when the target changed and the mean false alarm rate when it did not. Because A' corrects for potential differences in response bias in the percentage correct data, it forms the primary data for interpreting the results of these experiments.[2] We also report the percentage correct data. To provide a percentage correct measure that corrects, at least approximately, for response bias, we collapsed over

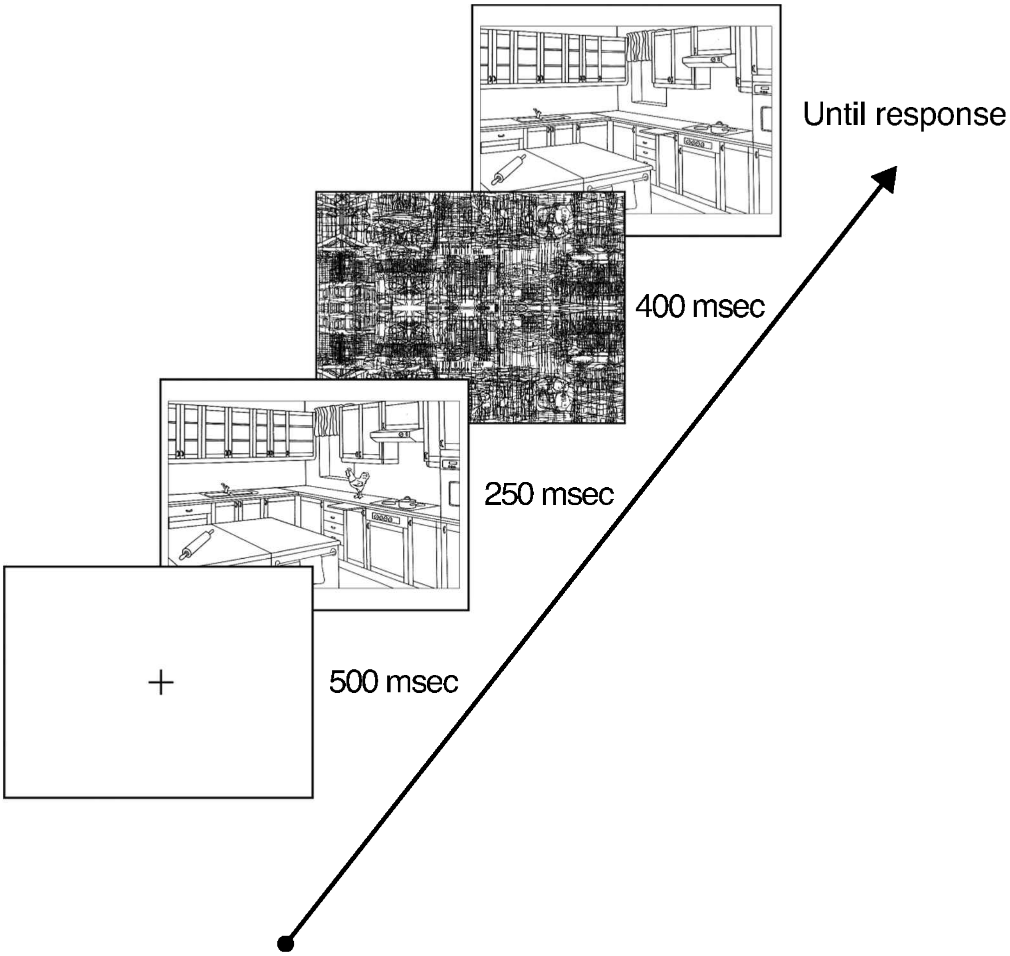

Figure 2. Schematic illustration of a trial in Experiment 1.

change trials (left–right orientation change or deletion) and no-change trials (same).

***A′* analysis**. Mean *A′* as a function of consistency and change type is presented in Table 1. There was a main effect of consistency [$F(1,23) = 19.78, MS_e = 0.0033, p < .001$], with better performance in the inconsistent condition (.822) than in the consistent condition (.770). There was also a reliable effect of change type [$F(1,23) = 206.03$, $MS_e = 0.0028, p < .001$], with deletions detected better than orientation changes (.874 vs .719, respectively). Finally, consistency and change type interacted [$F(1,23) = 14.27, MS_e = 0.0016, p < .005$]. This interaction was due to a larger inconsistent object advantage for left–right orientation changes (.083 difference) than for deletions (.022 difference), though both of these differences were reliable [$F(1,23) = 21.12, MS_e = 0.0039, p < .001$, and $F(1,23) = 5.49, MS_e = 0.001, p < .05$, respectively].

