Semantic memory for contextual regularities within and across scene categories: Evidence from eye movements

JAMES R. BROCKMOLE
University of Notre Dame, Notre Dame, Indiana

AND

MELISSA LE-HOA VÕ
Harvard Medical School, Boston, Massachusetts

When encountering familiar scenes, observers can use item-specific memory to facilitate the guidance of attention to objects appearing in known locations or configurations. Here, we investigated how memory for relational contingencies that emerge across different scenes can be exploited to guide attention. Participants searched for letter targets embedded in pictures of bedrooms. In a between-subjects manipulation, targets were either always on a bed pillow or randomly positioned. When targets were systematically located within scenes, search for targets became more efficient. Importantly, this learning transferred to bedrooms without pillows, ruling out learning that is based on perceptual contingencies. Learning also transferred to living room scenes, but it did not transfer to kitchen scenes, even though both scene types contained pillows. These results suggest that statistical regularities abstracted across a range of stimuli are governed by semantic expectations regarding the presence of target-predicting local landmarks. Moreover, explicit awareness of these contingencies led to a central tendency bias in recall memory for precise target positions that is similar to the spatial category effects observed in landmark memory. These results broaden the scope of conditions under which contextual cuing operates and demonstrate how semantic memory plays a causal and independent role in the learning of associations between objects in real-world scenes.

Although the visual world is complex, it is typically not random. In addition to the laws of physics and the social or cultural norms that constrain the arrangement of objects in a scene, other contextual regularities may provide order to our environment. The location of objects is often systematically, functionally, and meaningfully related to the position of other objects (e.g., televisions in front of sofas), certain classes of objects tend to co-occur (e.g., refrigerators, sinks, and food), and the actions afforded by specific objects are often engaged in particular temporal sequences (e.g., the movement of cars at an intersection). By encoding and maintaining such regularities in long-term memory, observers can reduce the cognitive complexity of visual environments, which in turn can result in more efficient action and behavior (see Chun & Turk-Browne, 2008, for a thorough review). This extraction of spatial and temporal regularities from the visual world in the service of enhancing performance on visually guided tasks is often called *visual statistical learning* (VSL).

A variety of experimental paradigms have been used to study the mechanisms behind and applications of VSL, but their common denominator as been an effort to characterize the representations that encode covariation among visual objects or events. Most efforts to elucidate the structure and content of the representations involved in VSL have investigated the types of regularities in an environment that can be learned (see, e.g., Baldwin, Andersson, Saffran, & Meyer, 2008; Chun & Jiang, 1998; Fiser & Aslin, 2001, 2002, 2005; Geng & Behrmann, 2002, 2005; Kirkham, Slemmer, & Johnson, 2002; Olson & Chun, 2001; Turk-Browne, Isola, Scholl, & Treat, 2008; Turk-Browne, Jungé, & Scholl, 2005; Turk-Browne, Scholl, Chun, & Johnson, 2009). A complementary approach has been to ask how readily one can transfer learned associations to novel contexts or situations (see, e.g., Brady & Oliva, 2008; Brockmole, Castelhano, & Henderson, 2006; Crist, Kapadia, Westheimer, & Gilbert, 1997; Ehinger & Brockmole, 2008; Jiang & Song, 2005; Turk-Browne & Scholl, 2009). Whereas the first approach defines human sensitivity to regularities in the environment, the second defines the flexibility with which statistical representations can be applied to situations that only partially replicate the original environmental circumstances present at the time of learning. By jointly considering what can be learned and how flexibly the resulting representation can be used to guide behavior, the contents and specificity

---

J. R. Brockmole, james.brockmole@nd.edu

of long-term visual memory gained through VSL can be determined.

Our focus in the present study was on both the acquisition and transfer of learning within a type of VSL called *contextual cuing*, in which repeated exposure to a specific arrangement of target and distractor items leads to a progressively more efficient search (Chun & Jiang, 1998). This improved efficiency has been evidenced by decreased search times (Brockmole & Henderson, 2006b; Chun & Jiang, 1998) and by fewer (Brockmole & Henderson, 2006a; Peterson & Kramer, 2001) and more direct eye movements (Brockmole & Henderson, 2006a) to the target. Although contextual cuing clearly demonstrates that memory for specific spatial configurations can be used to improve the efficiency of visually guided search, the nature of the information that people use to do this remains an open and important avenue of research. The aim of the present research was to investigate the degree to which learning can be abstracted away from specific learning contexts.

On the one hand, a body of evidence indicates that contextual cuing involves the learning of associations between consistently mapped visual features within specific scene exemplars. For example, the visual similarity of and spatial distance between targets and context-defining distractors (Jiang & Chun, 2001; Olson & Chun, 2002) and the proportion of a visual display that is repeated across viewing encounters (Brockmole et al., 2006) all affect contextual cuing. In essence, the greater the dissimilarity between repeating displays, the less contextual cuing is observed. These findings therefore suggest that contextual cuing is strongly item specific.

On the other hand, spatial associations among objects do not stem solely from learned associations between the consistently mapped visual features composing a display. In real-world scenes, visual features such as color can be altered without disrupting associative learning, provided that the changes do not alter the identity of the scene (Ehinger & Brockmole, 2008). Likewise, when visual features are held constant, one's ability to identify the objects, scenes, and their interrelations affects the speed and magnitude of contextual learning (Brockmole, Hambrick, Windisch, & Henderson, 2008; Brockmole & Henderson, 2006b). These studies indicate that learning can be at least partially abstracted away from specific featural aspects of a scene, provided that semantic knowledge including (but probably not limited to) categorical-identity information can be used to access memory for the spatial associations among objects in a scene.

From the evidence cited above, it appears that both item-specific features and abstracted category information can influence recall and recollection of learned statistical regularities. To date, however, the contribution of abstracted learning has not been partialed out from that of item-based memory, because the organizing principle of contextual cuing paradigms has been the repetition of context–target pairings that are specific to individual scenes. This leaves unanswered the questions of how statistical regularities common to multiple scenes are learned and how they are subsequently employed in the service of visually guided tasks. In this study, we aimed to directly address these issues.

