Does scene context always facilitate retrieval of visual object representations?

Ryoichi Nakashima · Kazuhiko Yokosawa

Published online: 5 January 2011
© Psychonomic Society, Inc. 2010

Abstract An object-to-scene binding hypothesis maintains that visual object representations are stored as part of a larger scene representation or scene context, and that scene context facilitates retrieval of object representations (see, e.g., Hollingworth, *Journal of Experimental Psychology: Learning, Memory and Cognition*, 32, 58–69, 2006). Support for this hypothesis comes from data using an intentional memory task. In the present study, we examined whether scene context always facilitates retrieval of visual object representations. In two experiments, we investigated whether the scene context facilitates retrieval of object representations, using a new paradigm in which a memory task is appended to a repeated-flicker change detection task. Results indicated that in normal scene viewing, in which many simultaneous objects appear, scene context facilitation of the retrieval of object representations—henceforth termed *object-to-scene binding*—occurred only when the observer was required to retain much information for a task (i.e., an intentional memory task).

Keywords Visual representation · Scene context · Flicker paradigm

Normally, in scene viewing, it is impossible to perceive information about all objects that are visible simultaneously in a visual field. Instead, we attend serially on small regions in this field. This raises the following query: Are we forming detailed representations of scenes in visual memory

during scene viewings? Recent studies (e.g., Hollingworth & Henderson, 2002) suggested we retain robust visual memories of objects and scenes. Furthermore, robust representations of individual objects in scenes can be formed either intentionally or incidentally during scene viewing (Castelhano & Henderson, 2005).

In visual environments, an object is usually surrounded by other objects that together form a scene context. In turn, such context information appears to facilitate object recognition (e.g., Davenport & Potter, 2004; but see Hollingworth & Henderson, 1998). Analogously, this raises a second question: Does scene context facilitate the retrieval of object representations stored during viewing? In other words, is an object representation stored as part of a larger scene representation or is it stored independently of scene context? Recent research has shown that individual objects, which contribute to scene context, become bound to scene context (Hollingworth, 2006, 2007). In the present study, we refer to this as the *object-to-scene binding hypothesis*.

Support for the object-to-scene binding hypothesis has been based on experimental results obtained from intentional memory tasks.¹ A common intentional memory task

R. Nakashima (✉) · K. Yokosawa
Department of Psychology, Graduate School of Humanities
and Society, The University of Tokyo,
7-3-1 Hongo,
Bunkyo-Ku, Tokyo 113-0033, Japan
e-mail: rnaka@L.u-tokyo.ac.jp

¹ Hayes, Nadel, and Ryan (2007) indicated that object-to-scene binding occurs automatically. However, they had observers view images that included a prominent foreground object with a background and tested their memory for the objects presented either within the original background or in isolation. In contrast, in normal scene viewings, many objects exist simultaneously, and a prominent foreground object cannot be determined. Their suggestion may be appropriate only to cases in which a prominent foreground object can be determined. In the present study, we aimed to expand their suggestion to a normal scene memory. We discussed this further in the General Discussion section.

has participants view scene images for preparing following memory tests (e.g., same or different?). Typically, to perform well in these memory tasks, participants must encode and accumulate local information sequentially, while retaining as much scene information as possible during the scene viewing on a given trial. As a result, in these tasks, it is difficult to ascertain whether a retained visual object representation is inevitably bound to the scene context or is conditionally bound contingent upon task demands, as stipulated by explicit retention instructions. In other words, it is possible that some tasks encourage a chunking strategy wherein viewers create larger units of visual information about a scene to facilitate retention of accumulated information. Such task demands, which are characteristic of intentional memory tasks, may effectively lead to object-to-scene binding.

In the present experiments, we expanded on the object-to-scene binding hypothesis by testing whether scene context facilitates retrievals of object representations that are formed and retained in a repeated-flicker change detection task (e.g., Rensink, O'Regan, & Clark, 1997), instead of an intentional memory task. In a repeated-flicker change detection task, to detect a localized scene change, observers may encode and compare small regions of a scene; in principle, then, the tasks do not demand a serial accumulation of information of a whole scene. By contrast, the task demands associated with most intentional memory tasks do require some accumulation of scene information, meaning that such a memory task is likely to impose a greater memory demand for the whole scene than is the repeated-flicker change detection task (Varakin & Levin, 2008). Consistent with this distinction, it has been found that jumbling a coherent scene reduces performance in an intentional memory task (Varakin & Levin, 2008), but not in a repeated-flicker change detection task (Yokosawa & Mitsumatsu, 2003). This suggests that properties of object representations can be different in these tasks, with respect to the relationship between object and scene context. Therefore, if object representations are always, regardless of a memory demand, bound to a scene context, a scene context should facilitate the retrieval of object representations retained in a repeated-flicker change detection task.

