Priming of fixations during recognition of natural scenes

Christian Valuch
Faculty of Psychology, University of Vienna, Austria

Stefanie I. Becker
School of Psychology,
The University of Queensland, Australia

Ulrich Ansorge
Faculty of Psychology, University of Vienna, Austria
School of Psychology,
The University of Queensland, Australia
Institute of Cognitive Science,
University of Osnabrück, Germany

Eye fixations allow the human viewer to perceive scene content with high acuity. If fixations drive visual memory for scenes, a viewer might repeat his/her previous fixation pattern during recognition of a familiar scene. However, visual salience alone could account for similarities between two successive fixation patterns by attracting the eyes in a stimulus-driven, task-independent manner. In the present study, we tested whether the viewer's aim to recognize a scene fosters fixations on scene content that repeats from learning to recognition as compared to the influence of visual salience alone. In Experiment 1 we compared the gaze behavior in a recognition task to that in a free-viewing task. By showing the same stimuli in both tasks, the task-independent influence of salience was held constant. We found that during a recognition task, but not during (repeated) free viewing, viewers showed a pronounced preference for previously fixated scene content. In Experiment 2 we tested whether participants remembered visual input that they fixated during learning better than salient but nonfixated visual input. To that end we presented participants with smaller cutouts from learned and new scenes. We found that cutouts featuring scene content fixated during encoding were recognized better and faster than cutouts featuring nonfixated but highly salient scene content from learned scenes. Both experiments supported the hypothesis that fixations during encoding and maybe during recognition serve visual memory over and above a stimulus-driven influence of visual salience.

Introduction

Our visual world is rich in detail. However, at each instance in time, humans select only a fraction of the surrounding visual information for purposes such as perception, action control, or memory storage and later recognition. This selection is called selective visual attention. The principles of selective visual attention are not yet fully understood, but understanding attention is key for many areas of cognitive research.

One way attention is studied is by tracking the human eyes (Kowler, 2011). Much has been learned about attention by studying human fixations during image viewing (Schütz, Braun, & Gegenfurtner, 2011). Fixations are the phases where human gaze is relatively stable, resting on a particular detail of the image. Humans tend to conduct two to three fixations per second. Fixations allow humans to resolve the visual information at the center of gaze with a higher spatial resolution because there is less convergence of retinal input at the fovea—the small area at the center of the retina that captures roughly 2° of viewing angle (cf. Henderson, 2003). There is evidence that most visual information is extracted during fixations and little visual information is taken up during saccades—the jumping eye movements that rapidly relocate gaze and result in another fixation (Bridgeman, Hendry, & Stark, 1975; McConkie & Currie, 1996).

Accordingly, there is a large body of studies that used photographs of natural scenes and recorded the participants' gaze behavior to understand where humans direct their attention (Henderson, 2003; Itti & Koch, 2000, 2001; Wilming, Betz, Kietzmann, &

Citation: Valuch, C., Becker, S. I., & Ansorge, U. (2013). Priming of fixations during recognition of natural scenes. *Journal of Vision, 13*(3):3, 1–22, http://www.journalofvision.org/content/13/3/3, doi:10.1167/13.3.3.

doi: 10.1167/13.3.3 Received December 3, 2012; published February 25, 2013 ISSN 1534-7362 © 2013 ARVO

König, 2011). These eye-tracking studies have demonstrated that principles like visual salience (Itti, Koch, & Niebur, 1998), top-down control of vision (Buswell, 1935; Yarbus, 1967), especially the top-down settings during target search (Einhäuser, Rutishauser, & Koch, 2008; Torralba, Oliva, Castelhano, & Henderson, 2006), and a preference for novel as compared to repeated information (Brockmole & Henderson, 2008; Itti & Baldi, 2009) all determine where humans fixate when viewing natural scene photographs.

To start with, consider visual salience. Salience denotes a metric of the strength of the representation of a visual image (Goferman, Zelnik-Manor, & Tal, 2010; Itti, Koch, & Niebur, 1998; Valenti, Sebe, & Gevers, 2009). The most influential model calculates salience as a topographic map of local visual feature contrasts within an image, separately for color, intensity, and orientation (Itti et al., 1998). These contrasts are summed at each location and topographically represented in one joint salience map. The higher the local salience value, the more likely it is that a fixation is directed to this position (Itti & Koch, 2000; Parkhurst, Law, & Niebur, 2002). In principle, visual salience could account for commonalities in fixated image areas across repetitions of images and between different viewers with above-chance accuracy (see Itti & Koch, 2000, 2001). Indeed, past research reported above-chance correlations between visual salience and human fixations (see Itti & Koch, 2000, 2001; Parkhurst et al., 2002).

Yet, factors beyond salience are responsible for a preference for repeated fixations on objects or locations (cf. Einhäuser, Spain, & Perona, 2008; Nuthmann & Henderson, 2010). Specifically, a tendency of a particular viewer to repeatedly look at previously observed positions or objects during a second examination of a familiar image could be driven by the importance of fixations for the representation of scenes in memory. For example, according to *visual memory theory*, scene recognition draws on local details, such as specific objects in a visual scene, which are picked up during fixations and encoded into memory (Hollingworth & Henderson, 2002).

In agreement with a supportive role of fixations for visual memory, the recognition of a local object or the detection of a local change in a scene can benefit from prior fixation on this particular detail (Hollingworth & Henderson, 2002; Hollingworth, Williams, & Henderson, 2001; Irwin & Zelinsky, 2002; Melcher & Kowler, 2001; Pertzov, Zohary, & Avidan, 2009). Also, in line with a role of fixations during encoding into memory, participants fixate more and longer on single objects during a scene-memory task than during the task of visually searching through a scene (Castelhano & Henderson, 2009), and participants are surprisingly accurate when asked to recognize local details from previously viewed scenes (Brady, Konkle, Alvarez, & Oliva, 2008; Castelhano, Mack, & Henderson, 2009; Konkle, Brady, Alvarez, & Oliva, 2010). All these findings suggest a supportive role of fixations on specific objects or locations for encoding and later memorization of natural scenes from photographs.