Our first goal was to demonstrate that contextual cuing can result when target–scene regularities exist across scenes with common identity but no perceptual overlap. To do this, we asked participants to search through photographs of bedrooms for an artificially embedded "T" or "L" (Brockmole & Henderson, 2006b). Letter targets were used because their locations could not be predicted a priori from scene content (e.g., their placement is not constrained by physical laws such as gravity), and thus we could evaluate the development of interobject associations independent of participants' preexisting knowledge of typical scene structures. For all participants, targets were placed in the same (x, y) coordinates on the display. By varying target–scene pairings (see Figure 1), we created a situation for one group of participants for whom the search target was always located on a bed pillow (*predictable* condition) and a situation for another group for whom the target was placed pseudorandomly with regard to local scene content (*random* condition). If observers are sensitive to covariation existant across scenes, search should become progressively more efficient in the predictable condition than in the random condition. Furthermore, because each scene is shown only once, any such increase in search efficiency will depend on the observers' ability to abstract associations common to all scenes rather than their ability to memorize target locations within specific scenes.

To quantify the efficiency of search, participants' eye movements were monitored, and gaze behaviors were used to measure search efficiency. We used gaze instead of response time (RT) as a measure of attentional guidance because there are current debates regarding the interpretation of RTs in contextual cuing tasks (see Kunar, Flusberg, Horowitz, & Wolfe, 2007). We examined three gaze measures: the elapsed time prior to fixating (i.e., finding) the target, the number of fixations made prior to fixating the target, and the scan-path ratio (see Brockmole & Henderson, 2006a). This last measure contrasts the total amplitude of all saccades prior to fixating the target with the linear distance between the participants' initial fixation point and the search target (Henderson, Weeks, & Hollingworth, 1999). A perfectly efficient search would yield a ratio of 1; as the eye-movement path becomes less direct, the scan-path ratio increases. In addition to these global measures of search efficiency that collapse across multiple eye movements, we also analyzed search efficiency in terms of individual fixations by computing the distance between sequential points of fixation and the target. Although not orthogonal, these measures characterize search behavior in different ways, and consistency across these dependent variables would provide converging indicators of contextual cuing and its behavioral effects.

In addition to using an analysis of search efficiency to measure learning, we later tested the extent to which participants could explicitly recognize old scenes and recall the target locations within them (see Brockmole & Henderson, 2006b; Chun & Jiang, 1998; Smyth & Shanks, 2008). The purpose of this additional analysis was twofold. First, it provided a measure of metacognitive awareness

Predictable   Random

Example Transfer Trials

Figure 1. Example scenes. The same 32 images were used in the predictable and random search conditions. In the predictable condition (left column), the target was always located on a bed pillow. In the random condition (right column), targets and scenes were shuffled so that the target was located randomly in each scene but in the same (x, y) coordinates across scenes (compare scenes within each row). Bull's-eyes have been added in this illustration to indicate target location (but not size), and they were not present in the experimental stimuli. Example transfer stimuli are depicted at the bottom of the figure.

pertaining to abstract associations. Since Chun and Jiang's (1998) seminal demonstration of implicit (i.e., independent of awareness) contextual cuing, this has been a commonly addressed question, with a variety of situational factors leading to the observation of implicit and explicit effects. Second, and more importantly, it provided a measure of the spatial precision with which targets were encoded into memory. A particular interest in this regard concerned target placement within falsely recognized scenes. In this case, no item-specific memory is available to guide target placement; however, generalized knowledge about target placement common to the studied scenes might be expected to bias spatial judgments regarding target placement.

Our second goal was to investigate the nature of the representations underlying abstracted learning by considering the degree to which the derived relational rules used to guide search can be applied to different scenes with varying levels of similarity to the initial learning context (see Figure 1). Following their initial search through the bedroom scenes, participants in both the predictable and random conditions were exposed to one of three different transfer scenarios. We exposed one group of participants to bedrooms that did not contain any pillows. In this case, the local landmarks predicting search targets were removed. However, we placed targets in locations where a pillow would normally appear. If observers abstract a perceptual rule on the basis of low-level visual features coincident with the presence of a pillow (e.g., small, rectangular, convex), then one might expect limited transfer of learning, since such features would now be absent from the area surrounding the target. If, however, guidance is more semantic in nature—that is, based on expectations that arise from scene category, identity, or gist information—then attention may be systematically directed to locations likely to contain a pillow. In this case, transfer of learning should be rather complete.

In order to determine the flexibility with which any such semantic representations can be applied, we also assessed whether the relational rules abstracted from different scenes are specifically linked to the global contexts in which they are learned. To this end, we showed observers either living rooms (second group) or kitchens (third group) in which the target was again located on a pillow. Although multiple scene categories might predict the presence and location of a pillow, it is possible that the application of abstracted associations will be limited to the scene category in which they were learned (in this case, bedrooms). If so, then any benefits of repetition should be lost on both sets of transfer trials. Such a result may be predicted by prior work demonstrating that scene-category information is a primary guide of attention allocation when item memory is available to guide search (Brockmole et al., 2006; Brockmole & Henderson, 2006a). Another possibility is that transfer of abstracted regularities can occur within any scene that predicts the presence and location of relevant landmarks. For example, living rooms clearly allow for the presence of pillows whose locations are in turn highly predictable (e.g., sofas), but pillows are anomalous in kitchens. If search is driven by a semantically derived rule, as hypothesized above, and

if global category identity is used to invoke these semantic associations (cf. Davenport & Potter, 2004), then one could predict that learning will transfer to living rooms but not to kitchens. This state of affairs may be adaptive, since category-based memory could be used to generate broad, but contextually appropriate, retrieval and search strategies. Note that if search is based purely on the perceptual characteristics of critical local objects (in this case, pillows), then the change to either scene category should cause little disruption in search.

METHOD

Participants

Sixty undergraduate students participated after providing informed consent.

Stimuli and Apparatus

A variety of stimuli were used in the experiment. First, we created 32 full-color photographs of bedrooms in which a single gray "T" or "L" (subtending 0.18 deg horizontally and vertically) was embedded. The luminance of each target was adjusted in each image so that its contrast with the local background was approximately equal in all scenes. This letter constituted the target in a visual search task. Two sets of stimuli were created. In the first, targets were placed on a bed pillow. The precise position of each target on each pillow was randomly chosen under the constraint that targets be clearly visible (i.e., they were not masked or camouflaged by edges, contours, textures, etc.). In the second, photographs and target positions were "shuffled," which had the effect of pseudorandomizing the placement of targets within scenes while keeping the (x, y) coordinates of all targets across the stimulus set unchanged. Care was taken to ensure that the overall visibility of targets in the shuffled image set was equal to that of targets in the predictable image set, and results reported below will confirm that at the beginning of the experiment, before substantial learning was possible, search through the shuffled image set was no more and no less difficult than it was through the predictable image set. In addition to the stimuli in these two image sets, 4 stimuli depicting bedrooms without pillows, 3 stimuli depicting living rooms, and 4 stimuli depicting kitchens were also created (see Figure 1).[1] Within the additional bedroom scenes, targets were located on the mattress near the head of the bed. In the living room scenes, targets were located on a sofa pillow. In the kitchen scenes, targets were located on a bed pillow. Finally, an additional 12 stimuli depicting bedrooms (with pillows) were created in which no targets were present. The manner in which all these experimental stimuli were used during the experiment is described below.