Experiment 1

Our aim is to examine the properties of object representations retained in a repeated-flicker change detection task. However, before assessing this, we must verify an important premise that scene context facilitates object retrieval in spite of the fact that scene images are presented with flickering displays. It is necessary to confirm this because a flickering image presentation is unnatural for

observers. In [Experiment 1](#), instead of the static presentation of images of scenes, participants viewed flickering images of scenes with no object change. Following this flickering image sequence, a test display was appended. The test display contained a target object that was either the same as or different from an object in the preceding scene image. In addition, the test display presented the target embedded either in its original scene context (background present) or in isolation (background absent). The task was to judge whether a target object was the same as or different from the object presented in the preceding flickering sequence. If a visual object representation is bound to its scene context, then presence of the original scene context presented during a memory test should facilitate object retrieval in this task, leading to higher performance in the background-present condition than in the background-absent condition.

In the test display, a target object was either the same as it had been in the preceding scene image, or was changed in one of the following two ways: either an orientation change or a token change. The orientation change rotated a target 90° in depth. The token change replaced a target with another object from the same basic category.

A facilitation of object retrieval by scene context could occur when scene images were studied with or without verbal suppression (Hollingworth, 2006, 2007). To rule out verbal encoding, in [Experiment 1](#), we used an intentional memory task with a verbal suppression task. This task minimizes the possibility of facilitation due to verbal encoding and permits clearer assessment of effects due to properties of visual representations.

Method

Participants Twenty-nine undergraduate students with normal or corrected-to-normal vision participated.

Stimuli and conditions Sixty scene images were rendered from 3-D models of 60 different real-world environments. Each scene contained about 15 discrete objects. In each model, one target object was chosen. To create orientation and token change images, the target object was rotated 90° in depth or was replaced by another object token. The background-absent images were created by applying a transparent texture, which was a uniform gray, to all objects except the target object (Fig. 1).

Background conditions applied to the test display. A test display presented a target object either within the original background (background present) or in isolation (background absent). In the background-absent condition, the background-absent image was presented in the test display. In the background-present condition, a background-present image was presented that contained a gray cue circle

Fig. 1 A sample scene illustrating the object manipulations in this study. The upper (lower) row shows the background present (absent) stimuli. In the experiments, stimuli were presented in color

surrounding the target object. The cue was necessary in the background-present condition to specify the target object, as in the background absent condition (Hollingworth, 2006).

Scene stimuli subtended $24^\circ \times 18^\circ$ of visual angle at a viewing distance of 57 cm, maintained by a forehead rest. Target objects subtended 4.4° on average along the longest dimension in the picture plane.

Apparatus Presentation of stimuli and recording of responses were controlled by MATLAB software. Stimuli were displayed at a resolution of 800 x 600 pixels by 24-bit color on a 22-in. monitor.

Procedure Participants were told that they would be presented with flickering images of scenes and that after this, they would receive a memory test in which they had to decide whether a cued object was the same as or different from an object presented in flickering displays. Digit repetition was used for verbal suppression. Each trial began with a screen containing one randomly chosen digit, from 1 to 7. Participants were instructed to murmur three consecutive digits, starting from the presented digit, repeatedly, and to continue this repetition throughout the trial. Following the digit presentation (1,000 ms) and a blank display (500 ms), the original image was presented for 240 ms repeatedly, separated by a black screen for 100 ms. The total duration of the flickering presentation was 10.88 s, which, on the basis of a pilot experiment, was determined to be adequate to view the complete scene. After the presentation, the instruction “Press a Button” was displayed. When participants pressed a response button, a test display was presented.

The task was to judge whether the cued object in the test display was the same as or different from the object presented in flicker displays. This is a typical intentional memory task. Participants pressed one of two buttons to indicate that the cued object was either the same or different. Six types of images, 2 background types (background present with postcue, background absent) x 3 target object types (original object, orientation change object, token change object) could occur as a test display. At the start of each experiment, each stimulus set was randomly assigned to one of the six test display conditions. Trial order was determined randomly.