Percentage correct analysis. Mean percentage correct performance as a function of consistency and change type is reported in Table 1. As in the *A′* data, there was a reliable effect of consistency [$F(1,23) = 22.02, MS_e = 21.86$, $p < .001$]; performance was better in the inconsistent condition (73.3%) than in the consistent condition (68.9%). In addition, there was a reliable effect of change condition [$F(1,23) = 212.77, MS_e = 32.72, p < .001$], with better performance in the deletion condition (79.6%) than in the left–right orientation change condition (62.6%). As in the *A′* data, there was a reliable interaction between consistency and change condition [$F(1,23) = 8.20, MS_e = 10.87$,

Table 1
Experiment 1: Mean *A′*
and Mean Percentage Correct (% Correct)
as a Function of Target Object Consistency and Change Type

Change Condition			
Deletion		Orientation Change	
A′	% Correct	*A′*	% Correct
Consistent			
.863	78.4	.677	59.4
Inconsistent			
.885	80.9	.760	65.8

$p < .005$]. The advantage for inconsistent target objects was larger in the left–right orientation change condition (6.4% difference) than in the deletion condition (2.5% difference), though both of these differences were reliable [$F(1,23) = 24.46$, $MS_e = 20.13$, $p < .001$, and $F(1,23) = 6.20$, $MS_e = 12.60$, $p < .05$, respectively].

Discussion

In Experiment 1, we replicated the inconsistent object change detection advantage found by Hollingworth and Henderson (2000). Changes to inconsistent target objects in scenes were detected more accurately than changes to consistent target objects. However, contrary to the prediction of the CSTM hypothesis, this effect was larger for visual changes (orientation change) than for changes that altered the semantic content of the scenes (deletion).

Before we proceed to test the next prediction of the CSTM hypothesis, a potential artifactual explanation for the inconsistent object advantage needs to be discussed. It is possible that the inconsistent object advantage is caused by strategic encoding of inconsistent objects. The participants might have developed such a strategy, either consciously or unconsciously, on the basis of the contingencies of the changes. Given that an inconsistent object was present in a scene, the inconsistent object changed on 50% of the trials (the other 50% were same trials). And, when an inconsistent object was present in a scene, a consistent object never changed. We have conducted an experiment to test the strategic coding hypothesis directly (Hollingworth & Henderson, 2000, Experiment 3). In that experiment, an inconsistent object was added to the scenes that contained the consistent targets, and a condition was included in which consistent distractors changed in the scenes that contained inconsistent targets. As a result, every scene contained an inconsistent object, but an inconsistent object was changed only 12.5% of the time, making a strategy of selectively encoding inconsistent objects suboptimal. Despite this design, change detection was clearly facilitated for the inconsistent targets over the consistent targets. Thus, the inconsistent object advantage does not appear to be caused by strategic encoding. In Experiment 2, trials were divided into two blocks as another test of the strategic encoding hypothesis. If the inconsistent object advantage arises because participants learn that inconsistent objects are more likely to change than consistent objects, the effect should be more pronounced in the second block of trials, after participants have had an opportunity to learn this relationship.

EXPERIMENT 2

Potter (1976) demonstrated that conceptual consolidation of a scene (i.e., the integration of conceptual codes within a structured scene representation and the transfer of this information from CSTM to a more stable memory store) requires approximately 400 msec. In Experiment 1, the pattern mask was presented for 400 msec, so that the total stimulus onset asynchrony (SOA) between initial and test scenes was 650 msec, presumably sufficient for consolidation. In Experiment 2, the pattern mask was presented either for 400 msec or for 30 msec. In the latter case, the SOA between initial and test scenes was only 280 msec. If the inconsistent object advantage arises from differential retention of conceptual codes, then that advantage should be larger when sufficient time is available to conceptually consolidate the initial scene before onset of the test scene. The change conditions in Experiment 2 were limited to the left–right orientation change and same conditions, because the orientation change condition produced the larger inconsistent object advantage in Experiment 1 and, thus, provided us with a better opportunity to observe an interaction between consistency and mask duration.