Stimuli were displayed at a resolution of 800 × 600 pixels in 24-bit color on a 17-in. CRT display with a refresh rate of 120 Hz. Chin-rests and forehead rests maintained a viewing distance of 81 cm. Under these conditions, the stimuli subtended 24 deg horizontally and 18 deg vertically. Throughout each trial, the position of each participant's right eye was sampled at a rate of 1000 Hz by an Eye-Link 2K eyetracking system (SR Research, Inc.) running in pupil and corneal-reflection mode, which resulted in an average spatial accuracy of 0.15°. A Microsoft video game controller was used to collect target-classification responses.

Design and Procedure

The experiment was divided into two main sessions. In the first session, participants engaged in a search task in which they located a target within a photograph of a real-world scene as quickly as possible. In the second session, participants were tested on the extent to which they could consciously recognize old scenes and recall the target locations within them. We describe the design and procedures related to each session in turn.

Search session. The search session crossed two search conditions (predictable and random) and three transfer conditions (bedroom, living room, and kitchen). Participants were randomly and equally assigned to one of the six resulting experimental conditions. All participants first viewed 32 bedroom scenes in a different random order. For participants in the conditions involving predictable search, targets were always located on the bed pillow. Participants in these conditions were not informed of this consistency, however. For participants in the conditions involving random search, targets did not covary with particular objects within the scenes even though they appeared in the same (x, y) positions in the display. Following these trials, one third of participants were shown bedroom scenes without pillows (targets on a bed), one third were shown living room scenes (targets on a sofa pillow), and one third were shown kitchen scenes (targets on a bed pillow). These *transfer trials* were unannounced and were not preceded by any break or pause.

Participants began the search session by completing a calibration routine that mapped the output of the eyetracker onto display position. Calibration was monitored throughout the experiment and was adjusted when necessary. At the beginning of each trial, a fixation cross was presented in the center of the display. Participants were instructed to look at this cross, whereupon the search scene was presented. Upon identifying the target, participants pressed one of two buttons corresponding to either "T" or "L." The trial was terminated if a response was not made within 15 sec of scene onset. Participants were told to respond as quickly, but as accurately, as possible.

Scene-recognition and target-recall session. Immediately following the completion of the search session, all participants completed the same recall session, in which they viewed 24 full-color photographs of bedrooms. Twelve of these photographs were randomly selected from the 32 bedroom images used in Experiment 1 (targets were removed, however). The remaining 12 photographs were of bedrooms not previously shown to observers. Participants viewed all 24 scenes in a different random order and classified each as either *old* (included in Experiment 1) or *new* (not included in Experiment 1). For those scenes that were judged to be old, participants then indicated where the target had been located by moving a cursor with a mouse trackball to the location at which they previously found the target. Participants were told that they were under no time constraints to make their responses. Eye movements were not recorded during this session of the experiment.

RESULTS

Whereas the method section is organized simply according to the order in which tests were administered, we organize our report of the experimental results into three thematic sections. First, we describe the effects of learning during visual search by focusing on the first 32 trials of the search session, in which all participants viewed the same bedrooms scenes. Second, we describe the effects of learning on one's ability to explicitly recall the position of targets within the previously viewed scenes. Finally, we consider the extent to which learning transfers to novel search contexts.

Analysis 1: Effects of Learning During Search

Initial analyses focused on the elapsed time, number of fixations, and scan-path ratio prior to fixating the target. For all of these analyses, the target was considered to be fixated when a point of fixation was within 0.85 deg of the center of the target. Trials on which this conservative criterion was not met were excluded from the analyses. Trials were also excluded if search was not completed within the allotted time, track-loss occurred, or calibration was inac-

curate. These trims resulted in the loss of 25% of trials.[2] The 32 bedroom trials were collapsed into four blocks of 8 trials each, and analyses contrasted search behavior for participants in the predictable and random conditions across this block factor.

Trends in the elapsed time prior to fixating the target, the number of fixations prior to fixating the target, and the scan-path ratio are illustrated in Figure 2. We initially sub-
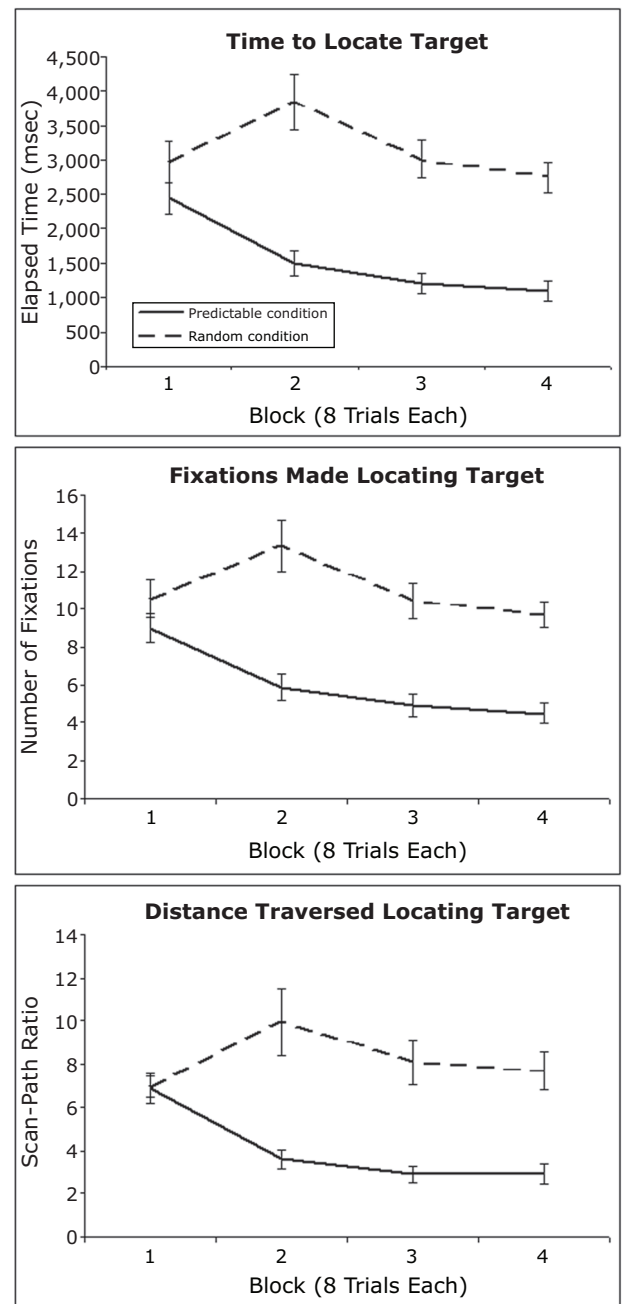
Figure 2. Trends in elapsed time prior to fixating targets (top), number of fixations prior to fixating targets (middle), and scan-path ratio of eye movements executed prior to fixating targets (bottom), illustrated as a function of block and search condition. Error bars depict standard errors of the means.

