Implicit scene learning is viewpoint dependent

KAO-PING CHUA and MARVIN M. CHUN
Vanderbilt University, Nashville, Tennessee

When novel scenes are encoded, the representations of scene layout are generally viewpoint specific. Past studies of scene recognition have typically required subjects to explicitly study and encode novel scenes, but in everyday visual experience, it is possible that much scene learning occurs incidentally. Here, we examine whether implicitly encoded scene layouts are also viewpoint dependent. We used the contextual cuing paradigm, in which search for a target is facilitated by implicitly learned associations between target locations and novel spatial contexts (Chun & Jiang, 1998). This task was extended to naturalistic search arrays with apparent depth. To test viewpoint dependence, the viewpoint of the scenes was varied from training to testing. Contextual cuing and, hence, scene context learning decreased as the angular rotation from training viewpoint increased. This finding suggests that implicitly acquired representations of scene layout are viewpoint dependent.

At any given moment, the visual system is flooded with a tremendous amount of complex stimuli. Because the brain has limited capacity, not all of this information can be used to guide thoughts and actions (Chun & Wolfe, 2001; Pashler, 1998; Simons & Levin, 1997; Sperling, 1960). Indeed, the visual system must focus primarily on information that has behavioral significance and ignore information that does not. To do this, the visual system employs selective attention mechanisms.

One of the main goals of attention research is to determine what factors draw attention. A classic method of addressing this issue is visual search, in which subjects search for a target object in an array of distractors (Treisman & Gelade, 1980; Wolfe, 1994a). By varying the properties of the targets and distractors, researchers have identified many properties that guide visual attention mechanisms and facilitate search. These properties include image-based, or *bottom-up*, factors, such as salient features (Bravo & Nakayama, 1992; Treisman & Gelade, 1980), abrupt onsets (Yantis & Jonides, 1984), and the presence of features that were formerly absent (Treisman & Gormican, 1988). Other guiding features include knowledge-based, or *top-down*, factors, such as automaticity effects (Schneider & Shiffrin, 1977), novelty effects (Johnston, Hawley, Plew, Elliott, & DeWitt, 1990), familiarity effects (Wang, Cavanagh, & Green, 1994), expectancy effects for locations where the target often appears (Miller, 1988; Shaw, 1978; Shaw & Shaw, 1977), and search templates that focus attention toward items that share features with the target (Egeth, Virzi, & Garbart, 1984).

Recently, Chun and colleagues showed that in addition to the factors listed above, global context plays an important role in guiding attention (Chun, 2000; Chun & Jiang, 1998, 1999; Chun & Nakayama, 2000; Olson & Chun, 2001). This guidance of attention by context, or *contextual cuing*, was demonstrated through a simple variation of the classic visual search task. In a prototypical experiment, subjects searched for a T among rotated L distractors. The spatial layout (configuration) of the items on the display defined the visual context of a target. A set of configurations was generated and repeated across blocks throughout the experiment, and targets appeared in consistent locations within their respective configurations from repetition to repetition. If subjects are sensitive to visual context, Chun and colleagues hypothesized that search performance should improve for targets appearing in repeated contexts (*old* condition), relative to targets appearing in novel contexts (*new* condition).

Over time, subjects did become increasingly faster at detecting the target in old displays than in new displays. This contextual cuing effect indicates that the subjects learned the global visual context and associated it with the target location, thus facilitating search (Chun & Jiang, 1998). More generally, the contextual cuing effect may reflect the visual system's sensitivity to regularities in scenes, which allows for the efficient deployment of attention toward relevant aspects of a scene (Chun, 2000; Chun & Jiang, 1998, 1999; Chun & Nakayama, 2000; Olson & Chun, 2001).

Three-Dimensional Contextual Cuing

One of the chief advantages of the contextual cuing paradigm is its potential ecological validity. Indeed, invariant contextual information is prevalent in the environment and may continually constrain attention mechanisms to focus on important objects in the visual world (Biederman, Mezzanotte, & Rabinowitz, 1982; Chun, 2000; Palmer, 1975; Rensink, O'Regan, & Clark, 1997). However, one problem

that potentially limits the generality of contextual cuing is that Chun and Jiang's (1998) stimuli lacked three-dimensionality. Indeed, the flat displays of visual search arrays clearly do not represent the rich depth and perspective of the real world.

Thus, to enhance the ecological validity of contextual cuing, it would be useful to establish that contextual cuing generalizes to three dimensions. To address this issue, we replicated the basic design of Chun and Jiang (1998), replacing the flat displays with artificial scenes that used pictorial cues to give an impression of apparent depth. On the basis of the work of Aks and Enns (1996), who demonstrated that search performance was sensitive to pictorial depth cues, such as background texture gradients, as well as the work of Wolfe (1994b), who showed that visual search tasks using isolated stimuli could be extended to naturalistic images, it is likely that subjects would exhibit a robust contextual cuing effect with our pseudonaturalistic scenes.

Contextual Cuing and Instance Theory

Another important issue concerns the nature of contextual representations that drive the cuing effect. One hypothesis is that memory traces for viewed contexts are stored as separate instances (Chun & Jiang, 1998). According to Logan's (1988) instance theory, automaticity (improved performance) results from the retrieval of episodic memory traces that are formed during each exposure to a particular stimulus. During the first few exposures, performance is mediated by slow, generic algorithms of attentional search. With repeated exposures, however, performance begins to rely on direct retrieval of the accumulating memory traces of previously viewed scenes. These memory traces cue attention to the associated target location and, thus, circumvent the slowness of memoryless attentional search. In contrast, for new displays, subjects continue to rely on algorithmic processing, since memory traces for novel displays do not exist. Thus, with training, search performance becomes faster on old displays than on new displays.

If search relies on memory traces of viewed displays (contexts), a fundamental question is how specific these memory traces are. If the memory traces are abstract, contextual cuing should readily generalize to transformed versions of learned displays. If the memory traces are specific, contextual cuing should be strongest in learned displays (Palmeri, 1997). Our use of three-dimensional (3-D) displays allowed us to address this question in a manner that should be informative to debates in the object/scene recognition literature.