Method

Participants. Twenty-four Michigan State University undergraduate students participated. All the participants had normal or corrected-to-normal vision, were naive with respect to the purposes of the research, and received partial credit for their participation. None of the participants had taken part in Experiment 1.

Apparatus, Stimuli, and Procedure. The apparatus, stimuli, and procedure were the same as in Experiment 1. The 160 trials were created by the within-participants factorial combination of 2 consistency conditions (consistent vs. inconsistent) × 2 change conditions (left–right orientation change vs. same) × 2 mask durations (30 msec vs. 400 msec) × 20 scenes. The mask duration manipulation was blocked; half of the participants saw the 30-msec mask condition first, and half saw the 400-msec mask condition first.

Results

***A′* analysis**. Mean A' as a function of consistency and mask duration is presented in Table 2. Block order did not produce a reliable effect ($F < 1$), nor did it interact with consistency [$F(1,22) = 1.23$, $MS_e = 0.0111$, $p = .28$]; thus, the following analyses collapsed across the block order factor. There was a reliable main effect of consistency [$F(1,23) = 9.73$, $MS_e = 0.0112$, $p < .005$], with better performance in the inconsistent condition (.743) than in the consistent condition (.675). There was also a trend toward higher performance in the 30-msec mask condition (.735) than in the 400-msec mask condition (.683) [$F(1,23) = 3.56$, $MS_e = 0.0183$, $p = .07$]. Mask duration and consistency did not interact ($F < 1$). The inconsistent object advantage was .073 in the 30-msec mask condition and .062 in the 400-msec mask condition.

Table 2
Experiment 2: Mean A' and Mean Percentage Correct (% Correct) as a Function of Target Object Consistency and Mask Duration

Mask Duration			
30 Msec		400 Msec	
A'	% Correct	A'	% Correct
Consistent			
.699	61.6	.652	59.2
Inconsistent			
.772	66.7	.714	64.1

In addition, we examined change detection performance as a function of the temporal order of the two blocks of trials, collapsing across mask duration. Change detection performance was significantly improved in the second block of trials ($A' = .748$) in comparison with the first ($A' = .657$) [$F(1,22) = 12.00$, $MS_e = 0.0124$, $p < .005$]. Critically, however, the inconsistent object advantage was numerically larger in the first block of trials ($A' = .098$) than in the second ($A' = .037$), though the interaction between block and consistency failed to reach significance [$F(1,22) = 2.00$, $MS_e = 0.0109$, $p = .17$].

Percentage correct analysis. Mean percentage correct as a function of consistency and mask duration is shown in Table 2. As in the A' data, there was a reliable effect of consistency [$F(1,23) = 14.34$, $MS_e = 83.70$, $p < .005$]. We again observed better performance in the inconsistent condition (65.4%) than in the consistent condition (60.4%). There was no main effect of mask duration [$F(1,23) = 2.01$, $MS_e = 149.5$, $p = .17$] and no interaction between mask duration and consistency ($F < 1$). The inconsistent object advantage was 5.1% in the 30-msec mask condition and 4.9% in the 400-msec mask condition.

An examination of the effect of temporal order of blocks revealed that change detection performance was significantly improved in the second block of trials (65.2%) in comparison with the first (60.6%) [$F(1,22) = 9.13$, $MS_e = 110.4$, $p < .01$]. The inconsistent object advantage was numerically larger in the first block of trials (6.35%) than in the second (2.81%), though the interaction between block and consistency failed to reach significance [$F(1,22) = 1.60$, $MS_e = 94.27$, $p = .22$].

Discussion

In Experiment 2, the duration of the pattern mask presented between the initial and test scenes was varied. However, there was no evidence that this manipulation influenced the size of the inconsistent object advantage. That advantage was just as large in the 30-msec mask condition, when less time was available for conceptual consolidation, as in the 400-msec mask condition, when more time was available for conceptual consolidation, and the numerical trend was toward a larger inconsistent object advantage in the 30-msec mask condition. Thus, this experiment does not provide support for the CSTM hypothesis.