One would therefore expect that the same fixations on specific objects or locations that the viewer made during image or scene learning or encoding would also be repeated by this viewer during image or scene recognition (cf. Noton & Stark, 1971; Stark & Ellis, 1981). However, at variance with this hypothesis, scene recognition can be accomplished with even a single fixation, too (cf. Helmholtz, 1867; see also e.g., Intraub, 1981; Potter, 1976; Schyns & Oliva, 1994; Thorpe, Fize, & Marlot, 1996). Presenting scenes for only a few tens of milliseconds is enough that participants recognize important characteristics of scenes (e.g., Bacon-Macé, Kirchner, Fabre-Thorpe, & Thorpe, 2007), although this time is too short to allow for more than one initial fixation (for reviews, see Intraub, 2012; VanRullen, 2007). The reason for human single-glance scene-recognition ability is that scene-specific characteristics are often contained in the low-spatial frequency band of a scene's image, so that participants can extract much of the gist of a scene from an image's periphery within a single fixation within the scene image (Oliva & Torralba, 2006). In addition, the supportive role of fixations for memory is doubtful in light of findings that participants fail to notice scene changes (e.g., made across single saccades) even after they have spent substantial time on the inspection of the scenes and have made many fixations within the scene (Ballard, Hayhoe, Pook, & Rao, 1997; Droll, Hayhoe, Triesch, & Sullivan, 2005; Friedman, 1979; Irwin & Zelinsky, 2002; McConkie & Currie, 1996). Also, research on the phenomenon of change blindness suggests that detailed memory representations of scenes, objects, or features perceived from the visual environment are incomplete, short-lived, and prone to dynamic overwriting (Rensink, 2002; Simons & Levin, 1997; Simons & Rensink, 2005).

Against the background of these inconsistencies, we wanted to test two predictions that follow from the hypothesized role of fixations for memory. First, we wanted to know whether a recognition task indeed increases the number of fixations on particular details (e.g., objects or locations) that repeat from learning to recognition. If it is true that fixations on details can be helpful for memory, then we expected that participants would show a tendency to fixate the same details first during learning (or encoding) and subsequently again during recognition of the same scenes.

This prediction seems to be borne out by findings supporting *scanpath theory* (Foulsham & Underwood, 2008; Noton & Stark, 1971). Scanpath theory claims that viewers use the same image-specific sequence of fixations during recognition that they have used during initial encoding or learning of the image. According to Stark and Ellis (1981) the reason for this is that an image-specific set of local visual features is encoded along with the oculomotor commands to conjointly constitute the scanpath (for an alternative model, which explains scanpaths without involving motor memory, see Didday & Arbib, 1975). However, in its strictest form, scanpath theory is probably too rigid because it requires that a viewer is able to repeat her/his exact sequence of fixations on a specific image during recognition (Foulsham & Underwood, 2008). During many everyday instances of scene recognition, an exact repetition of the scanpath would be of little use because the angle of view on a scene changes between encoding and recognition (e.g., Sanocki, 2003; Sanocki & Epstein, 1997). The relative distances between the objects present in the scene undergo a transformation between the different viewpoints, which would make it impossible to repeat a previous scanpath. Some features and objects that were present during encoding even disappear from sight, either at the fringe of the image or because they are occluded by other objects within a scene. Thus, if recognition would critically depend on an exact repetition of scanpaths, it would not be possible after changes of the perspective. Probably even more important: Because salience alone can account for the hypothesized repetition effect on fixations (Didday & Arbib, 1975; Kaspar & König, 2011), it needs to be clarified whether salience can equally account for any predicted tendencies of human viewers to repeat their fixations on objects or locations during learning and recognition (Foulsham & Underwood, 2008; Underwood, Foulsham, & Humphrey, 2009). A suitable test of the assumed role of fixations for scene memory was conducted in Experiment 1 of the present study.

Secondly, we wanted to make sure that fixating on a scene's details during learning or encoding of a scene improves later recognition. This hypothesis follows from the assumed supportive role of fixations for visual memory, too, and it was tested in Experiment 2.

Experiment 1

We wanted to know whether the task of recognizing a scene increases the participants' tendency to look at previously fixated scene content that repeats across learned (or encoded) and recognized scene images (see Foulsham & Underwood, 2008). We recorded the eye movements during scene learning and during scene recognition. If fixations on particular details during scene learning are helpful for encoding of a scene into memory, we expected that the same image details and locations were fixated during scene recognition as during scene learning. Importantly, this should be also true where the view of a scene changes because scenes can be recognized across different angles of view, and even where information about the change of the viewing angle is only based on visual input (Hirose, Kennedy, & Tatler, 2010; Sanocki, 2003; Sanocki & Epstein, 1997).

In the most informative experimental conditions for the role of fixations during scene recognition, we therefore shifted the perspective of the images from learning to recognition (in the old/shifted images), so that only half of the scene images were repeated during recognition. To achieve this aim, all of our scene images were cutouts taken from larger photographs. In the old/shifted images, this cutout was shifted during recognition so that only 50% of the learned image was repeated during recognition. For example, if the cutout was shifted to the left border of the original source image in order to produce a perspective shift, the left half of the learning scene image became the right half of the corresponding recognition image, and the recognition image thus showed a novel left half featuring scene content that was not visible during learning. If it is true that the recognition of a natural scene benefits from looking at repeated image information during encoding and recognition, we expected a clear preference of our participants to look at the repeated part of the old/shifted images during recognition.

Also, we wanted to know whether this expected repetition effect on the basis of the participants' aim to recognize a scene exceeded the commonalities of the fixations that were explained by salience alone. As mentioned in the Introduction, even in a visual system without memory, stimulus-driven salience would lead to repeated fixations when the same image would be presented twice (here: during learning and recognition) (cf. Didday & Arbib, 1975; Foulsham & Underwood, 2008; Kaspar & König, 2011). The reason for this prediction is that the same degree of visual feature salience would be realized when one particular image would be shown twice, first during a learning block and later during a recognition block.

To test whether the expected fixation-repetition effects across learning and recognition block were due to salience alone, we varied the task (between participants). We used a control group with a free-viewing task. Free viewing is a sensible comparison task because salience is a significant predictor of gaze direction during free-viewing experiments (Itti & Koch, 2001; Parkhurst et al., 2002). If the participant's task in the recognition group leads to increases of her/his repeated fixations on the same areas during learning and recognition, the number of fixations at repeated scene areas in the recognition group should exceed the number of fixations at repeated scene areas in the free-viewing group. The same hypothesis holds when the old (learned) image is fully repeated during the subsequent recognition block: An individual observer's overlap of fixated scene regions (across the two presentations) should be higher in the recognition group than in the free-viewing group.