mitted these trends to separate 2 (search condition: random and predictable) × 3 (transfer condition: bedrooms without pillows, living rooms with pillows, and kitchens with pillows) × 4 (block: 1–4) mixed model ANOVAs. Because the learning trials were same for all participants, in no analysis were the main effects or interactions involving transfer condition reliable (all $ps > .22$). Subsequent analyses therefore collapsed across transfer condition, and a 2 (search condition) × 4 (block) mixed model ANOVA was conducted for each dependent measure. The results of these analyses were entirely consistent, and we therefore report them in parallel.

We first note that in Block 1, all three measures were equivalent across the predictable and random search conditions (all $ps > .17$), indicating that search difficulty was initially equal in both conditions. Despite this initial equivalence, main effects were observed for search condition [time, $F(1,58) = 40.8, p < .001$; fixations, $F(1,58) = 35.9, p < .001$; scan-path ratio, $F(1,58) = 29.0, p < .001$], and block [time, $F(3,174) = 7.88, p < .001$; fixations, $F(3,174) = 7.71, p < .001$; scan-path ratio, $F(3,174) = 2.78, p < .05$]. Critically, these factors interacted [time, $F(3,174) = 7.30, p < .001$; fixations, $F(3,174) = 6.41, p < .001$; scan-path ratio, $F(3,174) = 12.1, p < .001$]. In the random condition, little change in search efficiency was observed across blocks. In Block 1, search required 2,971 msec and 10.5 fixations, and the eyes traversed a path 7.0 times greater than the shortest possible path to the target. Similarly, in Block 4, 2,750 msec, 9.7 fixations, and a scan-path ratio of 7.7 were required. In contrast, in the predictable condition, substantial decreases in all three measures of search efficiency were observed. When targets were consistently located on a bed pillow, an average 2,451 msec, 9.0 fixations, and a scan-path ratio of 6.9 were needed to find the target in Block 1, whereas only 1,097 msec, 4.5 fixations, and a scan-path ratio of 2.9 were required in Block 4. By any measure, then, search efficiency improved approximately

50% by the end of the learning trials when target location was predicted by scene context.

In addition to the above global-level measures of gaze behavior that collapsed across the entire viewing period between scene onset and fixation on the target, we also quantified search behavior at the level of individual eye movements at the very early stages of search (cf. Brockmole & Henderson, 2006a; Peterson & Kramer, 2001). The goal of this analysis was to determine the speed with which learning begins to guide attention. If learning can exert an early influence on search, then evidence for a systematic shift of gaze toward the targets should be evident in the first few fixations on a scene. The point during the experiment at which such a tendency was evident further provides an estimate for the speed with which contingencies were learned. For each of the first five fixations on a scene, we calculated the linear distance between the locus of gaze and the position of the target. Because this analysis concerns search behavior prior to finding the target, all trials were included.

Results are illustrated in Figure 3, broken down by search condition and block. A 2 (search condition) × 4 (block) × 5 (fixation number) mixed model ANOVA revealed reliable main effects of search condition [$F(1,57) = 38.6, p < .001$], block [$F(3,171) = 11.1, p < .001$], and fixation number [$F(4,228) = 7.54, p < .001$]. Critically, the three-way interaction of these factors was also reliable [$F(12,684) = 2.74, p < .001$]. To clarify the nature of this interaction, we conducted separate 4 (block) × 5 (fixation number) repeated measures ANOVAs within each search condition. Within the random condition, the only reliable effect to emerge was a that of fixation number [$F(4,116) = 3.44, p < .05$; all other $ps > .40$]. However, this effect was driven by a small linear increase in the measured distance between the locus of each successive fixation and the target [$F(1,29) = 4.56, p < .05$]. Clearly, gaze showed no propensity to progressively shift toward the target either
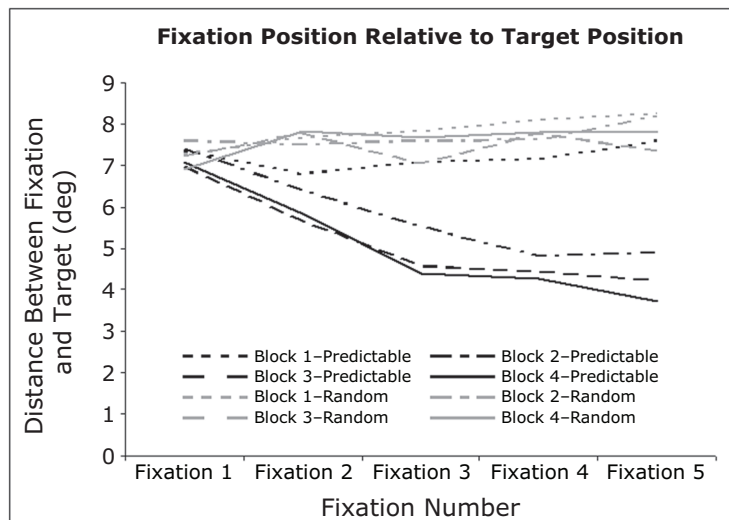

Figure 3. The linear distance between the point of fixation and the target as a function of search condition, block, and fixation number (the first five fixations on each scene) for the 32 bedroom scenes.

across fixations or across blocks. Within the predictable condition, however, the main effects of block [$F(3,84)$ = 16.7, $p < .001$] and fixation number [$F(4,112) = 19.4$, $p < .001$] were reliable, as was the interaction of these factors [$F(12,336) = 6.53$, $p < .001$]. Although the eyes showed no propensity to move toward the target in Block 1, paired t tests contrasting fixation–target distances across blocks and fixations showed that beginning in Block 2, the very first eye movement on the scene systematically moved toward the target ($p < .01$) and that this benefit was magnified in Blocks 3 and 4. These results indicate that the contingencies between targets and scenes were established quickly (i.e., between 9 and 16 repetitions), and, once learned, they exerted an immediate influence on attentional control during scene viewing by driving the very first saccade toward the target.