Results and discussion

We used A' as a measure of the memory performance. For each participant in each change condition, A' was calculated by using the hit rate (the rate of “same” response in the original object condition), and the false-alarm rate (the rate of “same” response in the change object condition).

Figure 2 shows the mean A' in each of the background and the object change conditions. In all conditions, A' values were above chance level (.5), $t_s > 5.02$, $p_s < .001$, demonstrating that object representations can be retained. The performance was better in the token change condition than in the orientation change condition, $F(1, 28) = 14.88$, $p < .001$. This is consistent with the expectation that orientation changes are more difficult to detect than token changes, because the orientation change does not affect object identity, whereas the token change does (Hollingworth & Henderson, 2002). The performance was better in

Fig. 2 Experiment 1: Mean A' as a function of background and change type. Error bars indicate standard errors

the background-present condition than in the background-absent condition, $F(1, 28) = 5.92, p < .03$. This is consistent with the object-to-scene binding hypothesis that scene context facilitates object retrievals (Hollingworth, 2006, 2007). There was no interaction, $F < 1$.

The results illustrate that object-to-scene binding occurs even when scene images to be remembered are presented in a flicker paradigm. Next, in Experiment 2, we addressed the main issue, namely whether or not a scene context facilitates retrieval of object representations retained when participants concentrate on a repeated-flicker change detection task; we used a modified change detection task which also assesses memory (as in Experiment 1).

Experiment 2

In Experiment 2, we considered one simple question: Does scene context facilitate the retrieval of object representations retained in a repeated-flicker change detection task? We used the typical repeated-flicker change detection task (Rensink et al., 1997) without a verbal suppression task. This change detection task, arguably, makes minimal demands for accumulation of scene information.

We used two types of flicker sequences. A one-change flicker sequence comprised alternations of an original image A and a modified image B displayed sequentially as A, A, B, B, A, A..., separated by blank screens (one-change trial). A no-change flicker sequence occurred on other trials; the no-change condition simply repeated A, A, A, A..., similar to Experiment 1. On both types of trials, a test display was appended after a flicker sequence. Participants were instructed to search for a scene change during a flicker sequence and to press a response button when they found a change. Then, they had to judge whether a cued object in the test display was the same as or different from that in the preceding sequence. In one-change trials, because the cued object in the test display was always the

changing object that they had to detect in the preceding flickered sequence, response to the test display was very easy. This encouraged participants to concentrate on searching for a scene change during flicker sequence.

The rationale for this design was to create a situation in which participants behaved as if they had to concentrate on searching for a scene change. Our primary interest, however, was in examining memory performance for the no-change flicker sequences. One-change and no-change trials occurred randomly within a trial block. However, on no-change trials, participants who could not discover an object change would be forced to rely on their memory of the preceding flickering scene because they could never know the cued object before a test display was presented.

To determine effects of scene context at the memory test, the two background conditions (background present, background absent) were used for the test displays as in Experiment 1. In order to assess memory with cued retrieval, we focused only on the no-change trials. On the trials, if object representations are bound to the scene context regardless of task demands, performance should be better in the background-present condition than in the background-absent condition.

Method

Participants Thirty-four undergraduate students with normal or corrected-to-normal vision participated; none had participated in Experiment 1. Data from six participants were discarded because they reported finding changes in many no-change trials (in more than 10 out of 13–44 trials; mean = 21.67 trials).

Apparatus The apparatus was identical to that in Experiment 1.

Stimuli and conditions The set of scene stimuli was expanded from 60 in Experiment 1 to 90, reflecting the experimental manipulation. In this experiment, 30 trials were one-change trials in which the original and change images were presented alternately in a flickering presentation; 60 trials were no-change trials in which the original image was repeated in a flickering sequence.

Procedure Participants were not informed beforehand that two types of trials (one-change and no-change trials) were mixed in a block (Fig. 3). In one-change trials, in which images changed in the flicker sequence, the original image and the change image—either orientation or token change image—alternated until the response or at most for 10.88 s. In the no-change trials, in which the original image repeated, the flicker sequence lasted 10.88 s, the same as in Experiment 1. In each trial, each image and blank durations

Fig. 3 A sample trial in one-change trials and no-change trials. In both trials, the images were presented with flicker, and the test displays were the same. In the test display, one image was chosen and presented in each trial

were the same as in [Experiment 1](#): an initial fixation (500 ms) and blank display (500 ms).