Method

Participants

Forty-eight observers (37 female) with a mean age of 24 years ($SD = 7$) were recruited from the student population of the Faculty of Psychology of the University of Vienna. They participated in exchange for partial course credit. Half of the observers participated in a "free-viewing" task and the other half participated in a "recognition" task. None of the observers participated in both groups. All participants were naive with respect to the research hypothesis. Prior to the actual experiment informed consent was obtained from all participants.

Stimuli

We used photographed outdoor real-world scenes of different categories. The complete set comprised 120 unique scenes that were carefully selected with regard to an intermediate recognition difficulty—that is, the photographs did not show any uncommon and peculiar objects, famous places, or specific individuals in the foreground, yet they contained some landmarks that should enable recognition even under conditions in which the perspective on a scene was shifted. The finally used stimuli were smaller cutout frames of originally larger images because one goal of the present experiment was to vary the degree of repeated visual information across successive views in a controlled way, while keeping the overall visual impression in all images comparable. From every source image, we therefore cropped the center frame which subtended the inner 50% of the original image's width and height. In a critical condition (i.e., the old/shifted scenes, see also Procedure and design) an alternative view on the scene was produced by shifting this inner 50% frame either to the right or the left border of the source photograph. For the creation of the old/shifted scenes, shifts to the right and to the left were equally frequent. In this manner, for a quarter of all images, we repeated only half of the old (i.e., previously presented) scene content in a successive presentation of the scene during the transfer/recognition block (see Figure 1). All photographs were resized to a resolution of 1024 × 768 pixels. Scene images were always displayed on full screen for 5 s, both during learning trials (or in the first block) and during transfer/recognition trials (or in the second block).

Apparatus

Scenes were displayed on a 19-in. color CRT monitor (Sony Multiscan G400) at a resolution of 1024 × 768 pixels and a vertical refresh rate of 100 Hz. Viewing distance was kept stable at 72 cm by chin and forehead rests, resulting in an apparent size of the full screen scenes of 28° × 21°. Eye movements were recorded using an EyeLink 1000 Desktop Mount eye tracker (SR Research Ltd., Kanata, Ontario, Canada) at a sampling rate of 1000 Hz. The eye tracker was calibrated using a 9-point calibration procedure on the observer's right eye (gaze position of the left eye was as well recorded and analyzed when samples of the right eye were missing). Prior to each trial there was a drift check that required participants to fixate on a centrally presented target circle. Recalibrations were performed if recorded fixation gaze average was outside a 1° radius of the pretrial drift check target circle. The experimental procedure was implemented in Experiment Builder (SR Research Ltd., Kanata, Ontario, Canada) and the experiment was run on a personal computer under Windows XP. Manual responses were recorded as button presses with the left or right index finger on a Microsoft Sidewinder game pad (key mappings were balanced across participants).

Procedure and design

Both tasks, recognition and free-viewing, used exactly the same stimuli and consisted of a learning block (in the recognition task) or a first block (in the free-viewing task), and a transfer/recognition block (in the recognition task) or a second block (in the free-viewing task). Task instructions differed between the two groups of participants. In the free-viewing group, observers were informed at the beginning of the first block that they were about to see 90 scenes in photographs, followed by a short break, and another 120 scenes in photographs in a second block. They were asked to simply "attentively examine" all of the photographs. In contrast, in the recognition group, observers were asked at the beginning of the learning block to attentively examine and *memorize* each of the 90 scenes for subsequent recognition in the transfer/recognition block. After completing the learning block, participants read the instructions for the transfer/recognition block and had time for a pause. At the beginning of the transfer/recognition block, the participants in the recognition group were informed that they would see a sequence of 120 photographs which could either be (a) *old/identical* (i.e., exactly the same photograph as presented in the learning block), (b) *old/mirrored* (i.e., a mirror-reversed version of a photograph presented in the learning block), (c) *old/shifted* (i.e., a familiar scene photographed from a different perspective), or (d) *new* (i.e., a photograph that had not

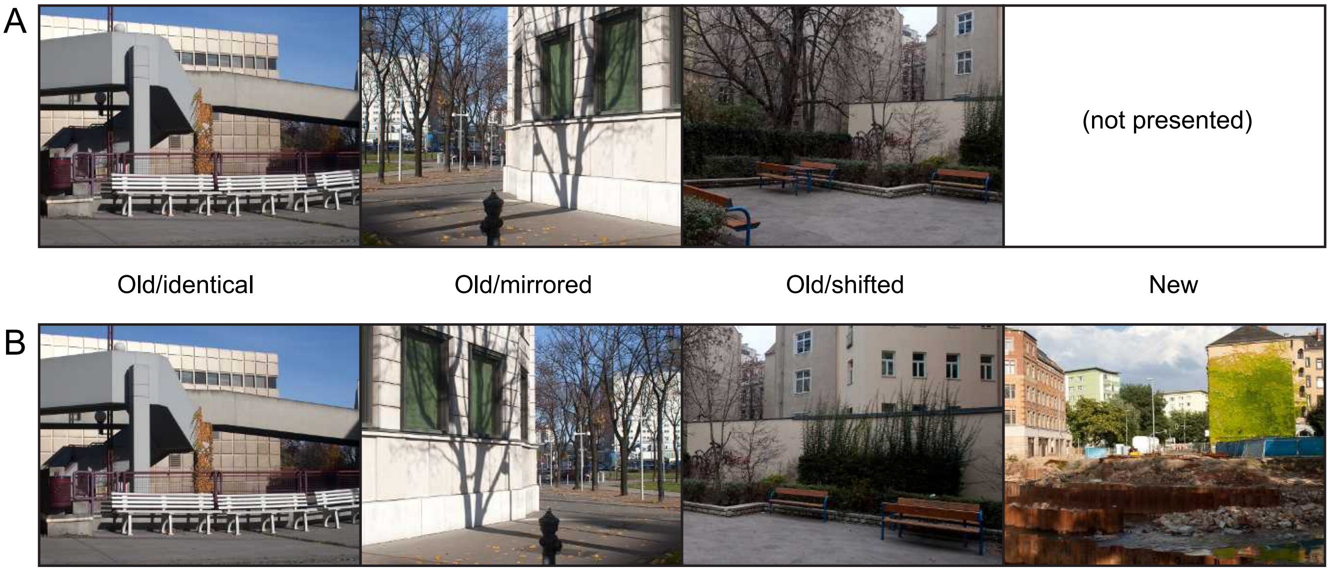

Figure 1. Example scenes from each of the experimental conditions. Each run consisted of two blocks (A, B). (A) Example scenes as presented in the learning block (or first block), which comprised 90 trials where each scene was presented once for 5 s (in randomized order). (B) Corresponding example scenes from the transfer/recognition block (or second block), which comprised 120 trials (including 30 new scenes).

even partly been shown as one of the learned scenes or the first block's scenes; see Figure 1).