In addition, there was a trend toward a larger inconsistent object advantage in the first block of trials than in the second. This result provides strong converging evidence against the possibility that the inconsistent object advantage results from participants learning the relationship between object consistency and probability of change. One potential explanation for the reduction of the inconsistent object advantage in the second half of the experiment is that for certain scene items, participants may have learned the position at which the target would appear. Changes to subsequent targets in those scenes could be detected accurately in both consistency conditions by a covert shift of attention to the critical location prior to the change. Such a strategy would raise overall performance with scene repetition but would reduce the size of the inconsistent object advantage through item-specific ceiling effects.

EXPERIMENT 3

To provide a final test of the CSTM hypothesis, we manipulated the type of mask appearing between the initial and test scenes. In this experiment, either a pattern mask or a conceptual mask (i.e., an image of a different scene type) was presented for 250 msec between presentation of the initial scene and that of the test scene. Potter (1976) demonstrated that presenting two different scenes in quick succession disrupts the conceptual analysis of the first. If the inconsistent object advantage derives from differential retention of conceptual codes, then the presentation of a conceptual mask between the initial and test scenes should interfere with conceptual processing, and thus should eliminate or attenuate the inconsistent object advantage. As in Experiment 2, the change conditions were limited to left–right orientation change and same conditions.

Method

Participants. Twenty-four Michigan State University undergraduate students participated. All the participants had normal or corrected-to-normal vision, were naive with respect to the purposes of the research, and received partial credit for their participation. None had taken part in Experiment 1 or 2.

Stimuli. The initial and test scene stimuli were the same as in Experiments 1 and 2. Intraub (1984) demonstrated that using the same conceptual mask on each trial did not produce significant conceptual interference. Thus, for the conceptual mask condition, 10 different additional line drawings of scenes were created. Each additional scene was used as a conceptual mask for two of the original experimental scenes. Assignment of conceptual mask scenes to experimental scenes was determined randomly. None of the conceptual mask scenes was of the same conceptual type as an experimental scene.

Apparatus and Procedure. The apparatus was the same as in Experiments 1 and 2. A 250-msec mask was used to ensure that the participants could process the conceptual content of the conceptual mask. The 160 experimental trials were created by the within-participants factorial combination of 2 consistency conditions (consistent vs. inconsistent) × 2 change conditions (left–right orientation change vs. same) × 2 mask types (pattern vs. conceptual) × 20 scenes.

Results

***A′* analysis**. Mean A' as a function of consistency and mask type is presented in Table 3. There was a reliable main effect of consistency [$F(1,23) = 5.45$, $MS_e = 0.0057$, $p < .05$], with an advantage for inconsistent targets (.728) in comparison with consistent targets (.682). There was also a main effect of mask type [$F(1,23) = 13.11$, $MS_e = 0.089$, $p < .005$], with better performance in the pattern mask condition (.740) than in the conceptual mask condition (.670). There was no interaction between mask type and consistency ($F < 1$). The inconsistent object advantage was .032 in the pattern mask condition and .060 in the conceptual mask condition.

Percentage correct analysis. Mean percentage correct as a function of consistency and mask type is shown in

Table 3
Experiment 3: Mean A' and Mean Percentage Correct (% Correct) as a Function of Target Object Consistency and Mask Type

Mask Type			
Pattern		Conceptual	
A'	% Correct	A'	% Correct
Consistent			
.724	62.3	.640	57.7
Inconsistent			
.756	65.4	.700	62.4

Table 3. As in the A' data, there was a reliable effect of consistency [$F(1,23) = 7.99, MS_e = 91.66, p < .01$]. We again observed better performance in the inconsistent condition (63.9%) than in the consistent condition (60.1%). There was also a main effect of mask type [$F(1,23) = 8.25$, $MS_e = 84.10, p < .01$], with better performance in the pattern mask condition (63.9%) than in the conceptual mask condition (60.0%). However, there was no interaction between mask type and consistency ($F < 1$). The inconsistent object advantage was 3.1% in the pattern mask condition and 4.7% in the conceptual mask condition.

Discussion

The presentation of a conceptual mask between the initial and test scenes was disruptive: Change detection performance was reliably worse in the conceptual mask condition in comparison with the pattern mask condition. However, there was no interaction between mask type and consistency. In fact, the inconsistent object advantage was larger for the conceptual mask than for the pattern mask, though not reliably so. Thus, the results of this experiment fail to provide support for the CSTM hypothesis.