Viewpoint Dependence/Independence and Contextual Cuing

Visual recognition of objects and scenes depends on a process by which stimuli are matched to an internal mental representation or memory of the stimuli. Presumably, these memories are formed from a certain vantage point. Yet, in our everyday lives, we are often able to recognize familiar stimuli from novel points of view. For example, we are able to recognize our bedroom from a number of vantage points, even though we have not necessarily experienced each of those vantage points previously.

The mechanism of recognizing novel views remains an important topic of debate (Biederman & Gerhardstein, 1993; Bulthoff & Edelman, 1992; Marr & Nishihara, 1978; Tarr, 1995). The key issue in this debate is the nature of object representations. According to viewpoint-dependent theories, mental representations change depending on the viewpoint from which they are made. By this account, the visual system encodes a large number of "snapshots," one for each viewpoint; each of these snapshots is constructed every time an object or scene is experienced from a particular vantage point. In contrast, viewpoint-independent theories argue that each stimulus has a general structural description that can be aligned to identify it from most viewpoints. The difference between the two points of view can be summarized as follows: According to viewpoint-dependent models, recognizing an object or a scene from a novel viewpoint will be more difficult than recognizing it from an experienced viewpoint (Tarr, 1995). In contrast, viewpoint-independent models predict that the stimulus will be equally easy to recognize from different viewpoints (Biederman & Gerhardstein, 1993).

Although the viewpoint dependency has been extensively researched for object recognition (Tarr, 1995), there have been fewer studies in which this issue has been investigated for scene recognition. One study to do so was conducted by Diwadkar and McNamara (1997). In their experiments, subjects learned a spatial layout of six different objects from a single vantage point. In a subsequent recognition task, the subjects experienced photographs of different layouts, some of which were rotated versions of the training layout. When the subjects were asked to determine whether the presented layout was the same as the training layout, response latency for rotated training layouts increased as the distance from the training view increased. This clearly supports the viewpoint-dependent hypothesis, since viewpoint independence predicts equal ease for all viewpoints. In addition, the results complement other studies that have suggested that spatial memories are viewpoint specific (Levine, Jankovic, & Palic, 1982; Shelton & McNamara, 1997).

Our use of pseudo–3-D scenes allows us to approach this debate by using a task quite different from the tasks generally used in this area of research. In the contextual cuing paradigm, subjects search for an embedded target and learn global scene (context) information incidentally. In other words, subjects are never instructed to encode the scene contexts. In contrast, past studies of scene recognition have typically required subjects to overtly attend to the entire training display. Hence, we will explore whether scene representations are viewpoint dependent or independent when scenes are encoded incidentally while performing a different primary task—that is, visual search.

In our experiment, subjects searched through pseudo–3-D scenes that were presented from a constant viewpoint

during training. These trials always consisted of scenes viewed from 0°, 15°, 30°, or 45° (see Figure 1). After training, the subjects completed a set of testing blocks in which displays were always presented from 0°. If incidentally acquired contextual representations are viewpoint dependent, the magnitude of the contextual cuing effect should decrease with increasing rotation. Such a result would be consistent with Lassaline and Logan (1993), who showed that transfer of learning in a counting task is disrupted across different orientations in the picture plane. In contrast, if contextual representations are viewpoint independent, there should be little effect of scene rotation.

Three-Dimensional Contextual Cuing and Implicit Learning

A final issue regards the nature of contextual learning. For two-dimensional (2-D) layouts, Chun and Jiang (1998) argued that global visual context can be learned implicitly. Indeed, a forced-choice recognition test administered after the search task revealed that subjects could not explicitly discriminate between old and new configurations, despite the fact that some form of specific memory for old configurations clearly improved their search performance (Chun & Jiang, 1998). Implicit learning refers to learning without explicit awareness that learning took place (Jacoby & Witherspoon, 1982; Schacter, 1987; Squire, 1992). The advantage of implicit learning is that it allows more information to be acquired than is possible through explicit channels (Berry & Dienes, 1993; Reber, 1989; Stadler & Frensch, 1998). Our aim was to establish that scene learning can occur implicitly, even for pseudonaturalistic displays that contain rich detail and a variety of perspective cues that may aid explicit recall.

To assess implicit learning in the contextual cuing task, we administered a sensitive explicit memory task introduced by Chun and Jiang (in press). In this task, subjects experienced a series of scenes in which the target was replaced by a distractor; these scenes included the old configurations used in the search task. The subjects were instructed to guess which quadrant of the display was most likely to contain the target if it were present. This memory test is better than the recognition test used in earlier studies of contextual cuing, because it taps directly into the information that supports faster search performance—namely, the approximate location of a target, given its repeated, associated context (Shanks & St. John, 1994). If subjects develop conscious knowledge of where to look in old displays, they should perform above chance for old displays in this guessing task. If contextual cuing is implicit, guessing performance should be at chance.

In summary, this study addressed the following three questions. First, does contextual cuing generalize to scene-like search displays with apparent depth? Second, are implicitly acquired scene context representations viewpoint dependent (specific) or viewpoint independent (abstract)? Finally, can contextual cuing for more naturalistic displays occur implicitly?

EXPERIMENT

Subjects performed 35 blocks of search for a bowling-pin–shaped target among an array of distractors. The distractors were shaped like dumbbells; half of them had small spheres at the end of the dumbbell, whereas the other half had large spheres (see Figure 1). Each block contained 16 trials. Eight of these trials were old, repeated displays that appeared once each per block. We predicted a facilitation for search on old displays relative to new displays, based on implicit learning of repeated contexts and associated target locations.

The first 30 blocks made up the *training* phase, and the last 5 blocks made up the *testing* phase. To examine whether scene context learning is viewpoint dependent, the training phase, varied between subjects, presented scenes viewed from 0°, 15°, 30°, or 45°; in contrast, the testing phase always presented scenes from 0°. If implicit scene learning is specific and viewpoint dependent, the contextual cuing effect would decrease in the testing phase as the rotation difference between training and testing increases. However, if implicit scene learning produces abstract layout representations, there should be no effect of training rotation on the size of contextual cuing in the testing phase.

Method

Subjects. A total of 66 subjects recruited from Vanderbilt University and the local community participated in this study as paid volunteers. One subject in the 45° condition was dropped and replaced because of a clear lack of effort in the guessing task (the same key was pressed throughout), and another subject in the 15° condition was dropped and replaced owing to an excessively high error rate in the search task. All the subjects had normal or corrected-to-normal vision.