Participants in the recognition group were instructed to press one of two response buttons whenever they thought that the currently presented scene is an *old/identical* or an *old/shifted* scene, and the other response button whenever they thought they had seen an *old/mirrored* scene or a *new* scene. No feedback about the correctness of the responses was given. Assignment of the different scenes to the four different conditions was counterbalanced across participants. Half of the participants saw scenes in the old/mirrored condition that the other half of the participants saw in the old/shifted condition (and vice versa). Analogously, half of the participants saw scenes in the new condition that the other half of the participants saw in the old/identical condition (and vice versa). The sequential order of the presented scenes was randomized for each participant in both, the learning block (or first block) and the transfer/recognition block (or second block). An experimental session lasted approximately 30 min, including the initial eye tracker setup and a short debriefing of the participants after they completed the transfer/recognition or second block.

Data analysis

Manual responses and eye movement data were preprocessed using Data Viewer (SR Research Ltd., Kanata, Ontario, Canada) software. For evaluating the recognition performance, we assessed hit rate (i.e., the rate of correct responses per condition) and reaction time (RT). Only the first button press after scene onset was evaluated. Using the SR Research algorithm, fixations were identified as the average gaze position during periods where the change in recorded gaze direction was smaller than 0.1°, eye movement velocity was below 30°/s, and acceleration was below $8000°/s^2$, respectively. Fixations below 100 ms and above 2000 ms were excluded from the analysis. Eventually, the first fixation in each trial was excluded from the analyses, too, because it always fell within an area of 2° around the screen center and thus reflected fixations from the pretrial drift check trailing into the scene presentation rather than scene examination. Valid data was subjected to further analyses in MATLAB (The Mathworks, Inc., Natick, MA) and SPSS (IBM, Inc., Chicago, IL). For all statistical tests, we set α at 0.05 and applied a Bonferroni-corrected α for post hoc comparisons.

Results

In the free-viewing group, we obtained 83,880 fixations altogether, whereof 4,383 (5.2%) were outside the valid duration range ($>$100 ms or $<$2 s) and thus excluded. From a total of 81,769 fixations obtained in the recognition group, 4,296 (5.3%) were excluded from the analysis by the same criteria.

First, we tested whether the task (free viewing vs. recognition) affected the number of fixations of the participants. For example, it could be that a recognition-task demanded observers to sample more information from the scene than a free-viewing task. It could further be that scene repetitions led to fewer fixations. A 2 × 2 mixed ANOVA, with the within-participant factor block (learning or first block vs. transfer/recognition or second block), and the between-participants factor task (recognition vs. free viewing) run on the mean number of fixations yielded no significant main effect and no interaction. On average, the observers made 3.0 fixations per second ($SD = 0.3$) in the free-viewing task, and 2.9 fixations per second ($SD = 0.4$) in the recognition task.

When confronted with a shifted viewpoint, do observers fixate repeated scene regions?

Of central interest for the present study were fixations on scene regions that repeated from learning to transfer/recognition across a varying viewpoint. We used the condition of old/shifted scenes to test whether a preference for old scene regions over new scene regions is contingent on the behavioral goal to recognize the scene as compared to a free-viewing situation. To that end, we evaluated the spatial distribution of fixations in the second block or transfer/recognition block: We compared the number of fixations that fell on the old half of the scene with the number of fixations that fell on the new half of the scene as a function of the viewing task (see Figure 2).

A 2 × 2 mixed ANOVA of the numbers of fixations during the transfer/recognition or second block, with the within-participant factor scene region of the old/shifted images (repeated side vs. novel side) and the between-participants factor task (free viewing vs. recognition) revealed a significant main effect of scene region, $F(1, 46) = 15.2$, $p < 0.001$, $\eta_p^2 = 0.25$. Furthermore, the analysis yielded a significant interaction between scene region and task, $F(1, 46) = 16.8$, $p < 0.001$, $\eta_p^2 = 0.27$. The main effect of task was not significant, $F(1, 46) = 0.3$, *ns*. We repeated the analysis for the overall dwell times in repeated versus novel sides of the old/shifted images and found the same pattern of results. Again, we identified a significant main effect of scene region, $F(1, 46) = 8.1$, $p < 0.01$, $\eta_p^2 = 0.15$, and a significant interaction of Scene Region × Task, $F(1, 46) = 18.8$, $p < 0.001$, $\eta_p^2 = 0.29$, whereas the main effect of task did not reach significance, $F(1, 46) = 1.95$, *ns*. Post-hoc pairwise comparisons showed that there were significantly more and longer fixations on repeated sides than on novel sides in the recognition task (mean number of fixations: old – new = 0.94, $t[23] = 6.18$, $p < 0.001$; mean overall dwell time: old – new = 249.8 ms, $t[23] = 5.49$, $p < 0.0010$). However, no corresponding effects were present in the free-viewing task (mean number of fixations: old – new = –0.24, $t[23] = –0.13$, *ns*; mean overall dwell time: old – new = –51.4, $t[23] = –0.98$, *ns*).