GENERAL DISCUSSION

In this study, we tested whether the inconsistent object change detection advantage can be explained by the differential retention of conceptual codes generated from a brief view of a scene. According to this CTSM hypothesis, conceptual codes for inconsistent objects receive priority when integrated with stored scene knowledge, and thus are more likely to be included in the evolving scene representation. None of the three experiments provided support for this hypothesis. In Experiment 1, changes that altered the conceptual content of the scene (deletion) did not exhibit a larger inconsistent object advantage than changes that did not alter conceptual content (orientation change). In Experiment 2, the inconsistent object advantage was no larger when ample time was provided for conceptual consolidation (400-msec mask) than when potentially insufficient time was available for consolidation (30-msec mask). In Experiment 3, the presentation of a conceptual mask between scenes should have disrupted conceptual processing of the initial scene relative to the presentation of a meaningless pattern mask (Potter, 1976). Although the conceptual mask did reduce detection performance in comparison with a pattern mask, it did not reduce the size of the inconsistent object advantage. The present experiments had sufficient sensitivity to observe modulations of the inconsistent object advantage, since that advantage was significantly larger for orientation changes than for deletions (Experiment 1). Thus, these results converge on the conclusion that the inconsistent object change detection advantage does not derive from differential conceptual processing during the short-term retention of scene information.

Before we discuss our preferred explanation for the inconsistent object advantage in terms of differential memory for visual object representations, it is necessary to eliminate one final alternative explanation. According to the *criterion modulation hypothesis* (Friedman, 1979; Palmer, 1975), the criterion for determining a match between the long-term memory representation of a consistent object type and current perceptual information is lowered for consistent objects. Thus, objects that are consistent with a scene should be identified more easily than inconsistent objects, since less perceptual information will need to be encoded to reach the identification threshold (Friedman, 1979; Masson, 1991; see also Johnston & Hawley, 1994). A criterion modulation model could produce the inconsistent object advantage in this study, under the additional assumption that perceptual analysis continues only until the criterion is reached and a match is found. More time will be required for inconsistent objects to reach criterion, and so more time will be devoted to constructing perceptual descriptions for those objects. Because a more completely elaborated object description would result when an object is inconsistent with its scene than when it is consistent, change detection based on a perceptual description would be facilitated for those objects, producing the inconsistent object advantage. The criterion modulation hypothesis predicts not only better encoding of perceptual detail for inconsistent objects, but also an advantage for the type-level identification of consistent objects. In other words, more perceptual detail will be encoded for an inconsistent object as a direct result of the fact that it is more difficult to identify. However, we have already demonstrated (using a similar set of stimuli) that when response bias is eliminated from object detection paradigms, there is no evidence that type-level identification is facilitated for consistent objects (Hollingworth & Henderson, 1998, 1999). Thus, the criterion modulation hypothesis cannot account for the inconsistent object advantage.

The failure of the CSTM hypothesis makes it likely that the inconsistent object advantage is caused by differential memory for visual object representations. Two strands of evidence directly support a visual locus for the inconsistent object advantage. First, the inconsistent object advantage has been observed in paradigms that require visual representation (such as orientation or token change detection). Second, Experiment 1 demonstrated that the inconsistent object advantage was actually larger for visual changes (orientation change) than for changes that did not require visual memory (deletion). Taken together, these data point toward a visual explanation. Thus, although scene context does not influence the initial identification of visual objects (Hollingworth & Henderson, 1998, 1999), it does appear to influence short-term memory for visual object representations. In other words, although perceptual identification appears to be functionally isolated from knowledge of typical scene composition, this isolation does not extend to visual memory.

The present evidence of context effects in visual memory corresponds to that of other recent studies demonstrating that VSTM is sensitive to contextual factors. First, VSTM for simple objects is influenced by the presence or absence of contextual elements. Jiang, Olson, and Chun (2000) presented an initial array of colored squares, followed by a short interstimulus interval and a test image in which a single target object either changed color or remained the same. This target object was presented in the test image either alone or within the original context of the other array objects. Change detection performance was more accurate when the context was maintained from study to test. In addition, recent studies on the inhibition of return phenomenon have demonstrated that short-term memory for object position is also context dependent. Signature effects of inhibition of return, such as slower perceptual judgments at previously attended locations (Posner & Cohen, 1984) or increased saccade latencies to previously attended objects (Abrams & Dobkin, 1994), are found only when the original context remains visible (Klein & MacInnes, 1999; Takeda & Yagi, 2000). Second, the spatial configuration and visual properties of contextual elements also influence short-term memory for objects. Jiang et al. found that the detection of changes to single objects was reduced when the spatial configuration of contextual objects or the color of those objects was changed. These results demonstrate that physical properties of the context influence short-term memory for visual object information.

