The Quarterly Journal of Experimental Psychology
Publication details, including instructions for authors and subscription information:
http://www.informaworld.com/smpp/title~content=t716100704

Object appearance and picture-specific viewpoint are not integrated in long-term memory

D. Alexander Varakin [a];Lester Loschky [b]
[a] Knox College, Galesburg, IL, USA [b] Kansas State University, Manhattan, KS, USA

To cite this Article Varakin, D. Alexander andLoschky, Lester(2010) 'Object appearance and picture-specific viewpoint are not integrated in long-term memory', The Quarterly Journal of Experimental Psychology, 63: 6, 1181 — 1200, First published on: 13 October 2009 (iFirst)

To link to this Article: DOI: 10.1080/17470210903254639

URL: http://dx.doi.org/10.1080/17470210903254639

THE QUARTERLY JOURNAL OF EXPERIMENTAL PSYCHOLOGY
2010, 63 (6), 1181–1200

Psychology Press
Taylor & Francis Group

Object appearance and picture-specific viewpoint are not integrated in long-term memory

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

Lester Loschky
Kansas State University, Manhattan, KS, USA

Previous work has demonstrated that visual long-term memory (VLTM) stores detailed information about object appearance. The current experiments investigate whether object appearance information in VLTM is integrated within representations that contain picture-specific viewpoint information. In three experiments using both incidental and intentional encoding instructions, participants were unable to perform above chance on recognition tests that required recognizing the conjunction of object appearance and viewpoint information (Experiments 1a, 1b, 2, and 3). However, performance was better when object appearance information (Experiments 1a, 1b, and 2) or picture-specific viewpoint information (Experiment 3) alone was sufficient to succeed on the memory test. These results replicate previous work demonstrating good memory for object appearance and viewpoint. However the current results suggest that object appearance and viewpoint are not episodically integrated in VLTM.

Keywords: Picture memory; Visual memory; Scene representation; Recognition memory.

People are very good at recognizing previously viewed pictures. In one demonstration of this ability, observers studied over 2,500 pictures and 3 days later were 90% accurate on a forced-choice recognition test for a subset of the studied scenes (Standing, Conezio, & Haber, 1970; see also Shepard, 1967; Standing, 1973). Despite decades of research, the nature of the representations that support such accurate performance on tests of visual long-term memory (VLTM) is still relatively poorly understood. One basic question that has yet to be adequately addressed is how information obtained from a picture is organized in VLTM vis-à-vis the picture itself. That is, there is still no clear answer regarding the degree to which scene-level representations in VLTM maintain the episodic details of pictures *as they were viewed*.

On the one hand, it is widely agreed that mental scene representations are not as detailed and stable as physical pictures. There is clear evidence that limits of visual acuity, attention, and working memory capacity restrict the amount of information that an observer can acquire at any given point in time (for reviews from various

---

Correspondence should be addressed to D. Alexander Varakin, Department of Psychology, Knox College, 2 E South St., Box K60, Galesburg, IL 61401, USA. E-mail: avarakin@knox.edu

Support for this research was provided by funds from Knox College awarded to D.A.V. We thank Jim Barrett, Christina Pfaff, Timothy J. Rairdon, and Tom Von Geldern for assisting with data collection. We also thank Stephen Killingsworth and two anonymous reviewers for commenting on earlier versions of this manuscript.

http://www.psypress.com/qjep
DOI:10.1080/17470210903254639

perspectives see Henderson & Hollingworth, 2003; Intraub, 1997; O'Regan, 1992; Rensink, 2002; Simons & Levin, 1997). Nevertheless, there has been some debate about the fate of episodically detailed visual representations after attention and visual working memory have been withdrawn from an object. Some theories posit that representations in VLTM are sparse. According to this view, the visual–cognitive system does not maintain scene-level representations that are "everywhere dense" (Intraub, 1997, p. 218) or "everywhere detailed and coherent" (Rensink, 2000, p. 17). An abstract representation of a scene's layout might be retained (e.g., Friedman & Waller, 2008); however, this representation would not contain detailed information about constituent objects' visual appearances (Rensink, 2000). These *sparse representation* theories can explain why people perform so well on long-term picture memory tests, because in many classic demonstrations of excellent picture memory it would have been possible to discriminate studied items from distractors on the basis of layout and/or conceptual gist. However, recent claims about VLTM contradict the claim that scene representations in VLTM are exceedingly sparse. In particular, *visual memory theory* (Henderson & Hollingworth, 2003; Hollingworth & Henderson, 2002) proposes that as attention is serially deployed to objects in a scene, detailed information about objects' appearances is indexed within a master map of locations and stored first in visual short-term memory (VSTM) and then in VLTM, effectively forming a relatively detailed scene-level representation (albeit, not a "sensory" copy).

Are these two proposals inconsistent with one another? As Peterson and Rhodes (2003) note, "One investigator's 'relatively detailed' representation may be another investigator's 'relatively sparse' representation" (p. 16). The purpose of the present experiments is to further explicate in what sense scene representations in VLTM are detailed. Such an explication seems necessary, because as reviewed below, much of the evidence that has been taken to support one view over the other is quite ambiguous.

A great deal of the evidence for the sparse view of scene representation has come from the literature on *change blindness* (CB), which is the difficulty observers have detecting large visual changes that occur simultaneously with some kind of visual or attentional disruption (for a review see Rensink, 2002; Simons & Rensink, 2005). CB can even occur for objects that have been attended (Levin & Simons, 1997; Varakin & Levin, 2006), so it clearly indicates a failure to use visual information to monitor the consistency of objects over time. However, it has been pointed out numerous times that a failure to detect a change to an object does not logically imply that an observer has a sparse memory for the object (e.g., Simons, 2000; Simons & Rensink, 2005). That is, CB may underestimate the extent to which episodically detailed visual information is retained in VLTM, and it therefore provides, at best, ambiguous evidence with respect to questions concerning the nature of VLTM.

Indeed, it is by now well established that fairly detailed information about object appearance can be reliably stored in VLTM. Many experiments have demonstrated that people perform quite well on two-alternative forced-choice (2-AFC) memory tests in which information about an object's visual appearance (e.g., token-level information of a basic-level object) is necessary to discriminate a studied picture from a distractor (e.g., Brady, Konkle, Alvarez, & Oliva, 2008; Hollingworth, 2004; Hollingworth & Henderson, 2002; Hollingworth, Williams, & Henderson, 2001). People can even succeed on memory tests for object appearance when the test is a surprise (Castelhano & Henderson, 2005; Varakin & Levin, 2006). Evidence of this sort has been used to argue against the view that scene representations in VLTM are sparse and in favour of the view that they are relatively detailed.

However, just as CB underestimates what is retained in VLTM, object appearance memory tests might overestimate the extent to which visual details in VLTM are integrated into an episodically detailed representation of a picture as it was viewed. When an observer is presented with two pictures in a forced-choice memory test, if the only difference between a studied picture and a never-before-viewed distractor is the visual appearance of a single object, then the observer only has to use *one* visual detail of the critical object in order to respond accurately. Of course, one might argue that the critical details, if selected randomly, are representative of the memory for details in the picture as a whole, and that memory for details therefore represents memory for the whole picture. But this raises the Gestalt psychologist's rhetorical question: Is the sum of the parts equal to the whole? We think not. People may have a good memory for many individual details about previously viewed scenes, but this does not necessarily mean they know which details originally appeared in which scene with which other details.

Thus, the general idea underlying so-called sparse representation theories—that is, that scene representations are not episodic instantiations of pictures as they were viewed—is not fatally compromised by evidence that object appearance information is reliably stored in VLTM. It could be the case that scene representations in VLTM are detailed but fragmented, or at least fragmentable. Definitively ruling out the position that scene representations are fragmented requires evidence that detailed object appearance information is episodically integrated with other information from a source picture, such as layout (e.g., Sanocki, 2003; Sanocki, Michelet, Sellers, & Reynolds, 2006) or viewpoint (e.g., Hock & Schmelzkopf, 1980), and that the result of such integration is maintained in VLTM.