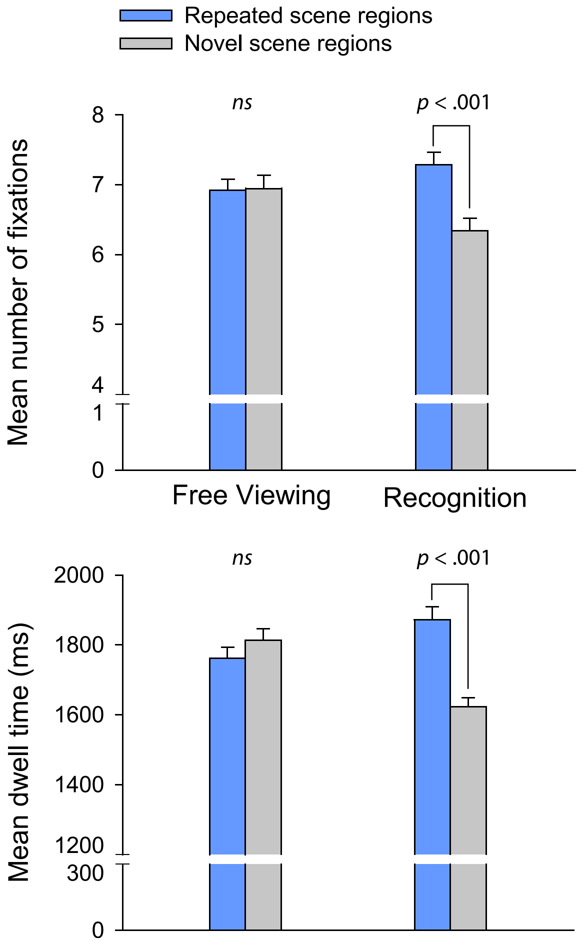

Figure 2. Mean number of fixations and overall dwell time (in milliseconds) on repeated (blue bars) versus novel (gray bars) scene regions in old/shifted scenes only during the second or transfer/recognition block, as a function of the task in Experiment 1. Error bars represent 1 *SEM*.

How does a preference for repeated versus novel scene content develop over time?

While the distribution of fixations over the whole presentation duration showed a significant preference for repeated scene content only under recognition instructions, a preference for either repeated or novel information could also vary over time (cf. Kaspar & König, 2011). Therefore, we additionally ran a time-binned analysis and counted the fixations that fell on repeated sides and novel sides of the old/shifted scene across five successive phases of scene examination (each lasting 1 s). Mean values are depicted in Figure 3 and illustrate that in the recognition group, repeated scene areas were consistently more frequently fixated than novel image areas over the entire duration of a trial. Mean numbers of fixations were submitted to a 5 × 2 ×

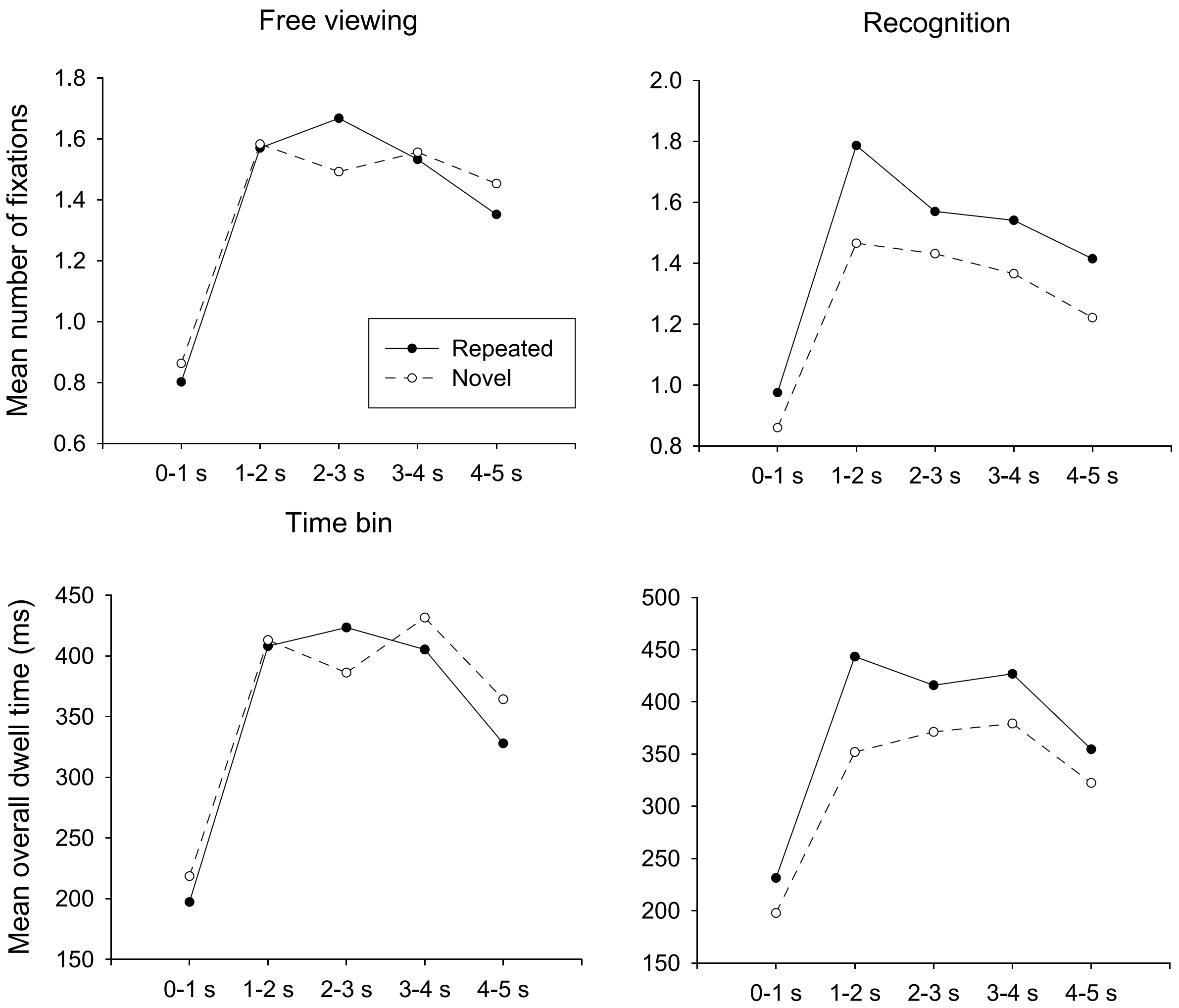

Figure 3. On the abscissae: Mean numbers of fixations (upper panels) and overall dwell times (in milliseconds; lower panels) in repeated versus novel scene regions (straight vs. broken lines) in the old/shifted condition only, as a function of task (free viewing, left panels, or recognition, right panels), and time bin (of fixation onset) on the ordinates in Experiment 1.