Recent studies have demonstrated that participants can quite rapidly learn statistical regularities within complex arrays of visual objects, forming the type of scene-level visual representation that could potentially produce context effects on visual memory (Chun & Jiang, 1998, 1999; Fiser & Aslin, 2001). Chun and Jiang (1998) manipulated the spatial context of distractors in a repeated search task and found that target detection was speeded when the target was embedded within a repeated distractor configuration in comparison with a novel configuration of the same distractors. In addition, Chun and Jiang (1999) demonstrated that participants can also learn the co-occurrence relationships between different visual shapes in complex arrays. Such learning appears to allow participants to predict the presence or position of an object that has been consistently presented within a particular context. It is therefore possible that such learning during people's everyday experience with common environments leads to the formation of visual memory representations coding the co-occurrence relationship between scenes and particular objects as well as the spatial relationships between constituent objects (Chun, 2000).

Although this is a promising hypothesis for the source of scene context effects such as the inconsistent object advantage, the generalization from learning of abstract arrays in the Chun and Jiang (1998, 1999) studies to learning of statistical regularities within real-world scenes is not simple. In the Chun and Jiang studies, the repeated contexts perfectly predicted a target attribute. Such perfect (or at least nearly perfect) correlations may be found in repeated encounters with a particular scene token (the objects in one's office and the arrangement of those objects may not change much from one encounter to the next), but there is a much lower correlation between a scene type (offices in general) and any particular object (see Henderson, 1992). Consider the class of office scenes. Although a three-hole punch may be perfectly consistent within an office (and much more likely to be found in an office than in any other scene), a three-hole punch is actually fairly unlikely to be visible within any particular office. Office scenes as a class place even less constraint on the spatial relationships between objects: Pens and staplers may both typically appear on top of a desk, but the spatial relationship between a pen and a stapler is only loosely constrained at best. If visual statistical learning is to explain influences of general scene knowledge on visual object processing (Chun, 2000), then further evidence will be required to demonstrate that real-world scenes contain sufficient statistical regularity and that the human visual system is sensitive to that regularity.

Finally, if the inconsistent object advantage derives from differential availability of visual object representations, what mechanisms account for this difference in memory? In the long-term memory literature, researchers have generally concluded that the inconsistent object advantage derives from differences in initial encoding into memory: Inconsistent objects are more likely to be attended and/or are attended longer, and thus more accurate visual information is encoded from them (Friedman, 1979; Pedzek et al., 1989). An explanation in terms of initial encoding may account, in part, for the inconsistent object advantage, but such an explanation cannot account for the full range of effects. In the free-viewing paradigms of Hollingworth et al. (2001) and Friedman, inconsistent objects were fixated longer than consistent objects during study, suggesting that the inconsistent object advantage might derive from differences in attentional selection and initial memory encoding. However, in both these studies, a robust inconsistent object advantage remained when fixation time was controlled, demonstrating that context also influenced postencoding processes, such as the retention or retrieval of visual object information.

Experiments such as those of the present study, in which the initial scene is presented for 250 msec or less (Hollingworth & Henderson, 1998, 1999, 2000), do not allow for differential selection by means of eye movements. It is possible that attention was oriented covertly to inconsistent objects during the brief scene presentation, but initial evidence does not provide support for this hypothesis. We (Hollingworth & Henderson, unpublished data) conducted two experiments designed to examine the covert allocation of attention as a function of object consistency. In the first, a scene was presented for 300 msec, and participants then made a speeded discrimination judgment for a probe symbol ("%" or "&") that appeared either in the position of a consistent object or in the position of an inconsistent object. In the second, a consistent or inconsistent object shifted slightly either up or down after 250 msec of viewing, and participants reported the direction of movement. If inconsistent objects attract or hold attention covertly, then perceptual judgments about those objects (shift in position) or at those object locations (symbol discrimination) should be facilitated, yet they were not. Null effects of consistency were observed in both experiments, each with sufficient power to detect an RT effect of less than 12 msec. Finally, the fact that inconsistent objects do not attract the eyes during free viewing (Henderson et al., 1999, using a set of stimuli similar to those used in the present study) suggests that inconsistent objects do not attract attention, given strong evidence that attention and eye movements are functionally coupled during free viewing (see, e.g., Hoffman & Subramaniam, 1995; Shepherd, Findlay, & Hockey, 1986; Theeuwes, Kramer, Hahn, & Irwin, 1998). These results do not eliminate the possibility that differences in attentional allocation contribute to the inconsistent object advantage, but they suggest that it may be fruitful to explore other possible explanations, such as differences in the retention or retrieval of visual object representations.