Search task: Design. The three main variables were configuration (old vs. new) and block (1–35), manipulated within subjects, and rotation (0°, 15°, 30°, and 45°), varied between subjects.

Each block contained 16 trials, 8 for each configuration type. The old set consisted of eight configurations that were repeated across blocks. A randomly chosen target always appeared in the same location within any particular old configuration. Thus, the spatial context of the target in each old configuration was predictive of the target's location. The new set consisted of eight configurations that were randomly generated for each block. To control for target location repetition effects in old displays, targets in new displays also appeared in one of 8 spatial locations throughout the experiment. In other words, 16 spatial locations were used equally often for the targets in both conditions (8 locations for old, 8 locations for new). To permit the guessing task at the end of the study, each set of stimuli (i.e., new and old) contained an equal number of target locations in each quadrant of the display. Importantly, the average eccentricity (near or far) and the positioning of target locations (up, down, left, or right) in each set were equal. Thus, any search performance differences between old and new configurations could not be due to probabilistic or saliency effects but, rather, must be attributed to the learned associations between global contexts and embedded target locations.

For training, viewpoint rotation was varied from 0°, 15°, 30°, and 45° between subjects, and each rotation condition contained 16 subjects. The range and increment of rotation was based on Diwadkar and McNamara (1997), as well as on pilot studies in our lab. The
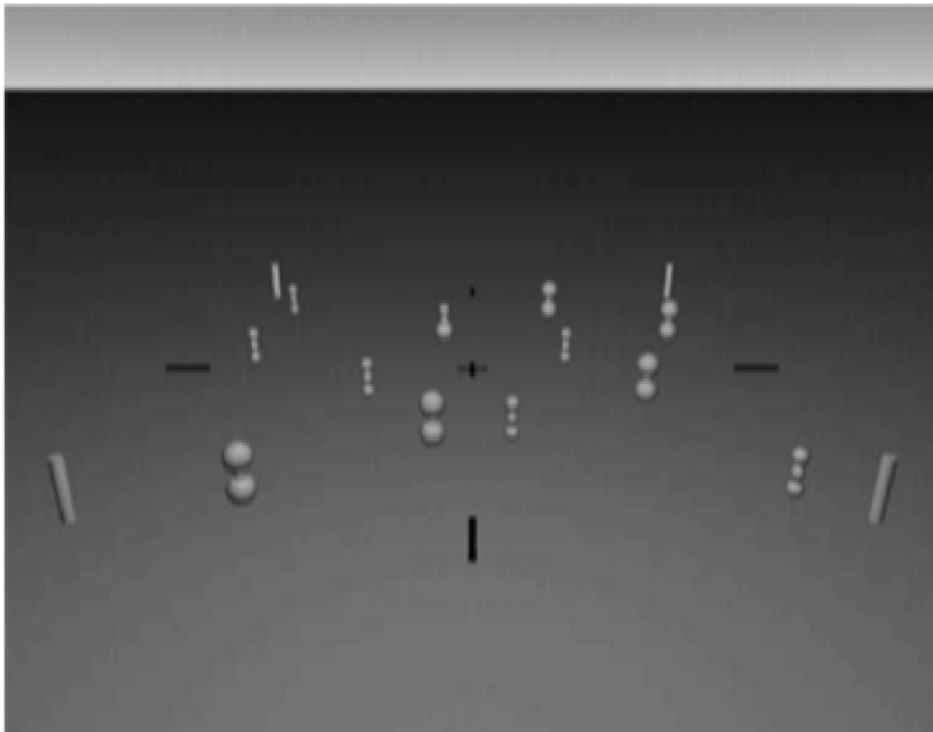
TRAINING
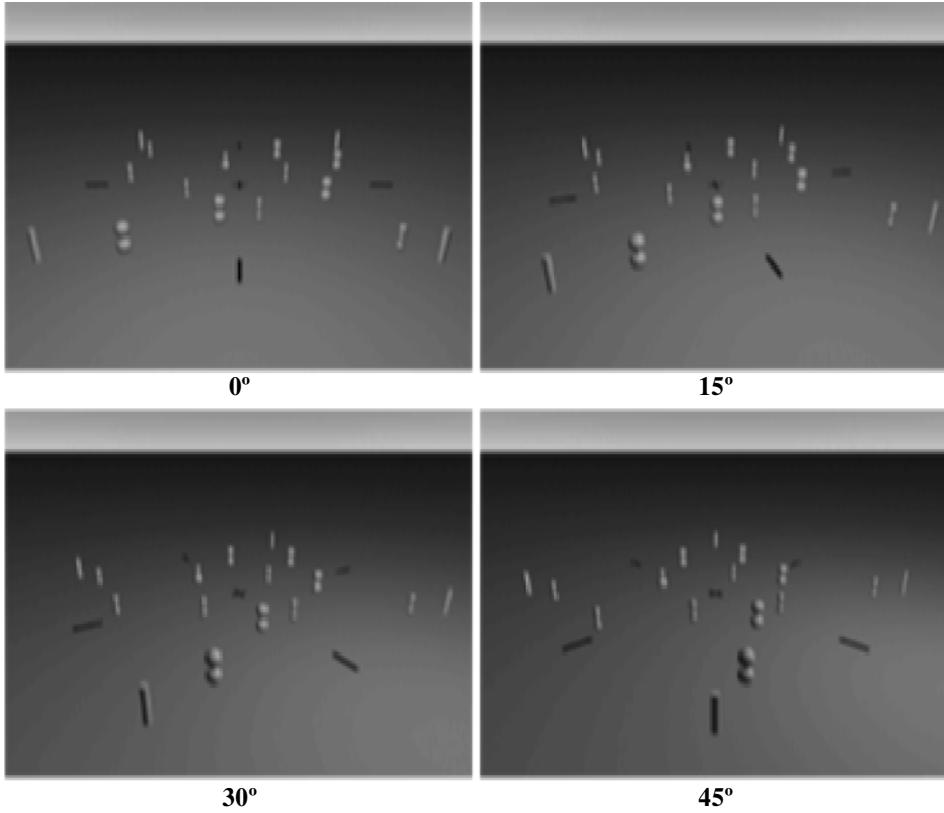
TESTING

Figure 1. Example scenes used in the experiment. Each picture contains the same scene viewed from a different angle.

between-subjects manipulation of rotation differed from past scene recognition tasks, which typically trained subjects on a single perspective and compared performance on different rotations within subjects (e.g., Diwadkar & McNamara, 1997). Such a design was not feasible for our study, because the large number of trials required to obtain a reliable contextual cuing effect discouraged us from testing multiple viewpoints within subjects. Also, we felt that presenting too many different viewpoint rotations from trial to trial would be disorienting, as well as nonecological, in light of the importance of perceived viewer-centered perspective (Huttenlocher & Presson, 1973; Rieser, 1989; Simons & Wang, 1998). Finally, presenting different viewpoint rotations across blocks within observers would not be optimal, because of learning effects. As such, viewpoint was manipulated as a between-subjects variable.