In fact, there is no clear evidence on the question of whether detailed object appearance information is or is not episodically integrated with other specific information about a picture in VLTM. Before proceeding, it is important to note that there are two ways that multiple pieces of information might be stored together in memory. The first way that two pieces of information may be stored together is through *integration*. If two pieces of information are integrated in memory, it means that they are treated as if they are a single piece of information—that is, they are "fused" (Hayes, Nadel, & Ryan, 2007). Thus, if two pieces of information are integrated, then accessing one piece of information necessarily means accessing the other, and losing access to one necessarily means losing access to the other. For example, hue and saturation are integrated in colour perception (e.g., Burns & Shepp, 1988). The second way in which two pieces of information may be stored together is through *association* (Hayes et al., 2007). If two pieces of information are associated in memory, it means that at least one of the pieces of information "points" to the other piece. However, the two pieces of information can be functionally separated, so forgetting one does not necessarily entail forgetting the other, and remembering one does not necessarily entail remembering the other (for a discussion of related ideas, see Murnane, Phelps, & Malmberg, 1999).

There is good evidence that object appearance information is associated with other specific information about a picture in VLTM. For example, using retention intervals of 40 s to a couple of minutes, both Hayes et al. (2007) and Hollingworth (2006) demonstrated that performance on an object appearance recognition test is better when the object is presented in the same visual context at both study and test than when the visual contexts are different at study and test. These results clearly demonstrate that pictorial context can be an effective retrieval cue for object appearance information in VLTM (Hayes et al., 2007; Hollingworth, 2006). Thus, object appearance information and aspects of visual context are at least associated in memory.

However, these experiments are ambiguous with regard to such object appearance information being episodically integrated with other information encoded from a particular picture. The reason for this ambiguity is that association alone may be sufficient for one type of information (e.g., scene context) to serve as an effective retrieval cue for another type of information (e.g., object appearance)—integration is not necessary. Indeed, Hanna and Remington (1996) have argued that colour can be an effective retrieval cue for shape information, even though colour and shape are not integrated in long-term memory. In their experiments, observers studied arrays of geometric shapes, each shape having a different colour. On a subsequent memory test, performance was better when the test array was presented in colour than when it was presented in black and white. However, reassignment of specific colour–shape conjunctions from study to test did not attenuate colour's ability to provide retrieval cues. Thus, Hanna and Remington's experiments demonstrate that one kind of information can provide retrieval cues for another kind of information even when the two kinds of information are not integrated.

In contrast, Hayes et al. (2007) interpret results from a functional magnetic resonance imaging (fMRI) study as evidence that objects are integrated with context. Specifically, they demonstrated that brain activity in the parahippocampal cortex (PHC) was greater during retrieval of objects that were studied as part of a scene and subsequently recognized in isolation than during retrieval of objects that were encoded and retrieved in isolation. The PHC contains the parahippocampal place area (PPA), which is thought to be important for processing scene-level information (e.g., Epstein & Kanwisher, 1998; O'Craven & Kanwisher, 2000). Thus, Hayes et al. interpreted the above result as evidence that objects are episodically integrated with scene context in VLTM. However, using Hanna and Remington's (1996) logic, these data are also ambiguous, because Hayes et al.'s (2007) study did not include the reassignment of objects and contexts. Thus, it is not clear to what extent the observed fMRI activity reflected an accurate representation of the conjunction of specific object information with specific context information, or simply reflected participants' general knowledge that an object had originally been studied in one of the contexts, but not necessarily a representation of the specific context in which the object actually appeared.

This critique raises the question: Can observers accurately distinguish between previously viewed and novel conjunctions of scene information? The results of previous work by Hock and Schmelzkopf (1980) suggest that the conjunction of visual features that constitute a viewpoint can be remembered for at least several minutes (the time-course of interest for current purposes, and within the range of VLTM). In their Experiment 2, participants studied a set of pictures that were different viewpoints of the same larger scene. In one condition, participants' ability to recognize the studied pictures was assessed a few minutes after study. Critically, the distractor pictures on the recognition test were viewpoints of the larger studied scene that were in between previously studied viewpoints (e.g., studied viewpoints of 0° and 60° would have a distractor of 30°). In other words, the distractor pictures contained information that participants had viewed before, but not within the same picture. Thus, responding accurately on the memory test required discriminating between novel and previously studied conjunctions of visual features that constitute a viewpoint. When the memory test was administered several minutes after study, when the information would be in VLTM, accuracy was 74% (though after a 1-week delay, it fell to 50.5%). The fact that participants could discriminate novel viewpoints from studied viewpoints in this design suggests that the features of a picture that define viewpoint can be recognized in conjunction. However, these experiments did not assess memory for other kinds of information, such as object appearance.

In summary, the extent to which information in VLTM is episodically integrated into representations of pictures as they were viewed remains unclear. There is good evidence that detailed object appearance information can be retained in VLTM; however, good memory for the details of single objects does not imply that observers remember where those details came from. Recent evidence suggests that object appearance information and aspects of pictorial context are associated in VLTM; however, these studies have not convincingly demonstrated that object-appearance representations are episodically integrated with other information about a picture in VLTM.

THE CURRENT STUDY

The question the current study addresses is whether performance on object appearance memory tests overestimates the extent to which scene representations in VLTM maintain the episodic details of pictures *as they were viewed*. Specifically, the current experiments tested whether object appearance and picture-specific viewpoint information are episodically integrated in VLTM. To this end, we constructed 2-AFC tests in which previously viewed pictures had to be discriminated from distractors on the basis of the conjunction of two pieces of information: an object's visual appearance and information specific to the viewpoint of the picture in which the object originally appeared (Experiments 1a, 1b, 2, and 3). We compared performance in these cases to conditions in which previously viewed pictures could be selected on the basis of object appearance information alone (Experiments 1a, 1b, and 2), or picture-specific viewpoint information alone (Experiment 3).

The basic assumption underlying this method is as follows: If two pieces of information are integrated in long-term memory, it should be possible in a recognition memory test to discriminate between old and new conjunctions of previously viewed features (based on the definition of integrated representations; see Hanna & Remington, 1996; Hayes et al., 2007; Murnane et al., 1999). Thus, if scene representations in VLTM maintain episodic integrations of object appearance and picture-specific viewpoint information (to the extent that each is remembered at all), then participants should be able to accurately discriminate old and new conjunctions of object appearance and viewpoint information on a 2-AFC test. However, if scene representations in VLTM are fragmented or fragmentable, then participants might be able to select previously viewed pictures on the basis of object appearance information or viewpoint information, but not necessarily the conjunction of both.

EXPERIMENT 1A

Experiments 1a and 1b were designed to test the hypothesis that the conjunction of object appearance and picture-specific viewpoint is automatically maintained in VLTM. Thus, participants were given a cover task to perform during an initial study phase and were not told about the memory test at the beginning of the experimental session. The cover task served two purposes: (a) to give participants a task to do while incidentally learning the pictures, and (b) to ensure that participants consciously attended to the critical objects. Thus, in the cover task, participants were asked to respond each time an object in a picture was highlighted (by an increase in luminance), while withholding responses whenever empty locations were highlighted. Participants viewed several pictures, some of which were different viewpoints of the same larger scene (see Figure 1a versus Figure 1b). The complementary viewpoints overlapped, so some of the objects in the larger scene were visible in both

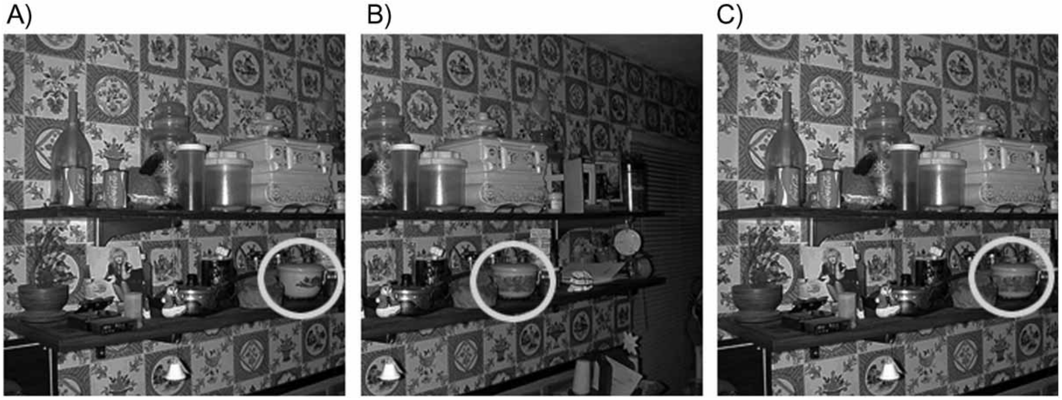

Figure 1. *Example of stimuli. A and B have different viewpoints and critical objects (in the rings, which were not visible in the actual stimuli); A and C have the same viewpoint and different critical objects; B and C have different viewpoints and the same critical object.*

viewpoints. However, one object's visual appearance was manipulated across the viewpoints (henceforth the *critical object*; see Figures 1a and 1b). On a subsequent 2-AFC test (administered after the study phase), participants had to choose between previously viewed pictures (e.g., Figure 1a) and distractors that depicted novel conjunctions of previously studied object appearance and picture-specific viewpoint information (e.g., Figure 1c, when observers had studied Figures 1a and 1b).