Summary of findings. Trends in all of the dependent variables reported above demonstrate that search gradually became more efficient in the predictable condition but not in the random condition. The time and number of fixations needed to find the target as well as the total amplitude of the eye movements decreased over trials in the predictable condition, in which targets were always placed on a bed pillow. This increase in the efficiency of search was also observed at the level of individual fixations very early during search, since the eyes moved more directly toward the target as the experiment unfolded.

Analysis 2: Effects of Learning on Target Recall After Search

Initial analyses pertaining to scene recognition and target placement confirmed that none of the dependent measures reported below varied across transfer conditions (all ps > .33). Analyses reported below therefore collapsed across this factor.

Scene recognition. Explicit scene recognition was measured by computing A', which contrasts hit and false-alarm rates and ranges from .5 (chance performance) to 1 (perfect discrimination). A' scores observed for participants assigned to the predictable and random conditions of Experiment 1 were .70 and .73, respectively, indicating equivalent [$t(54) < 1$] and above-chance recognition of old displays in both groups.

Target placement. Although participants in the predictable and random conditions discriminated between old and new scenes with equal accuracy, memory for target position clearly differed between these groups. First, the average error in target localization (i.e., the distance between the actual and participant-indicated target locations) in the predictable group was 3.97 deg, whereas for the random group this error was, on average, 8.16 deg [$t(53) = 6.39$, $p < .001$; 1 participant had a hit rate of 0 and was excluded from this analysis]. By comparison, the expected error in target placement on the basis of randomly selected responses was approximately 10.3 deg.[3] Second, on false-alarm trials (i.e., new scenes classified as old), participants in the predictable group placed targets within 2.78 deg of the bed pillows' center of mass (defined as the middle of the smallest imaginary rectangle that could be drawn around the bed pillows), whereas this

distance was, on average, 8.17 deg for the random group [$t(49) = 7.29$, $p < .001$; 5 participants had no false alarms and were excluded from this analysis]. Clearly, memory for target–scene contingencies improved explicit recall of target position. It is interesting to note, however, that in the predictable condition, participants' memory for target position was biased toward the pillows' center of mass. Using the pillows' center of mass as a reference point instead of actual target positions in old scenes, the average error in target localization was 2.49 deg. This value is 37% less than that obtained when measuring the distance between target placements and the actual location of the targets [$t(27) = 8.54$, $p < .001$] and is equivalent to the distance between target placements and the pillows' center of mass observed on false-alarm trials [$t(27) < 1$]. This result suggests that participants in the predictable group did not remember exact locations of the targets, but instead extrapolated a *general covariation rule* such as "the target is on the pillow" to guide their positioning of targets. We discuss this finding in more depth in the discussion, below.

Summary of findings. The results of Analysis 2 indicate that in addition to search efficiency, learned regularities are revealed by measures of explicit recall.[4] Despite equivalent scene recognition, participants in the predictable condition were able to indicate target positions in previously viewed scenes more accurately than were participants in the random condition. Target placement was systematically biased, however, toward the pillows' center of mass, indicating imprecise spatial memory for target position.

Analysis 3: Transfer of Learning to Novel Search Environments

Unlike the learning trials described above, the transfer trials were not the same for all participants. We therefore considered search performance separately within each of the three transfer conditions. Before reporting these analyses, we note that we did not analyze global gaze behavior prior to fixation on the target, because some participants (primarily those in the random condition) did not fixate the target on one or more of the transfer trials. We therefore focused on the distance analysis reported above, which does not require that the target be ultimately fixated. To avoid empty cells in the data matrix, we restricted our analysis to the first three fixations, because participants in the predictable condition often fixated the target by the third fixation (in contrast, none of the participants in the random condition fixated the target by even the fifth fixation). These results are illustrated in Figure 4.

Transfer Test 1: Bedrooms without pillows. We initially submitted the data to a 4 (trial) × 3 (fixation number) × 2 (search condition) mixed model ANOVA to determine whether search behavior varied across the four transfer trials. No main effects or interactions involving the trial factor were reliable (all ps > .71), indicating that search behavior on all four transfer trials was equivalent. Therefore, we conducted an ANOVA that collapsed across transfer trials. In this analysis, main effects of search con-
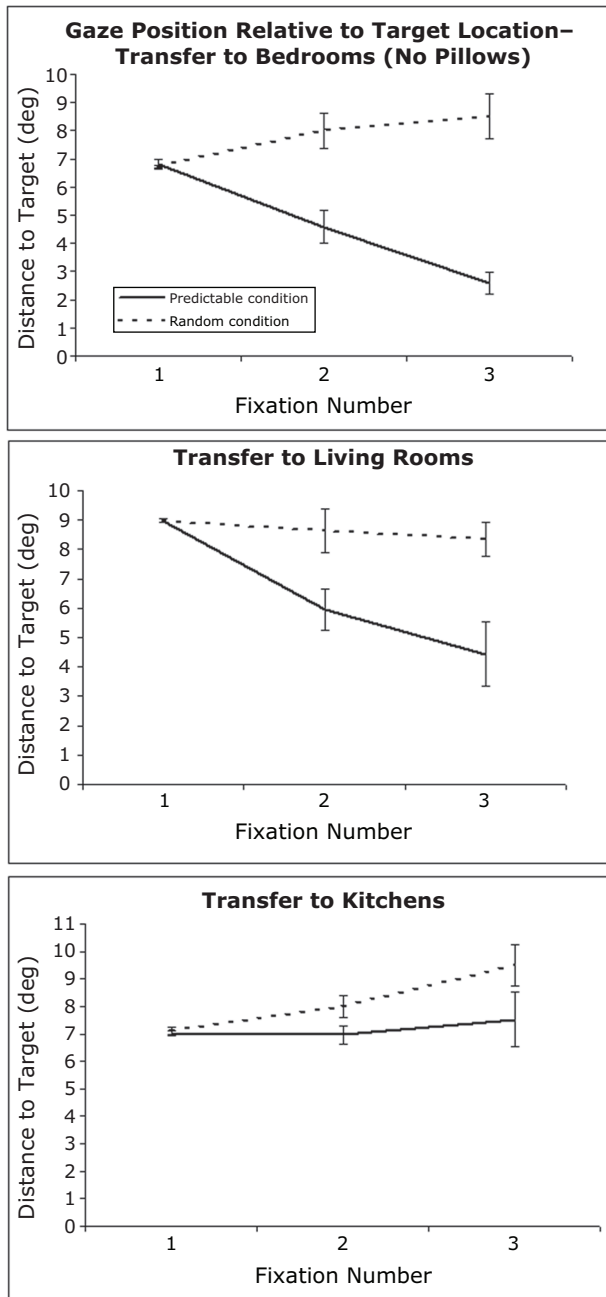
Figure 4. The linear distance between the point of fixation and the target on the transfer trials as a function of transfer condition, search condition, and fixation number (the first three fixations on the scene). Error bars depict standard errors of the means.