Within each rotation condition, the same set of old and new displays was used for all the subjects. Different new display sets were used in each of the 30 training blocks. The order of new displays was randomized for each observer, so that each subject experienced a different sequence of new sets. There are no known asymmetries between clockwise versus counterclockwise rotations (Shepard & Cooper, 1982), so we chose to rotate viewpoints in the clockwise direction only, in order to economize the design.

In the testing condition, we used a common viewpoint (0°) for all the subjects in order to minimize noise and to maximize comparability. The testing blocks used the same old configurations as those used in training, viewed from the 0° viewpoint. Old configurations were intermixed with a set of new configurations that were also viewed from the 0° viewpoint. Unlike training, new displays were repeated across all five testing blocks (Chun & Jiang, in press). The rationale for presenting the same new set throughout the testing phase was as follows: We had to repeat old configurations during the testing phase to obtain enough trials for a reliable response time (RT) analysis. However, the subjects would presumably become faster on *old* trials within the testing phase, owing to increased repetitions. To control for such within-testing session learning, we repeated the new configurations in the testing phase as well. Thus, any benefit in search for old displays relative to new displays could be attributed to the repetitions during training, rather than to the repeated exposure during testing.

Search task: Procedure. The subjects searched for a bowling-pin–shaped target among an array of distractors. Each trial displayed a scene containing one target and 11 distractor objects, forming a spatial layout in pseudo–3-D space (see Figure 1).

The target contained a small sphere at one end and a large sphere at the other end. If the large sphere appeared on the bottom, like an upright bowling pin, the target was labeled *up*. If the large sphere appeared on the top, like an upside-down bowling pin, the target was labeled *down*. The number of up and down targets was equated between *old* and *new* trials in each block.

The distractors were shaped like dumbbells; half of them had small spheres at the ends of the dumbbell (*thin* distractors), whereas the other half had large spheres (*fat* distractors). Scenes contained an invisible 8 column × 6 row grid, for a total of 48 spatial locations in which objects could appear. In each image, six distractors appeared in both the front three rows and the back three rows, to control for size effects across different scenes. Each image also contained an equal number of fat and thin distractors.

The experiment began with an instruction screen, followed by a practice block of eight trials. The spatial configurations of the practice block displays were not identical to any of the spatial configurations used in the real experiment. After the practice block was completed, the subjects pressed any key to proceed to the actual experiment.

Each trial began by displaying the *blank* field, which contained an empty green field, a blue sky, and four red poles positioned at the corners of the 8 × 6 grid. The display was divided into four quadrants by a set of black marks placed outside the perimeter of the grid,

in between each of the four poles. A black fixation cross was also placed in the middle of the scene, drawn on the surface plane. After 513 msec, the search array appeared in the blank field. Upon detection of the target, the subjects pressed 5 on the numerical keypad if the target was pointed up or 1 if the target was pointed down. After their responses, the subjects received verbal feedback (i.e. "Correct!" or "Error") in the form of a message that appeared at the top of the screen in the sky. The message lasted 500 msec, after which the next trial began with the presentation of the blank screen. Each trial had a time limit of 4 sec, and any trial that reached this threshold was removed from the analyses. Between blocks, a white screen prompted the subjects to press any key when they wished to proceed.

Guessing task. At the conclusion of the search task, the subjects performed a memory test that was designed to determine whether conscious memory of repeated displays facilitated performance (Chun & Jiang, in press). The subjects were not told at the beginning of the experiment about this test, but they were informed that there would be an extra 5-min task at the end of the experiment. The test consisted of a question section and a guessing task. First, the computer displayed a question asking, "Did you notice that some configurations of the stimuli were repeated from block to block ("y" or "n")?" If the answer was "no," the subject immediately proceeded to the guessing task. If the answer was "yes," the subject answered two additional questions. The first question was, "Around when do you think you started to notice repetitions (Block 1–35)?" The second question was, "Did you explicitly try to memorize the patterns (press "y" or "n")?" After the subjects had finished answering these questions, they proceeded to the guessing task.

In the guessing task, the computer displayed 24 configurations in random order. Eight configurations were from the old set,[1] 8 configurations were from the new set used in the testing phase, and 8 configurations were randomly generated. The randomly generated configurations used the same target locations as the new set. Regardless of training condition, all the images were viewed from the 0° viewpoint in order to be consistent with the testing images.

In all the displays, the target was replaced with a distractor. The task was to guess which quadrant the target would be located in if it were present. The subjects indicated their response by pressing 1–4 on the alphanumeric keyboard, where 1 corresponded to the upper left quadrant, 2 to the upper right quadrant, 3 to the lower left quadrant, and 4 to the lower right quadrant. Since each quadrant contained an equal number of target locations, chance performance was 25%. Responses were not timed during this portion of the experiment, which lasted approximately 5 min.

Apparatus and Stimuli. The experiment was programmed and executed in MATLAB 5.2.1, using the Psychophysics Toolbox extensions (Brainard, 1997; Pelli, 1997). The stimuli were generated using POV-Ray 3.1, a raytracing computer graphics program. POV-Ray uses a left-handed coordinate system in which x denotes left and right, y denotes up and down, and z denotes in and out of the screen ("in" being positive and "out" being negative). Each unit on the POV-Ray scale corresponds to approximately 0.0184 in. on a 15-in. monitor set to 640 × 480 resolution.