Method

Participants

A total of 31 undergraduate students at Knox College volunteered for this experiment in exchange for course credit. A total of 8 participants were dropped from all analyses—4 participants failed to perform the cover task, and an instrumentation error led to dropping 4 participants, resulting in a final sample size of 23 participants.

Apparatus

Stimuli were presented on 15″ monitors set at a refresh rate of 75 Hz and a resolution of 1,024 × 768 pixels. SuperLab 4 (Cedrus Corporation, San Pedro, CA) loaded on Mac Mini computers controlled stimulus presentation and saved responses.

Stimuli

All of the studied pictures consisted of the left- or rightmost 70% of larger digital photographs. A total of 8 pairs of digital photographs (drawn from the set of photographs used in Varakin & Levin, 2008) depicting indoor and outdoor environments were used. Members of a given pair of (larger) digital photographs were exactly the same except for the features of a single object (the critical object). The features that could differ were the critical object's colour (3 scene pairs), the non-shape-defining patterns on the surface of the object (3 pairs), or both (2 pairs). The shape (defined in terms of pixels) of each version of the critical object in a given scene was exactly the same, ensuring that the local context (i.e., the context that intersected the critical object) was also exactly the same.

To create the stimuli for the experiment, the left- or rightmost 189 × 480 pixels were cropped from the original 640 × 480 photographs, creating two translation-defined viewpoints consisting of the left- and the rightmost 451 × 480 pixels (~13.9 × 14.8 cm), as illustrated in Figure 1. For a given viewpoint, 42% of the image pixels were unique context, and 58% were exactly the same as the complementary viewpoint, including the critical objects' immediate local context.

For each viewpoint, in addition to the critical object (width ranging from ~1.3–5.5 cm, height from ~1.5–4.5 cm), three noncritical objects (width ranging from ~0.2–7.6 cm, height from ~0.7–8.5 cm) and four empty locations were semiarbitrarily chosen as response relevant for the cover task. To highlight an object or empty location the brightness of the response-relevant object/location was increased (average increase of approximately 6.36 cd/m^2; see Figure 2). Across complementary viewpoints, the noncritical objects and empty locations did not overlap so that participants would attend to different sets of objects in cases where both viewpoints were viewed. For example, if one of the plastic containers that is visible in Figures 1a and 1b (above and to the left of the critical object) was cued in the cover-task for the viewpoint in Figure 1a, it would not be cued in the cover task for the complementary viewpoint in Figure 1b. This was done to give participants a richer set of cues to use in discriminating differing viewpoints of a scene on the 2-AFC. Note that the location of the critical objects differed only in terms of physical pixel coordinates—they were the same in terms of the scene-centred relative location coordinates. Thus, having participants attend to different objects and different relative locations across complementary viewpoints could add additional cues to help participants determine which object appeared in which viewpoint.

Procedure and design

During the cover-task phase, participants were asked to press a button whenever an object in the

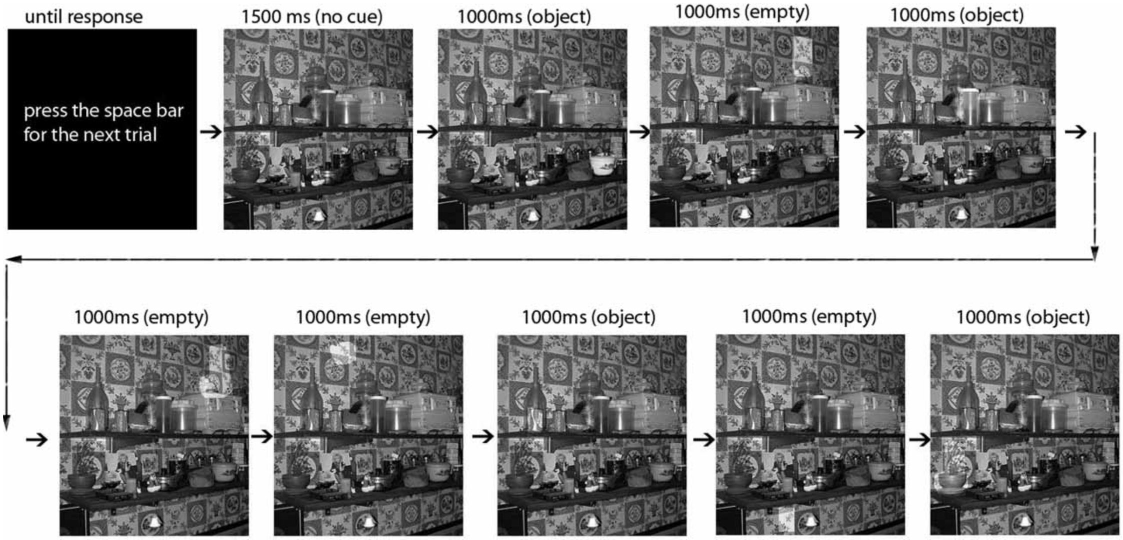

Figure 2. *Diagram of a cover task trial. In each frame, the brightness of a single object or empty location is increased relative to the preceding frame. The participants' task was to respond whenever an object "lit up" but to withhold responding when empty locations "lit up".*

picture "lit up", but to withhold responding when an empty location lit up (see Figure 2). Each trial began with a screen that read "press the space bar for the next trial". When the trial was initiated, a picture was presented, and after 1,500 ms the cueing sequence began. In the cueing sequence, the objects and empty locations were sequentially cued one at a time for 1,000 ms per cue (order was randomized); thus total viewing time for each picture was 9,500 ms. There were a total of 15 cover task trials. Immediately after Cover Task Trial 15 was completed, participants were given instructions for the 2-AFC memory test. There were 8 recognition test trials. On each trial, a previously viewed picture and distractor were presented side by side on the computer monitor (separated by about 15.5 cm centre to centre). The old picture and the distractor were exactly the same except for the appearance of a single object (i.e., the critical object, described in the *Stimuli* section above). The critical object in both choices was brighter than the background context (as in the frames on the cover task when the critical objects were cued), in order to cue participants' attention to it. Whether the distractor appeared on the left or right on each trial was randomly determined with equal probability. In the 2-AFC, participants used a mouse to select the "old" picture that they had viewed previously.

It is important to note that during the cover task phase, scenes were not presented in a completely random order. Figure 3 shows a schematic of the trials in the cover task and its relation to the 2-AFC test that followed. The picture presented on Cover Task Trial 1 was a filler picture and was treated as a practice trial, although participants were not made aware of this (the same filler pictures were used for all participants). In Cover Task Trials 2–9, the target pictures, which would appear as correct options in the (subsequent) 2-AFC recognition test, were presented (in a random order). On Trial 10, the same filler picture as that used in Trial 1 was presented again. There were two purposes of the second filler: (a) to ensure that *donor* pictures (described in the next sentence) did not appear one after the other, and (b) to expose observers to a picture that they had viewed previously, so that they would not be surprised when donor pictures appeared in Trials 11–14. Donor pictures depicted the complementary viewpoint of one of the scenes from Trials 2–9 and also contained the alternative