dition [$F(1,18) = 33.8, p < .001$] and fixation [$F(2,36) = 4.35, p < .001$] were observed. Critically, these factors interacted. [$F(2,36) = 26.6, p < .001$]. To explore the nature of this interaction, we separately considered the random and predictable search conditions. In the random condition, an effect of fixation number was observed [$F(2,18) = 3.91, p < .05$], but this was due to a modest tendency to shift gaze away from the target. In contrast, a clear shift of gaze *toward* the target was observed in the

predictable condition [$F(2,18) = 35.97, p < .001$], indicating some degree of transfer.

To assess whether transfer was complete, we contrasted performance on the transfer block with the last block of learning trials. A 2 (block) $\times$ 3 (fixation number) repeated measures ANOVA revealed a main effect of block because the distance between the initial fixation point and the target was not constrained to be equal across blocks [$F(1,9) = 13.7, p < .01$]. The expected effect of fixation number was also observed [$F(2,18) = 45.6, p < .001$]. Critically, these factors did not interact [$F(2,18) = 1.67, p = .22$], indicating equivalent guidance of gaze toward the target in the last block of learning trials and in the block of transfer trials. This finding suggests complete transfer of learning.

Transfer Test 2: Living rooms with pillows. Once again, a 3 (trial) $\times$ 3 (fixation number) $\times$ 2 (search condition) mixed model ANOVA was conducted to determine whether search behavior varied across the transfer trials. A main effect of trial was observed because the distance between the initial fixation point and the target was not constrained to be equal across the transfer trials [$F(2,36) = 3.69, p < .05$]. No interactions involving the trial factor were reliable, however, indicating that search behavior on all three transfer trials was equivalent. Therefore, we conducted an ANOVA that collapsed across transfer trials. In this analysis, main effects of search condition [$F(1,18) = 11.1, p < .001$] and fixation [$F(2,36) = 10.9, p < .001$] were observed. Critically, these factors interacted. [$F(2,36) = 6.33, p < .01$]. To explore the nature of this interaction, we considered the random and predictable search conditions separately. In the random condition, no effect of fixation number was observed [$F(2,18) < 1$], since the eyes showed no systematic tendency to move closer to the target. In contrast, a clear shift of gaze toward the target was observed in the predictable condition [$F(2,18) = 14.3, p < .001$], again indicating some degree of transfer of learning. Contrasting performance on the transfer block with that on the last block of learning trials in the predictable condition, a 2 (block) $\times$ 3 (fixation number) repeated measures ANOVA showed a marginal effect of block because the distance between the initial fixation point and the target was not constrained to be equal across trials or blocks [$F(1,9) = 3.22, p = .11$]. The expected effect of fixation number was also observed [$F(2,18) = 43.15, p < .001$]. Once again, these factors did not interact [$F(2,18) < 1$], indicating equivalent guidance of gaze toward the target in the last block of bedroom trials and in the block of transfer trials. This finding suggests complete transfer of learning from bedrooms to living rooms.

Transfer Test 3: Kitchens with pillows. As in the previous analyses, a 4 (trial) $\times$ 3 (fixation number) $\times$ 2 (search condition) mixed model ANOVA was conducted to determine whether search behavior varied across the transfer trials. No main effects or interactions involving the trial factor were reliable (all $ps > .11$), indicating that search behavior on all four transfer trials was equivalent. In an analysis that collapsed across transfer trials, a marginal effect of search condition was observed [$F(1,18) =$

3.58, $p = .075$]. The main effect of fixation was reliable [$F(2,36) = 4.63$, $p < .05$]. In contrast to prior analyses, however, these factors did not interact [$F(2,36) = 1.87$, $p = .17$]. In both the random and predictable conditions, gaze tended to shift away from the target, and the magnitude of these trends was the same across conditions. Additionally, for those participants in the predictable condition, gaze behavior on the transfer trials was indistinguishable from that on the first block of learning trials (all $ps > .54$). Thus, no evidence for any degree of transfer was obtained.

Summary of findings. Although targets were always located on a pillow during learning, the presence of pillows in the scene was not critical to search. When targets were positioned in a location in which one would reasonably expect to find a pillow (e.g., the bed), the benefits of learning were completely preserved. The effects of learning were also completely maintained when the scene category shifted from bedrooms to living rooms in which targets were located on a sofa pillow. When the scene category shifted to kitchens, however, the effects of learning were entirely lost, despite the fact that targets continued to be located on a pillow.

DISCUSSION

There is little dispute that visual behaviors including gaze control are strongly influenced by a variety of top-down factors (see, e.g., Buswell, 1935; Henderson, Brockmole, Castelhano, & Mack, 2007; Torralba, Oliva, Castelhano, & Henderson, 2006; Yarbus, 1967). Preexisting representations of scene structure (Castelhano & Henderson, 2007), likely constituent objects (De Graef, Christiaens, & d'Ydewalle, 1990; Friedman & Liebelt, 1981; Henderson et al., 1999), and likely locations of objects (Eckstein, Drescher, & Shimozaki, 2006; Henderson et al., 1999; Neider & Zelinsky, 2006; Torralba et al., 2006; Võ & Henderson, 2009) all constrain search and gaze behaviors. Little research, however, has explored the manner in which these semantic expectations are born. Our goal was to take a first step toward describing and evaluating the learning mechanisms that give rise to semantically based viewing behaviors.

To do this, we exploited contextual cuing, a type of visual statistical learning in which repeated exposure to particular context–target associations leads to progressively more efficient visual search. Typically, these learning effects are studied in situations in which specific visual displays are repeatedly shown to observers; however, the visual world is defined by contingencies that exist across, as well as within, scenes. In this article, we considered observers' sensitivity to such covariation and the nature of the statistical representations that are abstracted away from individual scenes. Because we did not provide prior instruction to our participants regarding the contingencies present in the scenes, we were able to investigate how these contingencies arose incidentally during scene viewing rather than through explicit instruction.