2 mixed ANOVA with the within-participant factors time bin (i.e., fixation onset within 0–1 s vs. 1–2 s vs. 2–3 s vs. 3–4 s vs. 4–5 s, relative to trial onset) and scene region (repeated side vs. novel side), and the between-participants factor task (free viewing vs. recognition). Mauchly's test indicated a violation of the model's sphericity assumption for the factor time bin, $\chi^2(9), = 39$, $p < 0.05$. As a consequence, we corrected the degrees of freedom for the significance test using Greenhouse-Geisser's epsilon. The analysis revealed a significant main effect of time bin, $F(2.87, 132.22) = 347.2$, $p < 0.001$, $\eta_p^2 = 0.88$, as well as a significant main effect of scene region, $F(1, 46) = 15.2$, $p < 0.001$, $\eta_p^2 = 0.25$. Significant interaction effects were present between time bin and task, $F(2.87, 132.22) = 7.2$, $p < 0.001$, $\eta_p^2 = 0.14$, as well as between scene region and task, $F(1, 46) = 16.8$, $p < 0.001$, $\eta_p^2 = 0.27$. Pairwise comparison of the number of fixations on repeated versus novel sides that we conducted separately for the two tasks and the five phases, revealed a significantly larger number of fixations on repeated sides than novel areas during all phases in the recognition task, all ps < 0.05, but only one corresponding significant difference in the free-viewing task (phase 3, from 2–3 s after onset, $p < 0.05$). All other effects were not significant. We repeated this analysis for overall dwell times in repeated versus novel scene regions and obtained qualitatively identical though weaker results.

Does the task also affect the correlation between successive fixation patterns in identical and mirrored scenes?

To assess how the task affected the correlation between the fixated image regions during learning (or first) block and transfer/recognition (or second) block we generated fixation maps for every participant. For every participant and trial, we constructed individual areas of interest (AOI) maps, by placing circular AOIs with a diameter of 2° at the fixated positions. Regions fixated during each scene examination were thus represented as two-dimensional logical matrices with a size equivalent to the pixel resolution of the stimulus. Labeling of fixated scene regions was done on a pixel-by-pixel basis: All pixels that fell within a circular area of 2° around the fixation coordinates were set to 1 (i.e., marked as "fixated"), whereas all cells that did not fall within a circular area of 2° around any of the fixation coordinates were set to 0 (i.e., marked as "unfixated"). We applied this procedure to all trials of the learning (first) block and the transfer/recognition (second) block from the old/identical, and the old/mirrored conditions, because in these two conditions 100% of the visual-scene content was repeated from the first to the second block. As a consequence, we could directly evaluate the overlap between the fixated image regions across two successive presentations (for old/mirrored scenes, we flipped the AOI maps from one of the observations along the horizontal axis). To quantify the congruence of individual AOI maps across the two successive presentations of an old/identical, or an old/mirrored scene, we calculated Pearson's phi-coefficient r_{φ}, which is a measure for the correlation of two dichotomous variables. All r_{φ} correlations were Fisher *z*-transformed (see Table 1) to allow for the calculation of mean correlations and parametric testing of differences between mean correlations.

The resulting *z* correlations were separately aggregated over participants for the old/identical, and the old/mirrored condition, respectively. Using one-sample *t* tests we first tested whether the resulting mean *z* correlations were significantly different from zero. This was the case for both tasks and in both conditions (all *t*s > 19, all *p*s < 0.001). A 2 × 2 mixed ANOVA with the within-participant factor condition (old/identical vs. old/mirrored) and the between participants factor task (free viewing vs. recognition) revealed a main effect of condition, $F(1, 46) = 25.3$, $p < 0.001$, $\eta_p^2 = 0.35$, showing that the correlation between successive presentations of a scene was higher in old/identical than in old/mirrored scenes. Importantly, the analysis also yielded a significant main effect of task, $F(1, 46) = 8.2$, $p < 0.01$, $\eta_p^2 = 0.15$, showing that correlations were higher in the recognition group than in the free-viewing group. There was no interaction between condition and task, $F(1, 46) = 0.02$, *ns*.

Recognition performance

Our analyses of recognition performance are solely based on data from the recognition group (i.e., half of the participants), because no decision about scene familiarity was collected from participants in the free viewing group. In 1.9% out of 2,880 trials, participants from the recognition group missed to give a response within the presentation duration of 5 s. The frequency of missed responses did not differ by image condition, χ^2 (3, $N = 54$) = 3.33, $p = 0.34$. These trials were therefore excluded from further analyses. Mean measures of recognition performance in valid trials are depicted in Figure 4. Mean rates of correct responses were subjected to a repeated-measures ANOVA, with the within-participant factor image condition (old/identical, old/shifted, old/mirrored, or new). Because Mauchly's test indicated a violation of the model's sphericity assumption, $\chi^2(5), = 26.7$, $p < 0.05$, we applied the Greenhouse-Geisser correction. The analysis yielded a main effect of image condition, $F(1.75, 40.33) = 17.3$, $p < 0.001$, $\eta_p^2 = 0.44$. Post-hoc pairwise comparisons showed that the rate of correct responses was significantly higher in old/identical scenes than in old/mirrored scenes ($p < 0.01$). Conversely, the rate of correct responses was significantly higher in new scenes than in all other conditions (all $p < 0.01$). No other differences between the conditions reached significance.

	Correlation of individual fixation patterns between blocks (*z*)			
	Free viewing		Recognition	
Condition	*M*	*SD*	*M*	*SD*
Old/identical	0.30	0.06	0.35	0.08
Old/mirrored	0.26	0.07	0.31	0.06

Table 1. Mean *z*-transformed correlations between the fixations of two successive scene views in Experiment 1's old/identical and old/mirrored conditions.

The same repeated-measures ANOVA computed over mean RTs yielded a significant main effect of image condition, $F(3, 69) = 16.2$, $p < 0.001$, $\eta_p^2 = 0.41$. Post-hoc pairwise comparisons showed that RTs were significantly faster in old/identical scenes than in old/shifted or old/mirrored scenes (both $p < 0.001$). Furthermore, RTs were significantly faster in new scenes than in old/shifted scenes ($p < 0.01$). No other differences between the conditions were significant. When the analysis of RTs was repeated including only trials in which observers responded correctly, the results were qualitatively identical, yielding a main effect of image condition, $F(3, 69) = 12.4$, $p < 0.001$, $\eta_p^2 = 0.35$. Correct RTs in old/identical scenes were faster than in old/shifted scenes ($p < 0.001$) or old/mirrored scenes ($p < 0.01$). Correct RTs in new scenes were faster than in old/shifted scenes ($p < 0.01$).