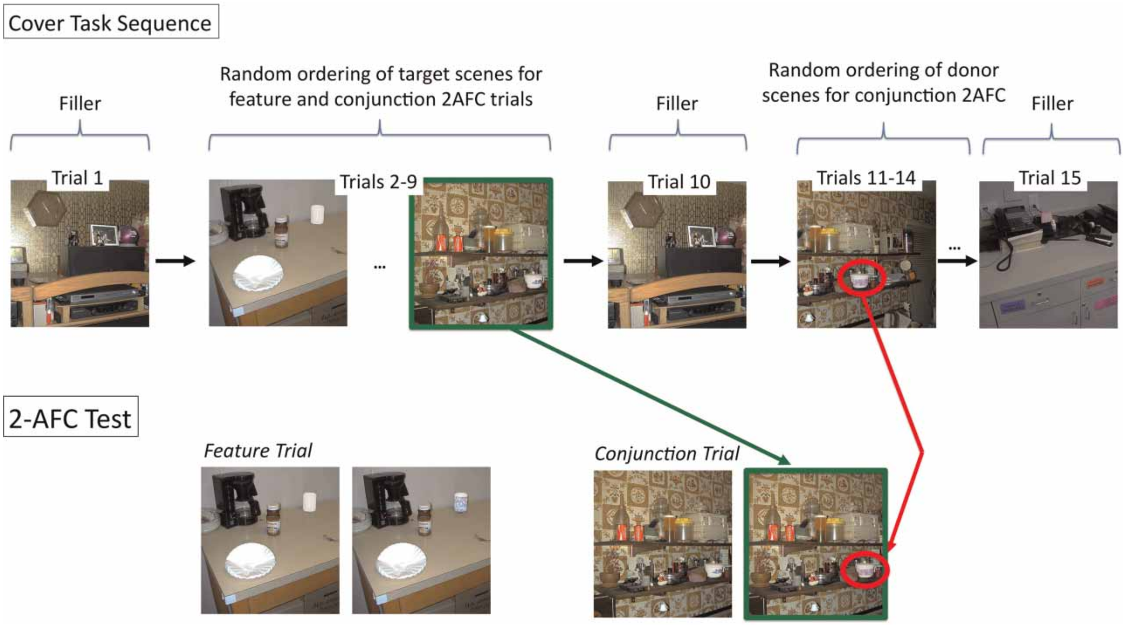

Figure 3. *The figure depicts the relationships among pictures in cover task trials to one another and to pictures in the recognition test for Experiment 1a, and the conjunction target before donor condition in Experiment 2. To view a colour version of this figure, please see the online issue of the Journal.*

version of the critical object, which would appear as a distractor in the subsequent memory test (see Figure 3). On Trial 15, a final filler picture (which was different than the picture from Trials 1 and 10) appeared to minimize VSTM's contribution to performance on the 2-AFC recognition test.

There were two kinds of 2-AFC recognition test trials (see Figure 3). On *conjunction memory trials*, the object appearance lure in the distractor scene was the same object as the one that appeared in the donor scene in the cover task phase, but was now presented within the same viewpoint as the correct option. Thus, the distractor scene was a novel conjunction of a previously seen viewpoint and critical object. Thus, both alternatives in the conjunction memory trial 2-AFC tests contained familiar objects. On *feature memory trials*, the distractor contained an object appearance lure that had not appeared during the cover task phase. Thus, the alternatives in the feature memory trial 2-AFC tests differed in terms of familiarity. There were four of each type of memory test trial. Order of presentation was randomized.

Eight versions of the experiment were created for Experiment 1a. Each version of each scene served in each role (i.e., target picture for feature memory trials, target picture for conjunction memory trials, donor scene, or distractor scene) an equal number of times. Within each version of the experiment, viewpoint (i.e., left vs. right) was counterbalanced for target pictures and donor pictures.

Results

Cover task

In all experiments, accuracy[1] on the cover task was near ceiling and did not vary significantly as a

---

[1] Cover task performance was also analysed in terms of the nonparametric sensitivity index A′ (Grier, 1971) and in terms of hits and false alarms. The pattern of results did not differ.

Table 1. *Mean accuracy on the cover task*

Experiment	*Accuracy*
Experiment 1a	.93 (.05)
Experiment 1b	.95 (.03)
Experiment 2	.94 (.02)
Experiment 3	.92 (.01)

Note: Standard deviations in parentheses.

function of picture type (see Table 1). Thus, we may assume that the luminance cue was effective at directing participants' attention to objects. Because we can assume that participants did attend to the critical object in each picture, any failures of recognition for critical objects cannot be attributed to failures of encoding (via attention), but instead must be due to problems in either storage or retrieval.

2-AFC recognition memory

Figure 4 presents the 2-AFC recognition accuracy data. Before the recognition test data were analysed, trials that contained critical objects to which the participant failed to respond during the cover task were removed (approximately 4% of recognition test trials) regardless of whether their choice on the 2-AFC was correct or incorrect.[2] A within-subject t test revealed that accuracy was higher on feature memory trials ($M = .64$, $SD = .27$) than on conjunction memory trials ($M = .46$, $SD = .31$), $t(22) = 2.25$, $p = .035$. Furthermore, accuracy was significantly better than chance (= .50) on feature memory trials, $t(22) = 2.51$, $p = .02$, but did not differ from chance on conjunction memory trials, $t(22) = 0.67$, $p = .51$. These results suggest that participants could remember object appearance information, but not in conjunction with picture-specific viewpoint information. Thus, these results are consistent with the hypothesis that object appearance and viewpoint are not integrated in VLTM.

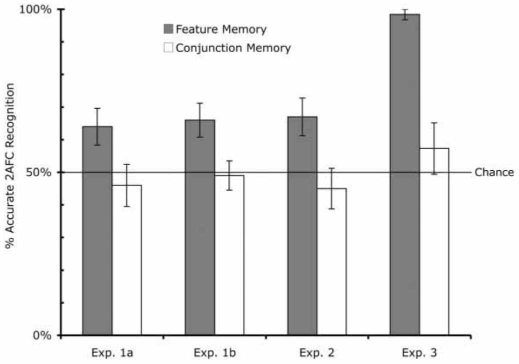

Figure 4. *Percentage recognition accuracy for Experiments 1a, 1b, 2, and 3. Error bars represent the standard error of the mean.*

EXPERIMENT 1B

In Experiment 1a, the pictures that would become correct options in the conjunction memory trials were always studied before their corresponding donor scenes (see Figure 3). Thus, it is possible that the donor scenes overwrote the representation of the target scenes for conjunction memory trials. If this were case, then both options on conjunction memory trials would have appeared incorrect, leaving participants to guess. In Experiment 1b, the pictures that would become correct options in the conjunction trials were always studied *after* their corresponding donor scenes (see Figure 5). If overwriting occurs in this case, then observers should be able to recognize the correct option in 2-AFC conjunction trials, as information about object appearance and viewpoint from the correct option would have overwritten the information from the donor scene.

Method

The method was exactly the same as that in Experiment 1a except as follows. Eight new versions were created for Experiment 1b in which the four target scenes for the feature recognition trials and the four donor scenes appeared in Cover Task

---

[2] The maximum number of trials removed for any given participant was two. The pattern of results does not change if these trials are included here and in subsequent experiments.

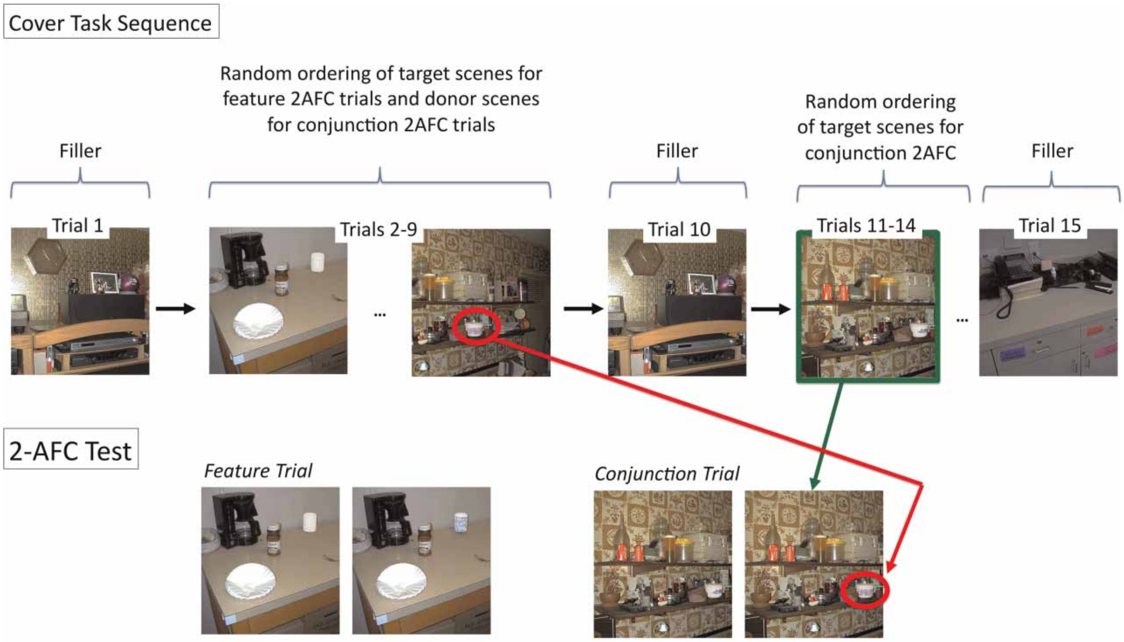

Figure 5. *The figure depicts the relationships among pictures in cover task trials to one another and to pictures in the recognition test for Experiment 1b, and the conjunction target after donor condition in Experiment 2. To view a colour version of this figure, please see the online issue of the Journal.*

Trials 2–9. The target scenes for the conjunction trials were presented in Trials 11–14 (see Figure 5).