In summary, in this study converging evidence from three manipulations was used to test a CSTM explanation for the inconsistent object change detection advantage in real-world scenes. Contrary to that explanation, it was found that the advantage was larger for visual than for conceptual changes, was unaffected by the duration of an intervening mask that appeared between presentation of an original scene and a changed scene and similarly was unaffected by the conceptual status of the intervening mask. Together, the results suggest that the locus of the inconsistent object advantage is more likely to be found in visual memory.

REFERENCES

Aaronson, D., & Watts, B. (1987). Extensions of Grier's computational formulas for A' and B' to below-chance performance. *Psychological Bulletin*, **102**, 439-442.

Abrams, R. A., & Dobkin, R. S. (1994). Inhibition of return: Effects of attentional cuing on eye movement latencies. *Journal of Experimental Psychology: Human Perception & Performance*, **20**, 467-477.

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

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.

Brewer, W. F., & Treyens, J. C. (1981). Role of schemata in memory for places. *Cognitive Psychology*, **13**, 207-230.

Chun, M. M. (2000). Contextual cueing of visual attention. *Trends in Cognitive Sciences*, **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.

Chun, M. M., & Jiang, Y. (1999). Top-down attentional guidance based on implicit learning of visual covariation. *Psychological Science*, **10**, 360-365.

Chun, M. M., & Potter, M. C. (1995). A two-stage model for multiple target detection in rapid serial visual presentation. *Journal of Experimental Psychology: Human Perception & Performance*, **21**, 109-127.

Fiser, J., & Aslin, R. N. (2001). Unsupervised statistical learning of higher-order spatial structures from visual scenes. *Psychological Science*, **12**, 499-504.

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

Goodman, G. S. (1980). Picture memory: How the action schema affects retention. *Cognitive Psychology*, **12**, 473-495.

Grier, J. B. (1971). Nonparametric indexes for sensitivity and bias: Computing formulas. *Psychological Bulletin*, **75**, 424-429.

Henderson, J. M. (1992). Object identification in context: The visual processing of natural scenes. *Canadian Journal of Psychology*, **46**, 319-341.

Henderson, J. M., & Hollingworth, A. (1999). High-level scene perception. *Annual Review of Psychology*, **50**, 243-271.

Henderson, J. M., Weeks, P. A., Jr., & Hollingworth, A. (1999). The effects of semantic consistency on eye movements during scene viewing. *Journal of Experimental Psychology: Human Perception & Performance*, **25**, 210-228.

Hock, H. S., Romanski, L., Galie, A., & Williams, C. S. (1978). Real-world schemata and scene recognition in adults and children. *Memory & Cognition*, **6**, 423-431.

Hoffman, J. E., & Subramaniam, B. (1995). The role of visual attention in saccadic eye movements. *Perception & Psychophysics*, **57**, 787-795.

Hollingworth, A. (2003). Failures of retrieval and comparison constrain change detection in natural scenes. *Journal of Experimental Psychology: Human Perception & Performance*, **29**, 388-403.

Hollingworth, A., & Henderson, J. M. (1998). Does consistent scene context facilitate object perception? *Journal of Experimental Psychology: General*, **127**, 398-415.

Hollingworth, A., & Henderson, J. M. (1999). Object identification is isolated from scene semantic constraint: Evidence from object type and token discrimination. *Acta Psychologica*, **102**, 319-343.

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

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. *Psychonomic Bulletin & Review*, **8**, 761-768.