Objects appeared on an invisible 8 × 6 grid whose corners were located at (−525, 0, −500), (525, 0, −500), (−525, 0, 500), and (525, 0, 500); these points corresponded to row 6 column 1, row 6 column 8, row 1 column 1, and row 1 column 8, respectively (row 1 is farthest back, column 1 is farthest left). The target was 54 units high, the thin distractor was 48 units high, and the fat distractor was 60 units high. The small sphere of the thin distractors and the target had a radius of 2 units, while the large sphere of the fat distractors and the target had a radius of 3.5 units. Four red cylindrical poles appeared just outside the corners of the grid. Each pole was 50 units high. The poles appeared at (−600, 0, −600), (600, 0, −600), (−600, 0, 600), and (600, 0, 600). The grid appeared on a flat green plane (with equation $y = −10$) that intersected with a blue sky near the "back" of the image ("back" corresponds to the top of the com-
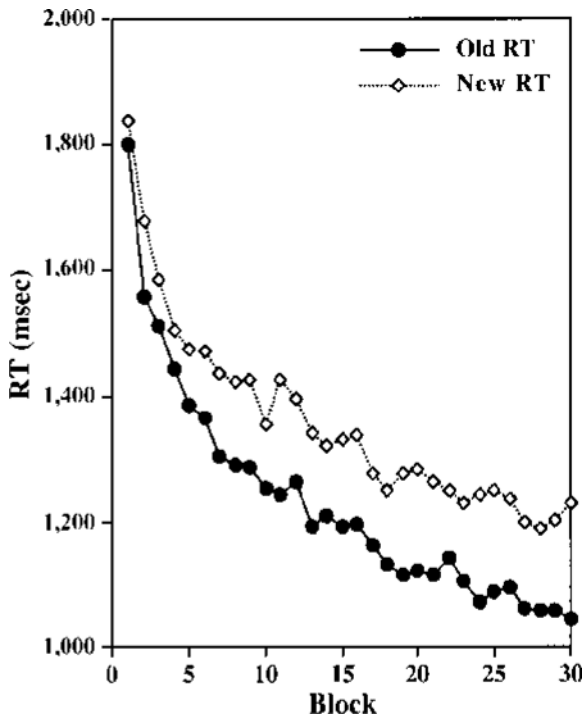
Figure 2. Training data collapsed across all training rotation conditions.

puter screen). Images were rendered at 640 × 480 pixels, and 15-in. monitor settings were adjusted to this resolution so that the images filled up the entire screen.

All the objects in the display were colored yellow to enhance visibility. A sense of depth was maintained by the fact that objects became slightly smaller the farther back they were (Aks & Enns, 1996). The images were rendered through the vantage point of a camera located at (0, 575, −1,500) that pointed its lens toward the point (0, 0, 0), subtending a camera viewing angle of 20.97°. These points were chosen so that the entire grid was visible and centered at the middle of the screen. A light source was placed behind the camera at the point (0, 750, −1,000) so that a light gradient from light to dark existed from front to back, further enhancing the sense of depth (Aks & Enns, 1996). The rotated displays were generated by rotating the camera the appropriate number of degrees. This was formally identical to rotating the subject counterclockwise around a circle with a radius of 1,606.43 POV-Ray units (i.e., the distance between [0, 575, −1,500] and [0, 0, 0]). Sample displays of the four rotations are shown in Figure 1.

Results

Search task: Errors. Overall error rates for the four rotation conditions were low (1.29%, 1.45%, 1.72%, and 1.61% for 0°, 15°, 30°, and 45°, respectively). An analysis of variance (ANOVA) revealed no significant effects of configuration or rotation on error rates during the training phase (all $Fs < 1$). There was a significant effect of block on error rate during training [$F(29,1740) = 3.78, p < .001$]. An analysis showed that this was due mainly to a higher proportion of errors during the first block, presumably because the subjects needed practice to master the task. Indeed, the error rate in the first block was

5.47%, whereas the average error rate in all the other blocks was 1.40%. There was also a significant interaction of rotation and block [$F(87,1740) = 1.45, p < .005$]. This was due to a higher error rate on the first block in the 30° and 45° conditions (10.2% and 7.03%, respectively, as compared with 1.56% and 3.13% for 0° and 15°). A separate analysis revealed that there were no significant effects of or interactions between rotation, block, and configuration on error rates in the testing phase (all $ps > .17$).

Search task: Response times. The mean RTs for all correct responses within each block were calculated and submitted to a repeated measures ANOVA, with configuration (old vs. new) and block (1–35) as within-subjects factors and rotation as a between-subjects factor. RTs that exceeded 4,000 msec were discarded; fewer than 1% of all the trials were discarded by this procedure in all conditions (0.50%, 0.71%, 0.52%, and 0.44% for 0°, 15°, 30°, and 45°, respectively). An ANOVA revealed no main effects of rotation or configuration on timeout rates (all $ps > .55$), but there was a main effect of block ($p < .001$) due to poorer performance at the beginning of training. A similar analysis revealed no effects of block, configuration, or rotation on timeout rates in the testing phase (all $ps > .25$). Subjects with combined error and timeout rates exceeding 10% were discarded; only 1 subject was removed by this criterion.

During the training phase, the subjects became faster at detecting targets in old displays. An ANOVA revealed main effects of configuration [$F(1,60) = 377.39, p < .001$], as well as of block [$F(29,1740) = 68.01, p < .001$]. There were no main effects of rotation group ($F < 1$), and all interactions with rotation group were not significant (all $ps > .36$). This implies that the training curves for the four groups of subjects were similar, and we therefore show the training data collapsed across all subjects (see Figure 2).

Across all subjects, although the interaction between configuration and block was not significant [$F(29,1740) = 1.16, p > .25$], an analysis restricted to Block 1 and Block 30 revealed a significant interaction [$F(1,63) = 7.29, p < .009$]. Thus, a significant contextual cuing effect was observed. Figure 2 clearly illustrates how performance reliably improved for the *old* condition with increases in number of blocks.