Participants

Participants were 28 undergraduate students at Knox College who participated in exchange for class credit. A total of 3 participants were dropped from all analyses—2 had participated in Experiment 1a, and 1 did not complete the cover task.

Results

Cover task

See Table 1.

2-AFC recognition memory

Figure 4 presents the recognition accuracy data. As in Experiment 1a, recognition trials that contained critical objects to which the participant failed to respond during the cover task were removed (approximately 2% of recognition test trials). A within-subjects t test revealed that accuracy was higher on feature memory trials ($M = .66$, $SD = .26$) than on conjunction memory trials ($M = .49$, $SD = .24$), $t(24) = 2.35$, $p = .03$. Furthermore, accuracy was significantly better than chance on feature memory trials, $t(24) = 3.09$, $p = .005$, but did not differ from chance on conjunction memory trials, $t(24) = 0.14$, $p = .892$.

Discussion: Experiments 1a and 1b

In Experiments 1a and 1b, recognition accuracy was at chance when observers had to recognize the conjunction of object appearance and viewpoint. It did not matter if the correct option was initially studied before or after the donor picture, thus ruling out overwriting as a cause for this failure. In contrast, performance was above chance in feature memory trials, when the distractor contained object features that had not been viewed before. Thus, object appearance information was retained, but it could not be accurately recognized in conjunction with picture-specific viewpoint information.

These results suggest that episodic representations of the conjunction of object appearance and picture-specific viewpoint information are not automatically maintained in VLTM. However, in Experiments 1a and 1b, participants were not initially informed about the memory test. Thus, it is possible that participants *can* reliably maintain VLTM of a conjunction of object appearance and picture-specific viewpoint information if they make a conscious effort to do so.

EXPERIMENT 2

The main purpose of Experiment 2 was to test whether observers can accurately recognize the conjunction of object appearance and picture-specific viewpoint information if they are fully informed about the memory test ahead of time.

Method

The method was exactly the same as that in Experiments 1a and 1b except that participants were given a detailed set of instructions about the nature of the memory test, including information about visually similar distractors, the donor scenes, and the conjunction memory trials. They were even shown a diagram of a conjunction memory trial, including the donor scene (similar to Figures 3 and 5).

Participants

A total of 17 undergraduate students from Knox College participated for course credit. One participant was excluded from all analyses for failing to complete the cover task, leaving a final $N = 16$.

Design

The versions from Experiments 1a and 1b were combined in Experiment 2. Thus, for half the participants, correct options for conjunction memory trials were studied before the donor pictures (as in Experiment 1a), and for half the participants, correct option scenes for the conjunction memory trials were studied after the donor pictures (as in Experiment 1b).

Results

Cover task

See Table 1.

2-AFC recognition memory

Figure 4 presents the recognition accuracy data. As in Experiments 1a and 1b, recognition trials that contained critical objects to which the participant failed to respond during the cover task were removed (approximately 4% of recognition test trials). An analysis of variance (ANOVA) with recognition trial type (feature or conjunction memory trial) as a within-subjects factor and target/donor picture study order (conjunction target pictures before or after donor pictures) as a between-subjects factor yielded a main effect of recognition trial type, $F(1, 14) = 8.90$, $MSE = 0.043$, $p = .01$. The study-order main effect and the interaction effect were not significant, $F(1, 14) < 1$. As in Experiments 1a and 1b, accuracy was significantly above chance in the feature memory trials ($M = .67$, $SD = .23$), $t(15) = 3.30$, $p = .008$, but did not differ from chance on the conjunction memory trials ($M = .45$, $SD = .25$), $t(15) = 0.74$, $p = .47$.

Discussion

Experiment 2 replicated the results of Experiments 1a and 1b, demonstrating again that object appearance and picture-specific viewpoint information cannot be used in conjunction to recognize previously viewed pictures, even when participants are fully informed about the memory test. Not only was performance at chance in the conjunction memory trials, but performance in the feature memory trials (67% accurate) was virtually identical to performance in Experiments 1a and 1b (64% and 66%, respectively) when encoding was incidental. Thus, in line with findings from Castelhano and Henderson (2005), it appears that the intention to encode object appearance information does not always improve memory for it.

This lack of an effect is consistent with the proposal that visual–cognitive processes that are engaged during the course of attending to and identifying objects impose limits on how well object appearance is retained in VLTM, and that these limits are not always overcome by intention to remember. Note that this idea leaves open the possibility that intention has robust effects on the retention and use of visual information in some task settings. For example, observers detect visual changes more frequently when they are intentionally looking for them (e.g., Beck, Levin, & Angelone, 2007; Levin & Simons, 1997; Varakin & Levin, 2006). In the current experiments, however, observers might have been too busy completing the cover task to engage in the sort of processes that would enhance memory for object appearance.

In addition, the cover task may have prevented observers from attending to and encoding viewpoint-specific information. Given the possibility that our viewers were biased to encode object features at the expense of viewpoint, we conducted a third experiment to test whether our cover task actually prevented encoding of viewpoint.

EXPERIMENT 3

A simple explanation for the results of Experiments 1a, 1b, and 2 is that the difference between the complementary viewpoints was too small to be discriminated on the basis of information in VLTM. In this view, the chance-level performance on conjunction memory trials would be entirely due to an inadequate representation of picture-specific viewpoint information in VLTM, rather than a failure to integrate object appearance and picture-specific viewpoint. Indeed, the cover task used in Experiments 1a, 1b, and 2 may have caused observers to focus so much on object appearance that information about scene layout and viewpoint was simply not encoded.

This explanation seems unlikely because "picture-specific viewpoint" information should include all information that is available in one viewpoint that is not available in the complementary viewpoint. Thus, even if the layout and borders of a particular viewpoint were not represented precisely (cf., Intraub, Bender, & Mangels, 1992; Intraub & Bodamer, 1993), participants in our experiments could have (in principle) relied on a viewpoint-specific *object* inventory. Each viewpoint contained objects that were not visible in its complementary viewpoint (e.g., the coke cans in Figure 1a and the objects on the right end of the shelf in Figure 1b), and the set of noncritical objects to which participants responded in the cover task did not overlap across complementary viewpoints. Thus, if object appearance information was episodically integrated with picture-specific viewpoint information, then an observer who viewed both left and right viewpoints of the same larger scene (e.g., Figures 1a vs. 1b) would have two viewpoint-specific object inventories. The critical object in one of the viewpoints (e.g., the circled bowl in Figure 1b) would not have been integrated in a representation containing the objects that were unique to the complementary viewpoint (e.g., the coke cans in Figure 1a), and vice versa. Nevertheless, one purpose of Experiment 3 was to test whether picture-specific viewpoint information would be encoded even though our cover task encourages a focus on individual objects. To this end, distractors on the 2-AFC differed from the correct option in terms of viewpoint (and both options depicted the same version of the critical object).

A second purpose for Experiment 3 was to test whether people could recognize which viewpoint appeared with a given object (see Figure 6). The conjunction memory trials of Experiments 1a, 1b, and 2 held viewpoint constant and tested whether people could recognize which object appeared within a given viewpoint (as shown in Figures 3 and 5). In Experiment 3, conjunction memory trials held the critical object constant and tested whether people could recognize which viewpoint appeared with a given object. If object appearance and viewpoint are not integrated in VLTM, then the results of Experiment 3's conjunction memory trials should replicate the results showing poor conjunction memory from Experiments 1a, 1b, and 2.