Because the amount of repeated visual information varied across image conditions (e.g., there was a 100% image repetition in the old/identical condition vs. 50% image repetition in the old/shifted condition), the number of fixations (or the fixation duration) necessary to solve the recognition task could differ as well. To test this, we counted the number of fixations per condition until the manual response. We assumed that any

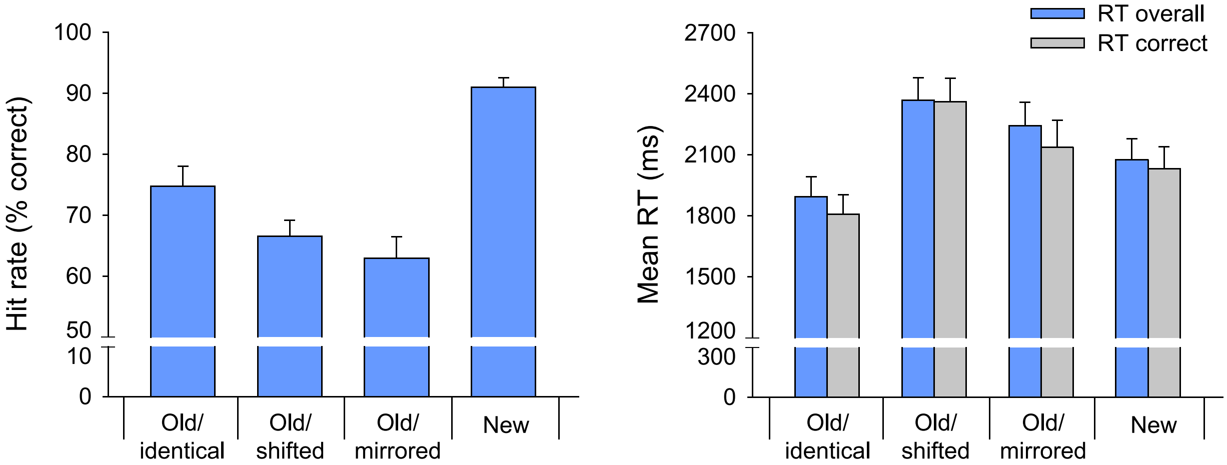

Figure 4. Recognition performance (hit rates, % correct; on the left) and reaction times (RTs; on the right) in the recognition task in Experiment 1. Error bars represent 1 *SEM*.

differences detected here reflected differences in task difficulty between the four conditions, as well as possibly differing necessities for sampling visual information from the scenes. Individual mean numbers of fixations until responses were submitted to a repeated-measures ANOVA which yielded a significant main effect of image condition, $F(3, 69) = 14.6$, $p < 0.001$, $\eta_p^2 = 0.39$. Post-hoc tests showed that the observers made significantly fewer fixations in old/identical scenes ($M = 7.4$, $SD = 1.78$) than in old/shifted ($M = 8.8$, $SD = 1.85$) or old/mirrored ($M = 8.2$, $SD = 1.88$) scenes (both $p < 0.01$). Furthermore, observers made significantly fewer fixations in new scenes ($M = 7.9$, $SD = 1.91$) than in old/shifted scenes ($p < 0.01$) before they made a button press. No other differences between the image conditions reached significance. (A similar analysis including only correct responses led to qualitatively identical results.)

Discussion

Results of Experiment 1 confirmed that the viewers' fixations were biased to locations (or objects) during recognition (or in the second block) that had also been fixated during learning/encoding (or in the first block). This bias was present in both groups of participants, but the bias was stronger in the recognition-task group compared to the free-viewing group. In the old/shifted images, this tendency was reflected in a higher number of fixations (and a higher overall fixational dwell time) on the old (repeated) image parts than on the novel image parts of the old/shifted images. This result is particularly interesting because it shows that fixations on the same objects or locations during learning and recognition could indeed support scene memory across changes of the perspective of a scene. The same results were found with the old/identical images and the old/mirrored images.[1] Because the same scene images were used in both the free viewing and the recognition group, visual salience of the stimuli was equated across tasks. Therefore, the pronounced tendency to fixate repeated scene regions must have been due to the influence of the observers' aim of recognition, which exceeded the stimulus-driven influence of salience alone. This conclusion is in line with the findings of Foulsham and Underwood (2008) and of Underwood et al. (2009) who reported that although the salience model could account for some of the variance in fixation patterns, the actual scan paths of observers were similar between learning and recognition, but this similarity was independent from the predictions of the salience model.

Also of interest would have been the recognition performance as a function of repeated fixations across learning and transfer/recognition blocks. However, the present experiment is not suited to test whether successful recognition crucially depends on fixations. One obvious shortcoming is the fact that no recognition data were available from the free-viewing group. Thus, we cannot decide whether any differences of recognition were associated with the differences of repeating fixations on particular objects or locations across the two blocks and between the two tasks. Also, it is unclear whether the differences in recognition performance between the different image conditions of the recognition group were in line with a supportive influence of repeated fixations on image memory. This is because our task required participants to correctly discriminate between old/identical, old/mirrored, old/shifted, and new images in order to select the correct of the two responses. However, it is difficult to assess whether repeating a fixation on an object or on a location in the old/mirrored images would have been helpful to discriminate this image from the alternative old images as it was required. For example, if participants have disregarded the relative location of an object during recognition, it is well possible that repeated fixations in the old/mirrored images could have lured the participants to give false alarms for old/identical images. Also, our use of old/shifted images made it likely that the participants disregarded relative locations during recognition. As a drastic example of how the use of old/shifted images might have led participants to disregard relative locations in the present experiment, consider the task of recognizing a scene from a learned frontal view of a lighthouse tower. Although it would probably aid recognition of this scene to fixate on the lighthouse tower, this tower would be looking very similar from the front and from the back—that is, it would look almost identical across very large changes of the perspective from learning to recognition. Even worse, a lighthouse viewed first from the front and later from the back would also be very similar to a lighthouse viewed first in correct cardinal view and then in a mirror-reversed image of this view. The fact that disregarding the location changes in (some of) the old/shifted images was helpful would thus be incentive to also disregard this information during the judgments about the old/mirrored images. As long as it is uncertain whether repeated fixations during learning and recognition would have been equally helpful for recognition in all image conditions, it would thus also be pointless to relate the number of repeated fixations to the quality of the recognition performance. To understand whether fixations on details during learning have their assumed supportive role for later recognition, we thus conducted Experiment 2.