The main question is whether training rotation affected contextual cuing for 0° testing displays. In the testing phase, there was a significant effect of configuration [$F(1,60) = 16.38, p < .001$], demonstrating the contextual cuing effect. Importantly, the interaction between configuration and rotation was significant [$F(3,60) = 2.88, p < .05$]. As can be seen in Table 1, the magnitude of the contextual cuing effect in the testing phase decreased with increasing distance from training. The size of the contextual cuing effect in the testing phase (new RT − old RT) was 137.9 msec [$t(15) = 3.48, p < .003$], 90.1 msec [$t(15) = 3.12, p < .007$], −3.73 msec ($p > .92$), and 55.7 msec ($p > .10$) for the 0°, 15°, 30°, and 45° rotation conditions, respectively. Although it appears that there

Table 1
Mean Response Times (RTs, in Milliseconds) for Old and New
Conditions in Each Rotation During the Testing Blocks

Rotation (°)	Old	New	New − Old
0	1,036.0	1,173.9	137.9
15	1,128.5	1,218.6	90.1
30	1,164.7	1,161.0	−3.7
45	1,185.2	1,240.9	55.7

Note—The difference between old and new RTs constitutes the size of the contextual cueing effect.

was a contextual cueing effect in the 45° condition, the positive value was mainly due to one outlier who showed a large cuing effect. When this outlier was removed, the average contextual cuing effect dropped to 35 msec ($p >$ .21). Furthermore, a contrast analysis revealed a significant linear trend [$F(1,60) = 4.97, p < .05$; Rosenthal & Rosnow, 1991], demonstrating that the contextual cuing effect decreased with increasing distance from training viewpoint. The main effect of block was significant [$F(4,240) = 8.78, p < .001$]. The main effect of rotation was not significant ($F < 1$), nor were there any interactions between rotation and block or configuration and block (all $ps > .18$).

Explicit recognition task. The results of the recognition test are summarized in Table 2. Mean accuracy for old displays was 26.5%, 28.9%, 28.9%, and 28.1% for 0°, 15°, 30°, and 45°, respectively; overall accuracy was 28.1%. None of these hit rates was significantly different from chance, which was taken to be a hit rate of 25% (for each condition, all $ps > .48$; for overall hit rate, $p > .18$). For the new testing phase displays and the randomly generated displays, the subjects also performed at chance (all $ps > .16$). Performance within each rotation was compared for old images, testing phase new images, and randomly generated new images. There was no significant difference between performance on any of the image types (all $ps > .232$). The same was true for data pooled across all subjects (all $ps > .475$).

Overall, 16 subjects indicated that they had not noticed repetitions, whereas 48 indicated that they had. Contextual cuing and the recognition test data were analyzed as a function of awareness of repetitions. We can summarize the contextual cuing effect by pooling across Blocks 16–30 (the second half of training), following prior convention (Chun & Jiang, 1998). According to this measure, the unaware subjects had a mean contextual cuing effect of 130.3 msec, whereas the aware subjects had a mean of 147.2 msec; this difference was not significant ($p > .25$). Overall on the guessing task, the unaware subjects performed at 32.0% correct for old displays, as compared with 26.8% for the aware subjects; however, this difference was not significant ($p > .33$). Similarly, there was no effect of awareness on the hit rate for new testing set images and the randomly generated images (all $ps > .43$). When the data were analyzed within each condition, performance of aware and unaware subjects on each of the three types of displays was not different from chance (all

$ps > .13$). Pooled across all aware subjects, performance in each type of display was not different from chance (all $ps > .24$); the same was true of unaware subjects (all $ps > .13$). Overall, the subjects in the aware and the unaware groups did not perform differently from each other on the three types of displays (all $ps > .37$).

Finally, 8 subjects overall indicated that they had explicitly tried to memorize the displays during the contextual cuing task. Because of the small number of data points, it was not feasible to perform formal statistical analyses on the performance of these subjects. Qualitatively, however, the magnitude of the contextual cuing effect for the subjects who reported explicit encoding strategies was essentially equal to the group mean, and their performance on the guessing task was not different from the mean either.

Discussion

During training, search performance became faster for old displays relative to new displays, demonstrating the contextual cuing effect. This indicates that the subjects encoded the spatial contexts of targets and used this information to guide search. Because objects were arrayed in apparent depth, the present finding indicates that contextual cuing generalizes to artificial scenes with pseudo–3-D layouts. Similar benefits from repeated contexts were also observed with real-world scenes (Sheinberg & Logothetis, 1998) and with virtual reality reaching tasks (Hayhoe, 2001), as well as with saccadic eye movement measures (Peterson & Kramer, 2001). All of these findings converge to suggest that contextual cuing has strong ecological validity in everyday visual perception and action.

In addition, the present study revealed contextual representations to be viewpoint dependent. After training on scenes viewed from a single viewpoint, the subjects were tested at various rotations away from the training viewpoint. Contextual cuing diminished with increased rotation from 0° to 15°, and at 30° and 45°, the contextual cuing effect became nonsignificant. These results parallel the results of Diwadkar and McNamara (1997), who found a decreased effect of training on a scene recognition task with increased distance from the training view-

Table 2
Hit Rates for Different Testing Conditions
in the Location Guessing Test

Rotation (°)	Old Set (%)	Repeated New Test Set (%)	Random New Set (%)
0	26.5	23.4	28.9
15	28.9	28.1	24.2
30	28.9	24.2	26.6
45	28.1	28.9	29.7
Overall	28.1	26.2	27.3

Note—"Old Set" refers to the repeated configurations used throughout the experiment. "Repeated New Test Set" refers to the new images that subjects had never experienced before (different from training) but were repeated throughout the testing phase. "Random New Set" refers to new images that were not repeated throughout the testing phase (i.e., newly generated in each block).

point. They are also corroborated by a pilot experiment in which the magnitude of contextual cuing decreased with increasing distance from training on a task very similar to that in the present experiments (the magnitude of contextual cuing was 147.9, 171.7, 110.9, and −28.0 msec for 0º, 15º, 30º, and 45º, respectively).

The demonstration of viewpoint dependency supports previous claims that contextual cuing is driven by instance-based representations of viewed displays (Chun & Jiang, 1998). In our study, subjects displayed a facilitation for old displays. This facilitation cannot be explained solely by simple improvements in search, which are reflected in the general decrease in RT for both old and new configurations. Rather, the contextual cuing effect is driven mainly by retrieval of specific memory traces for different contexts and their associated locations.