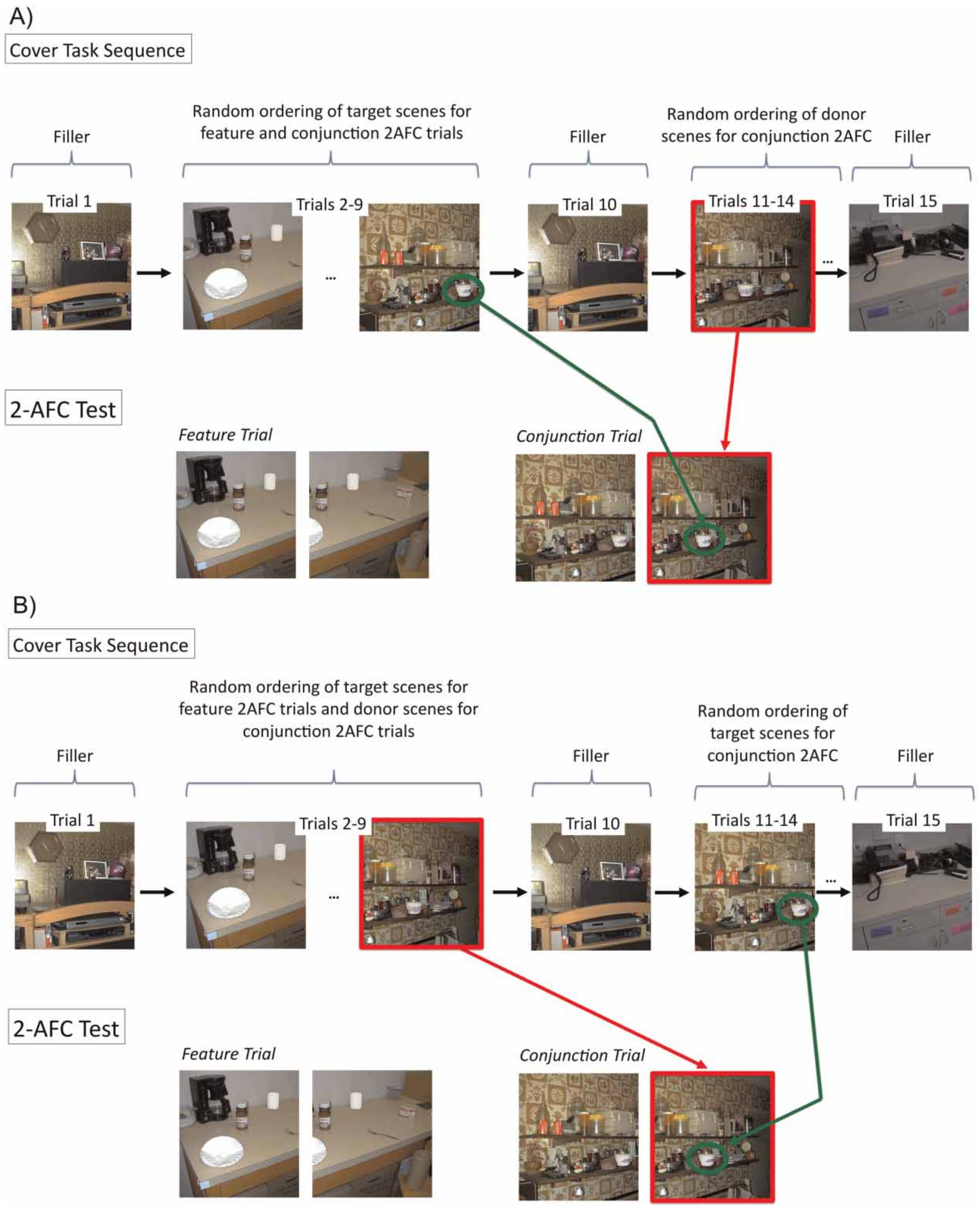

Figure 6. *The figure depicts the relationships among pictures in cover task trials to one another and to pictures in the recognition test for Experiment 3. In (A), the conjunction target before donor condition is depicted, and in (B), the conjunction target after donor condition is depicted.*

Method

The design of Experiment 3 was identical to that of Experiment 2 except that distractor scenes in the 2-AFC differed from the previously viewed pictures in terms of viewpoint (see Figures 6a and 6b). Thus, the critical object in both options was identical, and distractors on conjunction memory trials had the same viewpoint as donor pictures. Note that participants were even led to believe that the 2-AFC test would assess memory for the visual appearance of individual objects (as in Experiment 2). Thus, both the instructions about the memory test and the instructions about the cover task encouraged object-based processing, which, if anything, should work against the hypothesis that picture-specific viewpoint information will be encoded. Participants were, of course, given accurate instructions about the memory test, but only after the study phase was completed.

Participants

A total of 18 members of the Knox College community participated in exchange for course credit. Two participants' results were omitted from all analyses for failing to complete the cover task.

Results

Cover task

See Table 1.

2-AFC recognition memory

Figure 4 presents the recognition accuracy data. As in previous experiments, recognition trials that contained critical objects to which the participant failed to respond during the cover task were removed (approximately 3% of recognition test trials). An ANOVA with recognition trial type (feature or conjunction memory trial) as a within-subjects factor and target/donor picture study order (conjunction target pictures before or after donor pictures) as a between-subjects factor yielded a main effect of recognition trial type, $F(1,\ 14) = 24.95$, $MSE = 0.054$, $p = .001$. The study-order main effect and the interaction effect were not significant, Fs $(1,\ 14) < 1.15$, ps $> .30$. As in Experiments 1a, 1b, and 2, accuracy was significantly above chance in the feature memory trials ($M = .98$, $SD = .06$), $t(15) = 31.00$, $p < .001$, but not in the conjunction memory trials ($M = .57$, $SD = .32$), $t(15) = 0.92$, $p = .37$.

Discussion

The results of Experiment 3 strengthen the conclusions from Experiments 1a, 1b, and 2 in two critical ways.

First, Experiment 3's results suggest that the cover task used in Experiments 1a, 1b, and 2 did not prevent observers from encoding picture-specific viewpoint information in VLTM, or that the viewpoint differences were too small to discriminate accurately—observers performed almost perfectly on the recognition test when the same picture-specific viewpoint features present in Experiments 1a, 1b, and 2 were necessary to differentiate between the target and the distractor. It is important to note that in the feature recognition trials of Experiment 3, the novel viewpoint image contained novel objects. Therefore, as suggested in our introduction, it is possible that the good performance was based on recognizing novel objects and not recognizing viewpoint per se. In any case, the results of Experiment 3's feature recognition trials demonstrate that participants had access to information that enabled them to discriminate old and new viewpoints, be it a representation of viewpoint, or a viewpoint-specific object inventory.

Second, Experiment 3's results once again demonstrate that object appearance information and picture-specific viewpoint information cannot be used in conjunction to recognize previously viewed pictures, though, in this case, the conjunction task varied viewpoint and held the critical object constant.

GENERAL DISCUSSION

The current experiments tested whether object appearance and picture-specific viewpoint information are episodically integrated in VLTM.

If such representations are maintained, then participants should be able to perform at better than chance levels on 2-AFC tests that require recognizing the conjunction of both object appearance and picture-specific viewpoint information. In contrast to this prediction, the results showed that under both incidental (Experiments 1a and 1b) and intentional (Experiments 2 and 3) encoding conditions, participants performed no better than chance on 2-AFC tests for the conjunction of object appearance information and picture-specific viewpoint information (Experiments 1a, 1b, 2, and 3). Nevertheless, participants retained enough information in VLTM to recognize previously viewed pictures that differed from distractors only in terms of object appearance (Experiments 1a, 1b, and 2) or only in terms of viewpoint (Experiment 3). In summary, observers could recognize previously viewed pictures based on either object appearance or picture-specific viewpoint information, but could not discriminate new from old pictures based on the conjunction of object appearance and picture-specific viewpoint information.

There are at least three general ways to explain the finding that participants could not recognize the episodic conjunction of object appearance and picture-specific viewpoint. First, it is possible that participants could only retain the episodic conjunction of object appearance and picture-specific viewpoint for one of the two pictures that they studied (i.e., the picture showing the correct conjunction of object appearance and viewpoint or the donor picture). That is, the first picture may proactively overwrite, or be retroactively overwritten by the second picture. We call these "overwriting accounts". Second, it is possible that studying two similar pictures (i.e., the target and donor pictures) caused the episodic representation of each individual picture to become degraded by interference to the point that object appearance and/or picture-specific viewpoint information was not retained with sufficient fidelity to be useful for supporting performance in a 2-AFC test, which we call the "interference account". Third, it is possible that object appearance information and picture-specific viewpoint information of both studied pictures was retained in VLTM, but the episodic conjunction of object appearance and picture-specific viewpoint was not, which we call the "failed conjunction account". These three accounts are evaluated below.

The overwriting accounts predict that the order in which the target and donor scenes were initially viewed should affect performance on the 2-AFC recognition test on conjunction memory trials. However, this factor did not affect performance in any of the experiments. Thus, the overwriting accounts can be ruled out.