We first aimed to demonstrate that contextual cuing can result when target–scene regularities exist across scenes

with common identity but no perceptual overlap (Analysis 1 above). Participants searched for letter targets within 32 different bedroom scenes. For half of the participants these targets were randomly positioned within the scenes. For the other half of the participants, the targets were located on the bed pillow. Participants in this condition were sensitive to this contingency, and search became more efficient as the experiment progressed. This improvement in search demonstrates that observers are able to learn the regularities that exist across their visual environments, even when each environment is encountered only once.

When asked to identify the locations of targets after the search task was completed (Analysis 2 above), participants in the predictable search conditions placed targets on a bed pillow but with systematic imprecision. When measured against the center of the pillows, target placement was more accurate than when measured against the actual position of the target, which could appear anywhere on the pillow. There are two important implications for these findings. First, we can conclude that learning was explicit. This contrasts with the more usual finding that VSL involving multiple different (nonscene) displays is implicit (Chun & Jiang, 1999; Goujon, Didierjean, & Marmèche, 2007, 2009) but is consistent with superior explicit scene memory (Shepard, 1967; Standing, 1973; Standing, Conezio, & Haber, 1970) and explicit cuing typically observed during search through real-world scenes (Brockmole & Henderson, 2006b). As a result, it appears that metacognitive awareness of statistical regularities present in a display is more strongly linked with the stimulus materials used than with the nature of the learning itself. Second, explicit knowledge of statistical relationships in scenes leads to biased recall of target positions. This is reminiscent of prototype effects in categorical and spatial learning, in which the central tendency of a category introduces biases in the recall of specific exemplars (see, e.g., Huttenlocher, Hedges, & Duncan, 1991; Medin, Altom, & Murphy, 1984; Posner & Keele, 1968). In the present case, the abstracted contingencies enabled reasonably accurate recall, but this was likely driven by higher order spatial knowledge for the link between pillows and targets rather than discrete memory for precise target positions.

Our second aim in this study was to determine the nature of abstracted learning by considering the degree to which the derived relational rules used to guide search can be applied to different search contexts (Analysis 3 above). We first ruled out a perceptual learning effect based on low-level image properties by showing that when pillows were suddenly removed from bedrooms, participants in the predictable conditions continued to shift their gaze without hesitation to locations where pillows were most likely to appear—namely, the bed. This provides strong evidence that search was guided by semantic contingencies linking bedrooms and pillows rather than by a perceptual rule that relied on the visible presence of a pillow in the scene.

The use of these semantic regularities was not bound strictly to the particular category of scene in which they were learned, however. When the search context changed

from bedrooms to living rooms, participants who had found targets on bed pillows immediately shifted their attention toward the pillows on the sofa. This finding suggests that observers extrapolated a general rule that targets appear on a pillow, rather than a specific rule that they appear only on pillows in bedrooms, or even a moderately less constrained rule that they appear on pillows on beds. We can further exclude the possibility that the transfer we observed to the pillowless bedrooms was driven by a search for beds, because this strategy would have resulted in incomplete transfer to living rooms in which beds were absent. Basing knowledge of semantic regularities on local contingencies in this way provides flexibility to these representations, because they can be applied across multiple scenes with very different global properties.

The ability to employ local contingency rules was not completely independent of scene category information, however. When the search context changed from a bedroom to a kitchen, participants showed no propensity to shift their gaze toward the pillows that were present in the scene. In combination with the search behaviors we observed in the living rooms, this result suggests that guidance arises from an interaction between global and local contingencies: Scene identity and category information place limits on the application of local contingencies. We hypothesize that just as scene categories can constrain identification of canonical and noncanonical objects (cf. Davenport, 2007; Davenport & Potter, 2004), global scene-identity information can limit the access of local contingency information in memory. In the context of the present study, when global scene identity did not give rise to reasonable expectations concerning a pillow's presence or location, the localist "the target is on a pillow" rule was not invoked. Whereas basing abstracted contingency knowledge on local scene aspects allows learning to be applied to multiple scene categories, basing the access of such knowledge on global identity information limits its application to situations in which such knowledge is likely to be useful.

In conclusion, we have provided evidence for a novel kind of contextual cuing in which statistical regularities evident across scene exemplars are learned incidentally during scene viewing. The local contingencies underlying this learning can be flexibly applied to a variety of different scene categories insofar as they are consistent with global gist. Hence, we have shown how learning across scene exemplars and categories is guided by semantic expectations. Future research is now required to more closely examine these processes.

REFERENCES

Baldwin, D., Andersson, A., Saffran, J., & Meyer, M. (2008). Segmenting dynamic human action via statistical structure. *Cognition*, **106**, 1382-1407.

Brady, T. F., & Oliva, A. (2008). Statistical learning using real-world scenes: Extracting categorical regularities without conscious intent. *Psychological Science*, **19**, 678-685.

Brockmole, J. R., Castelhano, M. S., & Henderson, J. M. (2006). Contextual cuing in naturalistic scenes: Global and local contexts. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **32**, 699-706.

Brockmole, J. R., Hambrick, D. Z., Windisch, D. J., & Henderson, J. M. (2008). The role of meaning in contextual cuing: Evidence from chess expertise. *Quarterly Journal of Experimental Psychology*, **61**, 1886-1896.

Brockmole, J. R., & Henderson, J. M. (2006a). Recognition and attention guidance during contextual cuing in real-world scenes: Evidence from eye movements. *Quarterly Journal of Experimental Psychology*, **59**, 1177-1187.

Brockmole, J. R., & Henderson, J. M. (2006b). Using real-world scenes as contextual cues for search. *Visual Cognition*, **13**, 99-108.

Buswell, G. T. (1935). *How people look at pictures: A study of the psychology of perception in art.* Chicago: University of Chicago Press.

Castelhano, M. S., & Henderson, J. M. (2007). Initial scene representations facilitate eye movement guidance in visual search. *Journal of Experimental Psychology: Human Perception & Performance*, **33**, 753-763.

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., & Turk-Browne, N. B. (2008). Associative learning mechanisms in vision. In S. J. Luck & A. R. Hollingworth (Eds.), *Visual memory* (pp. 209-245). Oxford: Oxford University Press.

Crist, R. E., Kapadia, M. K., Westheimer, G., & Gilbert, C. D. (1997). Perceptual learning of spatial localization: Specificity for orientation, position, and context. *Journal of Neurophysiology*, **78**, 2889-2894.