Although we claim that contextual scene representations are viewpoint dependent in a 3-D representation, our rotation manipulations also increased the dissimilarity of the 2-D retinal images from the original trained scenes. Because of this, one may question whether our results reflect viewpoint effects in 3-D perceptual representations or dissimilarity effects in 2-D representations. Indeed, distorting a 2-D test image away from a 2-D trained image reduces the amount of learning transfer in counting tasks (Lassaline & Logan, 1993; Palmeri, 1997), as well as in contextual cuing search tasks (Jiang & Chun, 2001; Olson & Chun, 2002).

However, while acknowledging the contributions of 2-D image dissimilarity, we believe that our results reflect viewpoint dependency in a 3-D representation. First, past demonstrations of object or scene rotation effects also involved changing the similarity of the 2-D images to be matched. In fact, the logic of our study was identical to that used previously to claim viewpoint dependency in object or scene recognition (Diwadkar & McNamara 1997; Tarr, 1995). Put simply, the presence of rotation effects effectively rules out models that postulate viewpoint-independent representations for scene layout. Second, a recent study directly showed that depth information is encoded in contextual cuing tasks (Kawahara, 2002). Subjects were trained on search arrays that had front and back depth planes based on binocular disparity. The items on one depth plane were correlated with an embedded target location in order to produce a contextual cuing effect. After training, the disparity was switched so that the front plane appeared in the back and the back plane appeared in the front. Even though this manipulation did not affect the 2-D retinal image, the perceived change in depth relations reduced the contextual cuing effect, suggesting that contextual learning mechanisms encode depth. The combination of the Kawahara study and the present study suggests that when scenes depict depth, contextual representations encode 3-D layout in a viewpoint-dependent manner.

In regard to the mode of learning, our results confirm that contextual learning in naturalistic displays can be implicit (Chun & Jiang, 1998, in press). Although the subjects noticed repetitions, the vast majority reported that they did not try to encode the displays. In the guessing task used to directly assess conscious memory, the subjects were not able to accurately estimate the target locations above chance levels for old displays, despite significant benefits in search performance for the same target locations and associated contexts. Moreover, awareness of repetition or conscious efforts to memorize displays had no effect on performance in either the search task or the recognition test.

Although we emphasized the implicit nature of learning and memory in our task, we are not claiming that scene learning must occur implicitly. Undoubtedly, observers can explicitly learn novel scenes, as has been demonstrated in past studies of spatial layout (Diwadkar & McNamara, 1997; Levine et al., 1982; Shelton & McNamara, 1997). In addition, explicit learning of scenes is consistent with the intuition that one must actively search for landmarks when navigating in a novel environment.

In summary, our study complements prior studies by demonstrating scene context learning in a task in which subjects were not asked to attend to and encode global scene layouts. In everyday visual experience, much scene context information may be encoded implicitly to provide important cues for such visual behaviors as search and navigation (Chun, 2000; Chun & Nakayama, 2000). The present study suggests that implicitly acquired scene context representations are viewpoint dependent.

REFERENCES

Aks, D. J., & Enns, J. T. (1996). Visual search for size is influenced by background texture gradient. *Journal of Experimental Psychology: Human Perception & Performance*, **22**, 1467-1481.

Berry, D. C., & Dienes, Z. (1993). *Implicit learning*. Hove, U.K.: Erlbaum.

Biederman, I., & Gerhardstein, P. C. (1993). Recognizing depth-rotated objects: Evidence and conditions for three-dimensional viewpoint invariance. *Journal of Experimental Psychology: Human Perception & Performance*, **19**, 1162-1182.

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

Brainard, D. H. (1997). The Psychophysics Toolbox. *Spatial Vision*, **10**, 433-436.

Bravo, M. J., & Nakayama, K. (1992). The role of attention in different visual-search tasks. *Perception & Psychophysics*, **51**, 465-472.

Bulthoff, H. H., & Edelman, S. (1992). Psychophysical support for a two-dimensional view interpolation theory of object recognition. *Proceedings of the National Academy of Sciences*, **89**, 60-64.

Chun, M. M. (2000). Contextual cuing of visual attention. *Trends in Cognitive Sciences*, **4**, 170-178.

Chun, M. M., & Jiang, Y. (1998). Contextual cuing: Implicit learning and memory of visual context guides spatial attention. *Cognitive Psychology*, **36**, 28-71.

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

Chun, M. M., & Jiang, Y. (in press). Implicit, long-term spatial contextual memory. *Journal of Experimental Psychology: Learning, Memory, & Cognition*.

Chun, M. M., & Nakayama, K. (2000). On the functional role of implicit visual memory for the adaptive deployment of attention across scenes. *Visual Cognition*, **7**, 65-81.

Chun, M. M., & Wolfe, J. M. (2001). Visual attention. In B. Goldstein

(Ed.), *Blackwell handbook of perception* (pp. 272-310). Oxford: Blackwell.

DIWADKAR, V. A., & MCNAMARA, T. P. (1997). Viewpoint dependence in scene recognition. *Psychological Science*, **8**, 302-307.

EGETH, H. E., VIRZI, R. A., & GARBART, H. (1984). Searching for conjunctively defined targets. *Journal of Experimental Psychology: Human Perception & Performance*, **10**, 32-39.

HAYHOE, M. M. (2001, May). *Memory for spatial structure in saccadic targeting.* Paper presented at the annual meeting of the Vision Sciences Society, Sarasota, FL.

HUTTENLOCHER, J., & PRESSON, C. C. (1973). Mental rotation and the perspective problem. *Cognitive Psychology*, **4**, 277-299.

JACOBY, L. L., & WITHERSPOON, D. (1982). Remembering without awareness. *Canadian Journal of Psychology*, **36**, 300-324.

JIANG, Y., & CHUN, M. M. (2001). Selective attention modulates implicit learning. *Quarterly Journal of Experimental Psychology*, **54A**, 1105-1124.

JOHNSTON, W. A., HAWLEY, K. J., PLEW, S. H., ELLIOTT, J. M., & DEWITT, M. J. (1990). Attention capture by novel stimuli. *Journal of Experimental Psychology: General*, **119**, 397-411.