According to the interference account, attending to two similar pictures (i.e., two different viewpoints of the same larger scene) would have degraded the representation of information obtained from each individual picture through mutual interference. In other words, proactive and retroactive interference would effectively prevent participants from retaining episodically specific information about object appearance or viewpoint in VLTM in an accessible format. This sort of mechanism cannot be ruled out because previous work has demonstrated that long-term recognition performance declines as the number of studied items from a given similarity-class increases (e.g., Koutstaal & Schacter, 1997; Vogt & Magnussen, 2007). However, it seems unlikely that such interference would be strong enough to degrade the representation of object appearance and/or picture-specific viewpoint to such an extent that neither would be useful for supporting performance on a recognition test. For example, a recent study showed that accuracy on a 2-AFC memory test for 400 categorically similar pictures was still an impressive 83.7% after a 30-minute retention interval (Vogt & Magnussen, 2007). Thus, while interference could well have contributed to the current results, it seems unlikely that interference between the target and donor pictures could account for the chance-level performance in the current experiments' conjunction memory trials. In other words, it seems likely that participants retained some information about both versions of the critical object and both viewpoints (i.e., from the target and donor scenes).

According to the failed conjunction account, object appearance and picture-specific viewpoint information are both retained, but are not episodically integrated in VLTM. Thus, when confronted with a 2-AFC trial containing old and new conjunctions of object appearance and viewpoint, participants would be forced to reconstruct the conjunctions, which could lead to the formation of illusory conjunctions (c.f., Treisman, 1996). Illusory conjunctions of visual features in VSTM have been reported many times in previous research—for example, in transsaccadic memory experiments (Irwin & Gordon, 1998) and experiments using rapid serial visual presentation (Intraub, 1985, 1989). In these cases, illusory conjunctions in VSTM were attributed to limitations of spatial and temporal attention. However, illusory conjunctions of object appearance and viewpoint in VLTM in the current experiments cannot be attributed to limits of attentional mechanisms because cover task performance suggests that participants attended to the critical object, and each picture's presentation duration was long enough not to stress mechanisms that integrate visual information over brief periods of time. Thus, in the current experiments, the illusory conjunctions in VLTM would have stemmed from a failure to maintain and/or access properly conjoined representations in VLTM. The idea that object appearance and picture-specific viewpoint might consist of separate pieces of information needing to be conjoined is consistent with the idea that objects and scenes are processed in different brain areas and that coding of scenes in the PPA is viewpoint-specific (Epstein, 2005; Epstein & Kanwisher, 1998; O'Craven & Kanwisher, 2000). Furthermore, this account bears some similarity to the source-monitoring framework (Johnson, Hashtroudi, & Lindsay, 1993), which states that information in long-term memory is not typically retrieved with a source-identifying tag. Rather, source attributions are based on judgements about the type and quality of information a given memory trace contains. Our conjunction memory trials can therefore be conceived of as a kind of source attribution problem because the participants had to determine the source (e.g., picture-specific viewpoint) of some information (e.g., object appearance information). In the current experiments, the similarity between the sources of the target and distractor objects was high, thus increasing the likelihood of source confusions.

Interestingly, a strategy of identifying scenes on the basis of object features that are not bound to particular image locations is sometimes used for scene classification by "bag of words" models in computer vision[3] (see, e.g., Cao & Fei-Fei, 2007, for a discussion). The basic idea is that scenes can often be classified simply by registering different image features (e.g., image patches and object parts), while ignoring how they are spatially conjoined. The current results suggest that something similar may happen in recognition memory tests. That is, observers might base their decisions on whether or not image features (e.g., object appearance and viewpoint in this case) are present in VLTM, while ignoring how the features were originally conjoined. Such a strategy would work well in situations where old and new items could be discriminated on the basis of the mere presence of individual features (such as the current experiments' feature memory trials), but would lead to errors in cases where accessing the conjunction of features is necessary (such as the current experiments' conjunction memory trials).

Our claim has several important caveats. Most importantly, we are not claiming that object appearance is or is not associated with picture-specific viewpoint information in VLTM, only that it is not integrated. In fact, the logic underlying our interpretation of the current results does not speak to issues about association. Specifically, our interpretations of our experiments are based on the following modus tollens ($p \rightarrow q$, $\sim q$, $\sim p$) logic: If two pieces of information, here object appearance and picture-specific viewpoint, are integrated in episodic long-term memory (p), then viewers should be able to discriminate an

---

[3] We would like to thank an anonymous reviewer for mentioning the similarity between our claim and these models.

old from a new conjunction of those two pieces of information in a long-term memory recognition memory test (q). Our results show that viewers cannot make such discriminations (∼q). Thus, our results suggest that those two pieces of information, object appearance and picture-specific viewpoint, are not integrated in long-term memory (∼p). However, we cannot use this same logic to make any claims about association, because we have argued that association and integration are different. Specifically, whereas integrated representations are "fused" (Hayes et al., 2007), associated representations are fragmentable (Murnane et al., 1999). Thus, accessing one part of an association would not guarantee access to the other. Thus, we cannot use the results of the current study to argue that object appearance and viewpoint are, or are not, associated.

One might be tempted to argue that our results suggest that object details and viewpoint-specific information are not associated, because if two items are associated, and one is remembered, it seems reasonable that a person should be able to use the association to access the other. However, as our above distinction between integration and association suggests, it is possible that object appearance and viewpoint-specific information were associated in VLTM, but participants simply could not profitably use those associations in conjunction memory trials. Indeed, the results of Hayes et al. (2007) and Hollingworth (2006) suggest that object details and scene contexts are associated in long-term memory, which lends credence to the idea that some kind of viewpoint-specific information may be associated with object appearance. But if so, why was the association not used on conjunction trials? One possibility is that distractor conjunctions shared too many associations with correct conjunctions. For example, recall that the local context of critical objects in donor pictures was exactly the same as the local context of critical objects in conjunction targets. Thus, in conjunction memory trials, the shared associations may have interfered with participants' ability to utilize unique (i.e., nonshared) viewpoint-specific associations. In Hayes et al.'s (2007) and Hollingworth's (2006) experiments, distractor objects on the recognition test were novel and therefore had no prior association with any aspect of the context in which they were presented. Thus, in those studies, all associations between object and scene context could have served as retrieval cues, whereas, in the current experiments, only a subset of associations would have been helpful. Nevertheless, we note that this account of how associations are or are not used is highly speculative and should be addressed by future work.

A second important caveat is that we are not claiming that it is impossible to integrate object details with other sorts of information. Clearly, "context" encompasses more than picture-specific viewpoint information; thus, it remains possible that object appearance is (or can be) integrated with other kinds of information (e.g., semantic information, auditory information, task-related information, etc.) that were not tapped by the current methods. Moreover, it is possible that the processes that integrate object information with contextual information do not do so in a manner that maintains the episodic conjunction of features that are present in any particular picture. Instead, information obtained from a particular picture (which may or may not include object appearance and viewpoint) might be integrated within a more abstract (i.e., nonepisodic) representation of context, to form something akin to a schema (e.g., Brewer & Treyens, 1981) or context frame (e.g., Bar, 2004). A system that integrated information over a large number of episodes (as opposed to a single episode) may be able to take advantage of statistical regularities in the natural visual world, which could facilitate perception of novel scenes (e.g., Bar, 2004; Oliva & Torralba, 2007). For example, scene schema (or context frames) can be used to facilitate the identification of objects (e.g., Davenport, 2007; Davenport & Potter, 2004), guide attention and eye-movements (e.g., Brockmole & Henderson, 2008; Loftus & Mackworth, 1978), and influence the encoding of information into VLTM (e.g., Brewer & Treyens, 1981; Pezdek, Whetstone, Reynolds, Askari, & Dougherty, 1989), and schematic layout representations can facilitate processing of novel viewpoints of a particular environment (e.g., Friedman & Waller, 2008).