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

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

De Graef, P., Christiaens, D., & d'Ydewalle, G. (1990). Perceptual effects of scene context on object identification. *Psychological Research*, **52**, 317-329.

Eckstein, M. P., Drescher, B. A., & Shimozaki, S. S. (2006). Attentional cues in real scenes, saccadic targeting, and Bayesian priors. *Psychological Science*, **17**, 973-980.

Ehinger, K. A., & Brockmole, J. R. (2008). The role of color in visual search in real-world scenes: Evidence from contextual cuing. *Perception & Psychophysics*, **70**, 1366-1378.

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

Fiser, J., & Aslin, R. N. (2002). Statistical learning of higher-order temporal structure from visual shape sequences. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **28**, 458-467.

Fiser, J., & Aslin, R. N. (2005). Encoding multielement scenes: Statistical learning of visual feature hierarchies. *Journal of Experimental Psychology: General*, **134**, 521-537.

Friedman, A., & Liebelt, L. S. (1981). On the time course of viewing pictures with a view towards remembering. In D. F. Fisher, R. A. Monty, & J. W. Senders (Eds.), *Eye movements: Cognition and visual perception* (pp. 137-155). Hillsdale, NJ: Erlbaum

Geng, J. J., & Behrmann, M. (2002). Probability cuing of target location facilitates visual search implicitly in normal participants and patients with hemispatial neglect. *Psychological Science*, **13**, 520-525.

Geng, J. J., & Behrmann, M. (2005). Spatial probability as an attentional cue in visual search. *Perception & Psychophysics*, **67**, 1252-1268.

Goujon, A., Didierjean, A., & Marmèche, E. (2007). Contextual cuing based on specific and categorical properties of the environment. *Visual Cognition*, **15**, 257-275.

Goujon, A., Didierjean, A., & Marmèche, E. (2009). Semantic contextual cuing and visual attention. *Journal of Experimental Psychology: Human Perception & Performance*, **35**, 50-71.

Henderson, J. M., Brockmole, J. R., Castelhano, M. S., & Mack, M. (2007). Visual saliency does not account for eye movements during visual search in real-world scenes. In R. P. G. van Gompel, M. H. Fischer, W. S. Murray, & R. L. Hill (Eds.), *Eye movements: A window on mind and brain* (pp. 537-562). Amsterdam: Elsevier.

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

Huttenlocher, J., Hedges, L. V., & Duncan, S. (1991). Categories and particulars: Prototype effects in estimating spatial location. *Psychological Review*, **98**, 352-376.

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

Jiang, Y., & Song, J.-H. (2005). Hyperspecificity in visual implicit learning: Learning of spatial layout is contingent on item identity. *Journal of Experimental Psychology: Human Perception & Performance*, **31**, 1439-1448.

Kirkham, N. Z., Slemmer, J. A., & Johnson, S. P. (2002). Visual statistical learning in infancy: Evidence for a domain general learning mechanism. *Cognition*, **83**, B35-B42.

Kunar, M. A., Flusberg, S., Horowitz, T. S., & Wolfe, J. M. (2007). Does contextual cuing guide the deployment of attention? *Journal of Experimental Psychology: Human Perception & Performance*, **33**, 816-828.

Medin, D. L., Altom, M. W., & Murphy, T. D. (1984). Given versus induced category representations: Use of prototype and exemplar information in classification. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **10**, 333-352.

Neider, M. B., & Zelinsky, G. J. (2006). Scene context guides eye movements during visual search. *Vision Research*, **46**, 614-621.

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.

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

Posner, M. I., & Keele, S. W. (1968). On the genesis of abstract ideas. *Journal of Experimental Psychology*, **77**, 353-363.

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

Smyth, A. C., & Shanks, D. R. (2008). Awareness in contextual cuing with extended and concurrent explicit tests. *Memory & Cognition*, **36**, 403-415.

Standing, L. (1973). Learning 10,000 pictures. *Quarterly Journal of Experimental Psychology*, **25**, 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**, 73-74.

Torralba, A., Oliva, A., Castelhano, M. S., & Henderson, J. M. (2006). Contextual guidance of eye movements and attention in real-world scenes: The role of global features in object search. *Psychological Review*, **113**, 766-786.

Turk-Browne, N. B., Isola, P. J., Scholl, B. J., & Treat, T. A. (2008). Multidimensional visual statistical learning. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **34**, 399-407.

Turk-Browne, N. B., Jungé, J. A., & Scholl, B. J. (2005). The automaticity of visual statistical learning. *Journal of Experimental Psychology: General*, **134**, 552-564.

Turk-Browne, N. B., & Scholl, B. J. (2009). Flexible visual statistical learning: Transfer across space and time. *Journal of Experimental Psychology: Human Perception & Performance*, **35**, 195-202.

Turk-Browne, N. B., Scholl, B. J., Chun, M. M., & Johnson, M. K. (2009). Neural evidence of statistical learning: Efficient detection of visual regularities without awareness. *Journal of Cognitive Neuroscience*, **21**, 1934-1945.

Võ, M. L.-H., & Henderson, J. M. (2009). Does gravity matter? Effects of semantic and syntactic inconsistencies on the allocation of attention during scene perception. *Journal of Vision*, **9**(3, Art. 24), 1-15.

Yarbus, A. L. (1967). *Eye movements and vision.* New York: Plenum.

NOTES

1. The minor disparity in stimulus numbers for each scene type arose because the various conditions for which they were developed were completed at different times and for initially different purposes.

2. Uncharacteristically long eye movements have a disproportionately larger impact on scan-path ratio than on search time and fixation count, and scan-path ratios are more heavily influenced by outliers as a result. The analysis of scan-path ratio therefore additionally excluded trials with ratios greater than 2.5 *SD*s above the mean in the random condition. This resulted in the exclusion of an additional 10 trials (1%) from the predictable condition (mean scan-path ratio of excluded outliers = 35.3) and 22 trials (3%) from the random condition (mean scan-path ratio of excluded outliers = 43.6).

3. We determined this value by replicating the recall task 10 times with 60 "virtual observers" in each replication. The "responses" generated by these observers were randomly selected (x, y) coordinates. Across these replications, the average distance between responses and target positions ranged from 9.40–10.72 deg (SD = 0.38).

4. We also note that the same conclusion regarding explicit awareness was supported by participants' verbal reports. After the experiment, we asked participants in the predictable condition if they "noticed anything about the placement of targets in the scenes." Twenty-eight of the 30 participants spontaneously reported that the targets were on pillows. The remaining 2 participants spontaneously reported that the targets were on beds.