Participants were told to search for a change in each flickering image presented and to press the response button as soon as they found a change. If they could not find a change within the time limit, the instruction “Press a Button” was displayed. After they pressed the response button, a test display was presented.

The test display and the method of response required for the test display were identical to those in [Experiment 1](#) (i.e., respond “Same” or “Different”). Participants were informed that once a change was found and responded to, the cued object in the test display would be the changing object in the flickering presentation, which would be a very easy task. For example, in a given one-change trial, an original image and a token change image were presented alternately during flicker presentation. If an orientation change image is presented in the test display, participants should judge the object as “different;” otherwise, they should judge it as “same.” Participants were also informed that a cued object in the test display was chosen randomly in the scene if they failed to find a change within the time limit. In this case (the no-change trials), they were required to answer a memory test (as in [Experiment 1](#)). At the start

of each experiment, the stimulus set was randomly assigned to one of the 12 conditions (2 trial types x 6 test display conditions). Trial order was determined randomly.

Results and discussion

Detection performance in one-change trials We considered a trial in which participants responded correctly in the test display as correct change detection. The mean percentage of correct change detection on all one-change trials was 88.2% ($SD = 6.93$). Data were divided into the two background conditions in the test display. Accuracies were not different between two background conditions, $F < 1$. The mean reaction time for correct change detection was 5.00 s ($SD = .628$). This is much shorter than the time limit for a flickering presentation (10.88 s). These results suggest that participants indeed concentrated primarily on searching for a change in the flicker sequences.² We did not analyze data of the trials in which participants could not find a

² Although some participants conjectured that there were some no-change trials, no one reported that they primarily memorized a scene.

change within time limit, because the trials were very few ($M = 3.92$; $SD = 1.99$).

Memory performance in no-change trials The accuracy on no-change trials was of central interest. In this analysis, data of the trials in which participants responded within the time limit were removed (0–4 trials for each participant; $M = 1.46$ trials). Figure 4 shows the mean A' for each of the background and change conditions. In all conditions, the A' values were above chance, $t_s > 3.87$, $p_s < .001$. The performance was better in the token change condition than in the orientation change condition, $F(1, 27) = 7.43$, $p < .02$, replicating the result of Experiment 1. However, in contrast with Experiment 1, there was no significant difference between two background conditions, $F(1, 27) = 1.45$, $p > .2$, indicating that scene context does not affect the object retrievals. There was no interaction, $F < 1$.

Scene context did not facilitate the object retrieval in this experiment. This cannot be the result of a floor effect, because the performances in all conditions were above chance, and the performances in the token change conditions were higher than in the orientation change conditions. The performances in the token change conditions, at least, were not low; thus, this result is not derived from a floor effect.

The object-to-scene binding hypothesis (Hollingworth, 2006, 2007) was not replicated in this experiment. In this experiment, the participants' main task was a repeated-flicker change detection task, which alone does not require participants to accumulate much scene information (i.e., whole scene information) for the task (Varakin & Levin, 2008). Therefore, object-to-scene binding does not always occur.

General discussion

Recent studies (Hollingworth, 2006, 2007) have suggested that objects in scenes are bound to scene context in

memory, and scene context facilitates object retrieval. Our present study considered one simple question related to this suggestion: Does scene context always facilitate the object retrieval? In Experiment 1, we confirmed that the object-to-scene binding hypothesis was supported even when studied images were presented in a flicker paradigm.³ In Experiment 1, the primary task was a memory task that required participants to remember information about whole scenes. In Experiment 2, we examined whether scene context facilitates object retrieval in a repeated-flicker change detection task. The primary task in Experiment 2 was repeated-flicker change detection, which does not require the accumulation of scene information. We failed to find the evidence of object-to-scene binding. Taken together, these indicate that object representations are not always bound to the scene context; instead, they are bound only when the retention of much information is necessary (i.e., an intentional memory task).

Why does object representation bind to the scene context only when observers have to retain much information for the task? As was described in the introduction, this effect can result from observers' strategies of memorization. In intentional memory tasks, to retain much information effectively, observers use a strategy to create chunks of information. This can result in object-to-scene binding.