In conclusion, the current experiments have demonstrated that observers retain both object appearance and picture-specific viewpoint in VLTM. However, this information cannot be used in conjunction. Based on these results, we have argued that the memories underlying performance on picture recognition tests do not function as episodic instantiations of their source pictures. Physical pictures contain object appearance information and picture-specific viewpoint information. Thus, an episodically integrated representation of a picture in VLTM would also have both types of information (even if it were assumed to be less detailed than a sensory copy in iconic memory). The current results therefore suggest that mental scene representations in VLTM are episodically fragmented (or fragmentable) structures in comparison to the physical pictures on which they are based. Why, then, are these results important? As noted earlier, no one argues that mental scene representations are as detailed as actual pictures. However, so-called sparse representation theories have been largely rejected on the grounds that object details can be stored in VLTM. Thus, the pendulum of theory has been swinging away from the view that scene representations bear little resemblance to their source pictures and towards a view that may be consistent with the idea that episodic scene representations in VLTM are fully integrated instantiations of pictures as they were viewed, albeit with loss of sensory detail. Thus, the current results are important because they suggest that although various kinds of visually specific information (e.g., object appearance and picture-specific viewpoint information, etc.) may be retained in VLTM, such information may not be integrated within an episodic representation of the picture as it was viewed. This would explain why the whole of a particular scene memory is difficult to assemble from the sum of its detailed parts.

Original manuscript received 8 December 2008
Accepted revision received 29 July 2009
First published online 15 October 2009

REFERENCES

Bar, M. (2004). Visual objects in context. *Nature Reviews Neuroscience*, *5*, 617–629.

Beck, M. R., Levin, D. T., & Angelone, B. (2007). Change blindness: Beliefs about the roles of intention and scene complexity in change detection. *Consciousness and Cognition*, *16*(1), 31–51.

Brady, T. F., Konkle, T., Alvarez, G. A., & Oliva, A. (2008). Visual long-term memory has a massive storage capacity for object details. *Proceedings of the National Academy of Sciences*, *105*(38), 14325–14329.

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

Brockmole, J. R., & Henderson, J. M. (2008). Prioritizing new objects for eye fixation in real-world scenes: Effects of object-scene consistency. *Visual Cognition*, *16*, 375–390.

Burns, B., & Shepp, B. E. (1988). Dimensional interactions and the structure of psychological space: The representation of hue, saturation, and brightness. *Perception and Psychophysics*, *43*(5), 494–507.

Cao, L., & Fei-Fei, L. (2007). Spatially coherent latent topic model for concurrent object segmentation and classification. *Proceedings of the IEEE International Conference in Computer Vision*, *11*, 1–8.

Castelhano, M. S., & Henderson, J. M. (2005). Incidental visual memory for objects in scenes. *Visual Cognition*, *12*(6), 1017–1040.

Davenport, J. L. (2007). Consistency effects between objects in scenes. *Memory and Cognition*, *35*(3), 393–401.

Davenport, J. L., & Potter, M. (2004). Scene consistency in object and background perception. *Psychological Science*, *15*(8), 559–564.

Epstein, R. A. (2005). The cortical basis of visual scene processing. *Visual Cognition*, *12*(6), 954–978.

Epstein, R., & Kanwisher, N. (1998). A cortical representation of the local visual environment. *Nature*, *392*, 598–601.

Friedman, A., & Waller, D. (2008). View combination in scene recognition. *Memory and Cognition*, *36*(3), 467–478.

Grier, J. B. (1971). Nonparametric indexes for sensitivity and bias: Computing formulas. *Psychological Bulletin*, *75*(6), 424–429.

Hanna, A., & Remington, R. (1996). The representation of color and form in long-term memory. *Memory and Cognition*, *24*(3), 322–330.

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*(9), 873–889.

Henderson, J. M., & Hollingworth, A. (2003). Eye-movements, visual memory, and scene representation. In M. A. Peterson & G. Rhodes (Eds.), *Perception of faces, objects, and scenes* (pp. 356–383). Oxford, UK: Oxford University Press.

Hock, H. S., & Schmelzkopf, K. F. (1980). The abstraction of schematic representations from photographs of real-world scenes. *Memory and Cognition*, *8*(6), 543–554.

Hollingworth, A. (2004). Constructing visual representations of natural scenes: The roles of short- and long-term visual memory. *Journal of Experimental Psychology: Human Perception and Performance*, *30*(4), 519–537.

Hollingworth, A. (2006). Scene and position specificity in visual memory for objects. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *32*(1), 58–69.

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*(1), 113–136.

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 and Review*, *8*(4), 761–768.

Intraub, H. (1985). Visual dissociation: An illusory conjunction of pictures and forms. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *11*(4), 431–442.

Intraub, H. (1989). Illusory conjunction of forms, objects, and scenes during rapid serial visual search. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *15*(1), 98–109.

Intraub, H. (1997). The representation of visual scenes. *Trends in Cognitive Sciences*, *1*(6), 217–222.

Intraub, H., Bender, R. S., & Mangels, J. A. (1992). Looking at pictures but remembering scenes. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *18*(1), 180–191.

Intraub, H., & Bodamer, J. L. (1993). Boundary extension: Fundamental aspect of pictorial representation or encoding artifact? *Journal of Experimental Psychology: Learning, Memory and Cognition*, *19*(6), 1387–1397.

Irwin, D. E., & Gordon, R. D. (1998). Eye movements, attention and trans-saccadic memory. *Visual Cognition*, *5*(1–2), 127–155.

Johnson, M. K., Hashtroudi, S., & Lindsay, D. S. (1993). Source monitoring. *Psychological Bulletin*, *114*(1), 3–28.

Koutstaal, W., & Schacter, D. L. (1997). Gist-based false recognition of pictures in older and younger adults. *Journal of Memory and Language*, *37*, 555–583.

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

Loftus, G. R., & Mackworth, N. H. (1978). Cognitive determinants of fixation location during picture viewing. *Journal of Experimental Psychology: Human Perception and Performance*, *4*(4), 565–572.

Murnane, K., Phelps, M. P., & Malmberg, K. (1999). Context-dependent recognition memory: The ice theory. *Journal of Experimental Psychology: General*, *128*(4), 403–415.

O'Craven, K. M., & Kanwisher, N. (2000). Mental imagery of faces and places activates corresponding stimulus-specific brain regions. *Journal of Cognitive Neuroscience*, *12*(6), 1013–1023.

Oliva, A., & Torralba, A. (2007). The role of context in object recognition. *Trends in Cognitive Sciences*, *11*(12), 520–527.

O'Regan, J. K. (1992). Solving the "real" mysteries of visual perception: The world as an outside memory. *Canadian Journal of Psychology*, *46*(3), 461–488.

Peterson, M. A., & Rhodes, G. (2003). Analytic and holistic processing: The view through different lenses. In M. A. Peterson & G. Rhodes (Eds.), *Perception of faces, objects, and scenes* (pp. 3–19). Oxford, UK: Oxford University Press.

Pezdek, K., Whetstone, T., Reynolds, K., Askari, N., & Dougherty, T. (1989). Memory for real-world scenes: The role of consistent with schema expectation. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *15*(4), 587–595.

Rensink, R. A. (2000). The dynamic representation of scenes. *Visual Cognition*, *7*(1–3), 17–42.

Rensink, R. A. (2002). Change detection. *Annual Review of Psychology*, *53*(1), 245–277.

Sanocki, T. (2003). Representation and perception of scenic layout. *Cognitive Psychology*, *47*(1), 43–86.

Sanocki, T., Michelet, K., Sellers, E., & Reynolds, J. (2006). Representation of scene layout can consist of independent, functional pieces. *Perception and Psychophysics*, *68*(3), 415–427.

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

Simons, D. J. (2000). Current approaches to change blindness. *Visual Cognition*, *7*(1–3), 1–15.

Simons, D. J., & Levin, D. T. (1997). Change blindness. *Trends in Cognitive Sciences*, *1*(7), 261–267.

Simons, D. J., & Rensink, R. A. (2005). Change blindness: Past, present and future. *Trends in Cognitive Sciences*, *9*(1), 16–20.

Standing, L. (1973). Learning 10,000 pictures. *Quarterly Journal of Experimental Psychology*, *25*(2), 207–222.

Standing, L., Conezio, J., & Haber, R. N. (1970). Perception and memory for pictures: Single-trial learning of 2500 visual stimuli. *Psychonomic Science*, *19*(2), 73–74.

Treisman, A. (1996). The binding problem. *Current Opinion in Neurobiology*, *6*, 171–178.

Varakin, D. A., & Levin, D. T. (2006). Change blindness and visual memory: Visual representations get rich and act poor. *British Journal of Psychology*, *97*(1), 51–77.

Varakin, D. A., & Levin, D. T. (2008). Scene structure enhances change detection. *Quarterly Journal of Experimental Psychology*, *61*(4), 543–551.

Vogt, S., & Magnussen, S. (2007). Long-term memory for 400 pictures on a common theme. *Experimental Psychology*, *54*(4), 298–303.