KAWAHARA, J. (2002, May). *Contextual cuing effect in three dimensional layouts.* Poster session presented at the annual meeting of the Vision Sciences Society, Sarasota, FL.

LASSALINE, M. E., & LOGAN, G. D. (1993). Memory-based automaticity in the discrimination of visual numerosity. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **19**, 561-581.

LEVINE, M., JANKOVIC, I. N., & PALIC, M. (1982). Principles of spatial problem solving. *Journal of Experimental Psychology: General*, **111**, 157-175.

LOGAN, G. D. (1988). Towards an instance theory of automatization. *Psychological Review*, **95**, 492-527.

MARR, D., & NISHIHARA, H. K. (1978). Representation and recognition of the spatial organization of three-dimensional shapes. *Proceedings of the Royal Society of London: Series B*, **200**, 269-294.

MILLER, J. (1988). Components of the location probability effect in visual search tasks. *Journal of Experimental Psychology: Human Perception & Performance*, **14**, 453-471.

OLSON, I. R., & CHUN, M. M. (2001). Temporal contextual cuing of visual attention. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **27**, 1299-1313.

OLSON, I. R., & CHUN, M. M. (2002). Perceptual constraints on implicit learning of spatial context. *Visual Cognition*, **9**, 273-302.

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

PALMERI, T. J. (1997). Exemplar similarity and the development of automaticity. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **23**, 324-354.

PASHLER, H. (1998). *The psychology of attention.* Cambridge, MA: MIT Press.

PELLI, D. G. (1997). The VideoToolbox software for visual psychophysics: Transforming numbers into movies. *Spatial Vision*, **10**, 437-442.

PETERSON, M. S., & KRAMER, A. F. (2001). Attentional guidance of the eyes by contextual information and abrupt onsets. *Perception & Psychophysics*, **63**, 1239-1249.

REBER, A. S. (1989). Implicit learning and tacit knowledge. *Journal of Experimental Psychology: General*, **118**, 219-235.

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.

RIESER, J. J. (1989). Access to knowledge of spatial structure at novel points of observation. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **15**, 1157-1165.

ROSENTHAL, R., & ROSNOW, R. L. (1991). *Essentials of behavioral research: Methods and data analysis* (2nd ed.*).* New York: McGraw-Hill.

SCHACTER, D. L. (1987). Implicit memory: History and current status. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **13**, 501-518.

SCHNEIDER, W., & SHIFFRIN, R. M. (1977). Controlled and automatic human information processing: I. Detection, search and attention. *Psychological Review*, **84**, 1-66.

SHANKS, D. R., & ST. JOHN, M. F. (1994). Characteristics of dissociable learning systems. *Behavioral & Brain Sciences*, **17**, 367-395.

SHAW, M. L. (1978). A capacity allocation model for reaction time. *Journal of Experimental Psychology: Human Perception & Performance*, **4**, 586-598.

SHAW, M. L., & SHAW, P. (1977). Optimal allocation of cognitive resources to spatial locations. *Journal of Experimental Psychology: Human Perception & Performance*, **3**, 201-211.

SHEINBERG, D. L., & LOGOTHETIS, N. K. (1998). Implicit memory for scenes guides visual exploration in monkey. *Society for Neuroscience Abstracts*, **24**, (Pt. 2), 1506.

SHELTON, A. L., & MCNAMARA, T. P. (1997). Multiple views of spatial memory. *Psychonomic Bulletin & Review*, **4**, 102-106.

SHEPARD, R. N., & COOPER, L. A. (1982). *Mental images and their transformations.* Cambridge, MA: MIT Press.

SIMONS, D. J., & LEVIN, D. T. (1997). Change blindness. *Trends in Cognitive Sciences*, **1**, 261-267.

SIMONS, D. J., & WANG, R.-F. (1998). Perceiving real-world viewpoint changes. *Psychological Science*, **9**, 315-320.

SPERLING, G. (1960). The information available in brief visual presentations. *Psychological Monographs: General & Applied*, **74**, 1-29.

SQUIRE, L. R. (1992). Declarative and nondeclarative memory: Multiple brain systems supporting learning and memory. *Journal of Cognitive Neuroscience*, **99**, 195-231.

STADLER, M. A., & FRENSCH, P. A. (Eds.) (1998). *Handbook of implicit learning.* Thousand Oaks, CA: Sage.

TARR, M. J. (1995). Rotating objects to recognize them: A case study on the role of viewpoint dependency in the recognition of three-dimensional objects. *Psychonomic Bulletin & Review*, **2**, 55-82.

TREISMAN, A. M., & GELADE, G. (1980). A feature-integration theory of attention. *Cognitive Psychology*, **12**, 97-136.

TREISMAN, A. [M.], & GORMICAN, S. (1988). Feature analysis in early vision: Evidence from search asymmetries. *Psychological Review*, **95**, 15-48.

WANG, Q., CAVANAGH, P., & GREEN, M. (1994). Familiarity and pop-out in visual search. *Perception & Psychophysics*, **56**, 495-500.

WOLFE, J. M. (1994a). Guided Search 2.0: A revised model of visual search. *Psychonomic Bulletin & Review*, **1**, 202-238.

WOLFE, J. M. (1994b). Visual search in continuous, naturalistic stimuli. *Vision Research*, **34**, 1187-1195.

YANTIS, S., & JONIDES, J. (1984). Abrupt visual onsets and selective attention: Evidence from visual search. *Journal of Experimental Psychology: Human Perception & Performance*, **10**, 601-621.

NOTES

1. The spatial distributions of the fat and thin distractors within the old and repeated new test set configurations were preserved across repetitions during the search task. However, the thin and fat distractor distributions within the old and repeated new test displays were not preserved in the guessing task configurations, owing to a programming error. The global configurations were maintained, of course. Although this confound is not ideal, on the basis of two prior studies, there is little reason for concern. First, Chun and Jiang (1998) demonstrated that contextual cuing was robust even when the identities of the distractors were completely changed from one stimulus set to another, as long as the configurations and associated target locations were preserved. Second, Chun and Jiang (in press) employed exactly the same guessing task with displays that were identical to those used during the search task (except that the target was replaced by a distractor), and subjects performed at chance levels for both old and new displays in two separate experiments.