This raises questions regarding the difference between our findings and those reported by others. In particular, Hayes et al. (2007) suggested that object-to-scene binding can occur automatically, whereas the present results do not suggest this. Before discussing this, we must first consider the difference between the cases in which scene context facilitates object recognition (Davenport & Potter, 2004) and those in which scene context affects only an observer's response bias (Hollingworth & Henderson, 1998). Davenport and Potter (2004) indicated that scene background usually facilitates object recognition, but the effect of scene context may be lost when a target object is one of the objects in a scene, because observers could not know which object would be tested beforehand.

We now return to the question concerned with differences between findings reported by Hayes et al. (2007) and those of this study. The stimuli used by Hayes et al. always included a prominent foreground object, similar to those of Davenport and Potter (2004), whereas stimuli of the present study consisted of many objects, much like those used by Hollingworth and Henderson (1998). That

Fig. 4 Experiment 2: Mean A' as a function of background and change type on no-change trials. Error bars indicate standard errors

³ The background facilitation was replicated in an additional experiment ($N = 16$) in which a verbal suppression task was not conducted and the number of trials was 90. Accuracy was higher in the background-present condition ($A' = .84$) than in the background-absent condition ($A' = .80$), $F(1, 15) = 5.94$, $p < .03$. Therefore, the main reason for the difference between Experiments 1 and 2 must not have been either a verbal suppression or the number of tasks.

is, when observers view scene images containing many objects without heavy demands for accumulating information, object representations cannot be stored as larger scene representations. In contrast, when the target object is prominent in the studied image (as with Hayes et al., 2007), scene context facilitation of object retrieval are likely to occur, because the relationship between a salient foreground object and its surrounding background should be perceived. In this case, background information can work as an aid for perception or retrieval of a foreground object.

In all experiments in the present study, biases toward “same” response were higher in the background-present condition than in the background-absent condition, $F_s > 6.00$, $p_s < .03$. This indicates that participants tended to assume that they had viewed the target object when it was presented in the original scene context in the test display. This is similar to the results of Hollingworth and Henderson (1998). Therefore, scene context potentially has an effect to bias observers’ judgments and sometimes facilitates both object recognitions and object retrievals. The condition of scene context facilitation for object retrievals can be either an intentional scene memorization (i.e., accumulation of scene information) or the presentation of a prominent target object in a studied image.

In summary, we suggest that, in normal scene viewings, object-to-scene binding occurs only when the explicit retention of much information is necessary for the task. The results speculate that, in normal scene perception in which many objects exist, the effect of scene context can be very similar for object recognition (Hollingworth & Henderson, 1998) and object retrieval. This phenomenon can be explained in terms of the potential bias effect and facilitation effect.

Author note This study was supported by a grant from the Research Fellowship of the Japan Society for the Promotion of Science for Young Scientists to R.N., and by a Grant-in-Aid for Scientific Research from the Japan Society for the Promotion of Science awarded to K.Y. We thank Edward Awh and two anonymous reviewers for helpful comments on the manuscript.

References

- Castelhano, M. S., & Henderson, J. M. (2005). Incidental visual memory for objects in scenes. *Visual Cognition*, 12, 1017–1040.
- Davenport, J. L., & Potter, M. C. (2004). Scene consistency in object and background perception. *Psychological Science*, 15, 559–564.
- Hayes, S. M., Nadel, L., & Ryan, L. (2007). The effect of scene context on episodic object recognition: Parahippocampal cortex mediates memory encoding and retrieval success. *Hippocampus*, 17, 873–889.
- Hollingworth, A. (2006). Scene and position specificity in visual memory for objects. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 32, 58–69.
- Hollingworth, A. (2007). Object-position binding in visual memory for natural scenes and object arrays. *Journal of Experimental Psychology: Human Perception and Performance*, 33, 31–47.
- 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. (2002). Accurate visual memory for previously attended objects in natural scenes. *Journal of Experimental Psychology: Human Perception and Performance*, 28, 113–136.
- 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.
- Varakin, D. A., & Levin, D. T. (2008). Scene structure enhances change detection. *The Quarterly Journal of Experimental Psychology*, 61, 543–551.
- Yokosawa, K., & Mitsumatsu, H. (2003). Does disruption of a scene impair change detection? *Journal of Vision*, 3, 41–48.