Intraub, H. (1984). Conceptual masking: The effects of subsequent visual events on memory for pictures. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **10**, 115-125.

Irwin, D. E. (1992). Visual memory within and across fixations. In K. Rayner (Ed.), *Eye movements and visual cognition: Scene perception and reading* (pp. 146-165). New York: Springer-Verlag.

Jiang, Y., Olson, I. R., & Chun, M. M. (2000). Organization of visual short-term memory. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **26**, 683-702.

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

Klein, R. M., & MacInnes, W. J. (1999). Inhibition of return is a foraging facilitator in visual search. *Psychological Science*, **10**, 346-352.

Mandler, J. M., & Johnson, N. S. (1976). Some of the thousand words a picture is worth. *Journal of Experimental Psychology: Human Learning & Memory*, **2**, 529-540.

Masson, M. E. J. (1991). Constraints on the interaction between context and stimulus information. In K. J. Hammond & D. Gentner (Eds.), *Proceedings of the Thirteenth Annual Conference of the Cognitive Science Society* (pp. 540-545). Hillsdale, NJ: Erlbaum.

Palmer, S. E. (1975). The effects of contextual scenes on the identification of objects. *Memory & Cognition*, **3**, 519-526.

Pedzek, K., Whetstone, T., Reynolds, K., Askari, N., & Dougherty, T. (1989). Memory for real-world scenes: The role of consistency with schema expectations. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **15**, 587-595.

Posner, M. I., & Cohen, Y. (1984). Components of visual orienting. In H. Bouma & D. G. Bouwhuis (Eds.), *Attention and performance X* (pp. 531-556). Hillsdale, NJ: Erlbaum.

Potter, M. C. (1976). Short-term conceptual memory for pictures. *Journal of Experimental Psychology: Human Learning & Memory*, **2**, 509-522.

Potter, M. C. (1993). Very short-term conceptual memory. *Memory & Cognition*, **21**, 156-161.

Potter, M. C. (1999). Understanding sentences and scenes: The role of conceptual short-term memory. In V. Coltheart (Ed.), *Fleeting memories* (pp. 13-46). Cambridge, MA: MIT Press.

Shepherd, M., Findlay, J. M., & Hockey, R. J. (1986). The relationship between eye movements and spatial attention. *Quarterly Journal of Experimental Psychology*, **38A**, 475-491.

Takeda, Y., & Yagi, A. (2000). Inhibitory tagging in visual search can be found if search stimuli remain visible. *Perception & Psychophysics*, **62**, 927-934.

Theeuwes, J., Kramer, A. F., Hahn, S., & Irwin, D. E. (1998). Our eyes do not always go where we want them to go. *Psychological Science*, **9**, 379-385.

van Diepen, P. M. J., & De Graef, P. (1994). *Line-drawing library and software toolbox* (Psych. Rep. No. 165). Leuven: University of Leuven, Laboratory of Experimental Psychology.

NOTES

1. We use the terms *VSTM* and *CSTM* to distinguish between the format and the content of representation (visual as opposed to conceptual). We consider both hypotheses neutral with regard to whether short-term retention constitutes a separate memory store from long-term memory. In the visual domain, the format and content of VSTM is clearly different from early forms of iconic retention (see Irwin, 1992, for a review); it is in comparison with iconic memory that VSTM is defined as a different memory system. However, virtually no work has been conducted to investigate whether VSTM retention is different in format or content from visual representations retained over the long term, though existing evidence suggests that retention over the two time scales is supported by similar forms of representation (see Hollingworth, 2003, for further discussion). The fact that the inconsistent object advantage appears both in short- and long-term memory paradigms also suggests a close relationship between short- and long-term visual retention. Given these considerations, we use *short-term* to refer only to the temporal characteristics of the task.

2. For above-chance performance, A' was calculated as specified by Grier (1971):

$$A' = \frac{1}{2} + \frac{(y-x)(1+y-x)}{4y(1-x)},$$

where y is the hit rate and x the false alarm rate. In the few cases in which a participant performed below chance in a particular condition, A' was calculated using the below-chance equation developed by Aaronson and Watts (1987):

$$A' = \frac{1}{2} - \frac{(x-y)(1+x-y)}{4x(1-y)}.$$

(Manuscript received April 2, 2002;
revision accepted for publication April 4, 2003.)