Experiment 2

In Experiment 2 we used a scene-memory task. We tested whether looking at scene details during recognition that were also fixated during learning of the scene facilitated recognition of the learned scene. This prediction follows from an assumed supportive role of fixations for the encoding of the fixated details of the scenes into memory. As in Experiment 1, our participants again first viewed and encoded photographs of real-world scenes in a learning block. In a subsequent transfer/recognition block, we asked the participants to recognize the scenes that they had learned and to discriminate them as old scenes from hitherto not presented new scenes. Critically, recognition was now tested with much smaller cutouts from the originally encoded full-screen scene images. We used these cutouts rather than full-screen images of the whole scenes during recognition because with the images of the whole scenes, scene gist alone might have provided sufficient information for correct scene recognition (Oliva & Torralba, 2006), and this could have masked any supportive influence of fixated details on scene memory or scene recognition.

To understand the role of fixations for scene memory we used two types of cutouts from the old images: cutouts that showed scene content that the participants fixated during learning and control cutouts that showed salient scene content that was present in the old images but that was not fixated by the participants. We used visual salience to determine the locations from where control cutouts were taken. This had two interrelated reasons. First, as explained, salience maps predict the direction of gaze with above-chance accuracy (e.g., Foulsham & Underwood, 2008; Itti & Koch, 2000). Second and even more important in the present context, salience itself might improve scene recognition simply because of a relatively high amount of information at salient image regions (Elazary & Itti, 2008; van der Linde, Rajashekar, Bovik, & Cormack, 2009).

Method

Participants

Twenty-four observers (15 female) with a mean age of 25 years ($SD = 4$) were recruited from the student population at the Faculty of Psychology of the University of Vienna and participated voluntarily or in exchange for partial course credit. All had normal or corrected-to-normal vision. Prior to the actual experiment informed consent was obtained from all participants.

Stimuli

We used a set of 60 photographs of outdoor scenes. The scenes are shown in Figure 5. In the learning block scenes were displayed as in Experiment 1, on full screen at a resolution of 800 × 600 pixels and a vertical refresh rate of 100 Hz. In the transfer/recognition block, much smaller cutouts of the scene photographs (100 × 100 pixels, showing as 3.5° × 3.5°) were presented at screen center. Example cutouts (based on gaze data from one representative participant) are shown in Figure 6A.

Apparatus

This was the same as in Experiment 1, with the following exceptions. The experimental procedure was implemented in MATLAB with the Psychophysics Toolbox (Brainard, 1997; Pelli, 1997). Eye movements were monocularly recorded from the dominant eye, and RTs and response identities were registered via the "F"

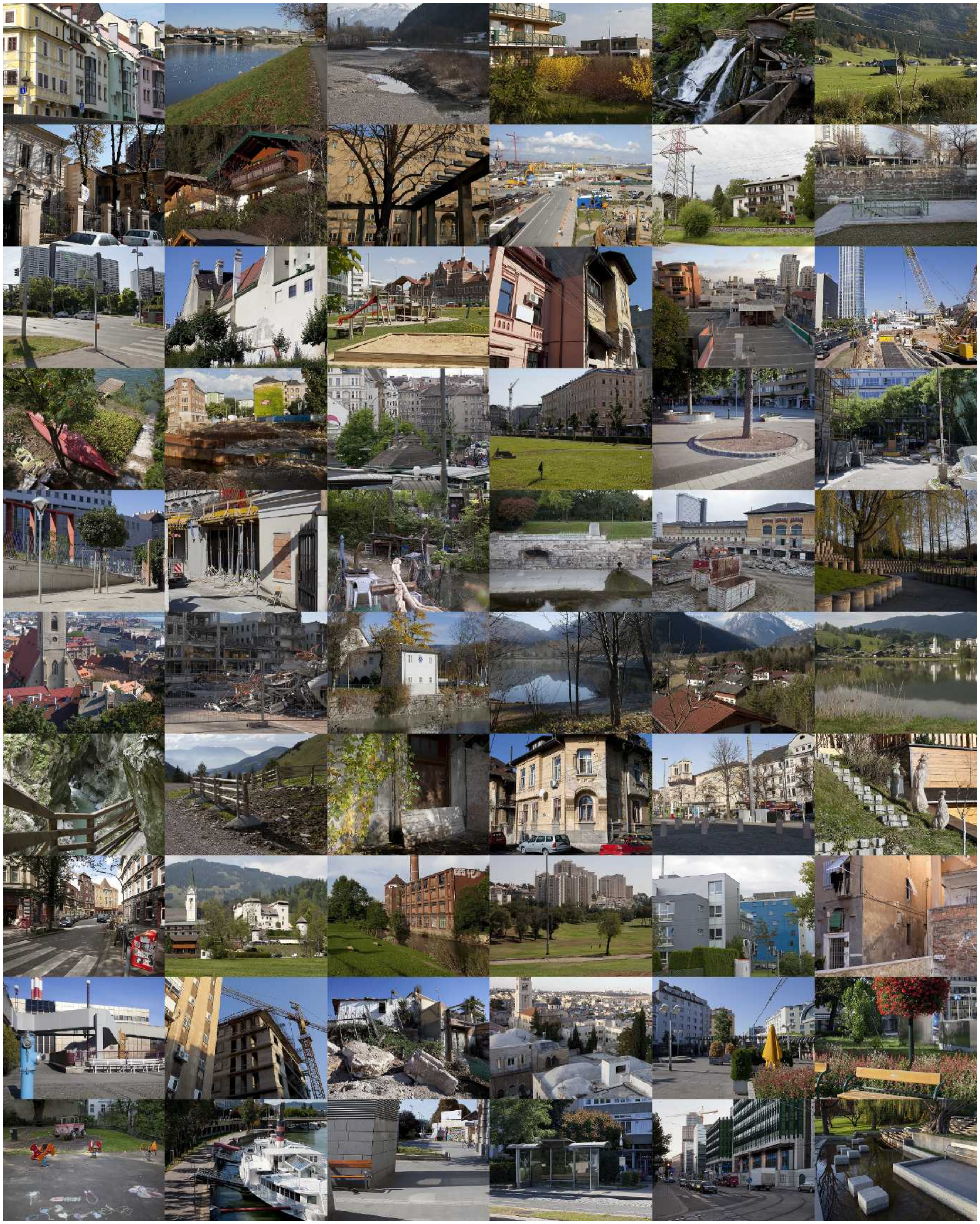

Figure 5. The complete set of scenes used in Experiment 2. Only half the scenes were learned by the observers. The second half of the scenes was used for the creation of new cutouts.

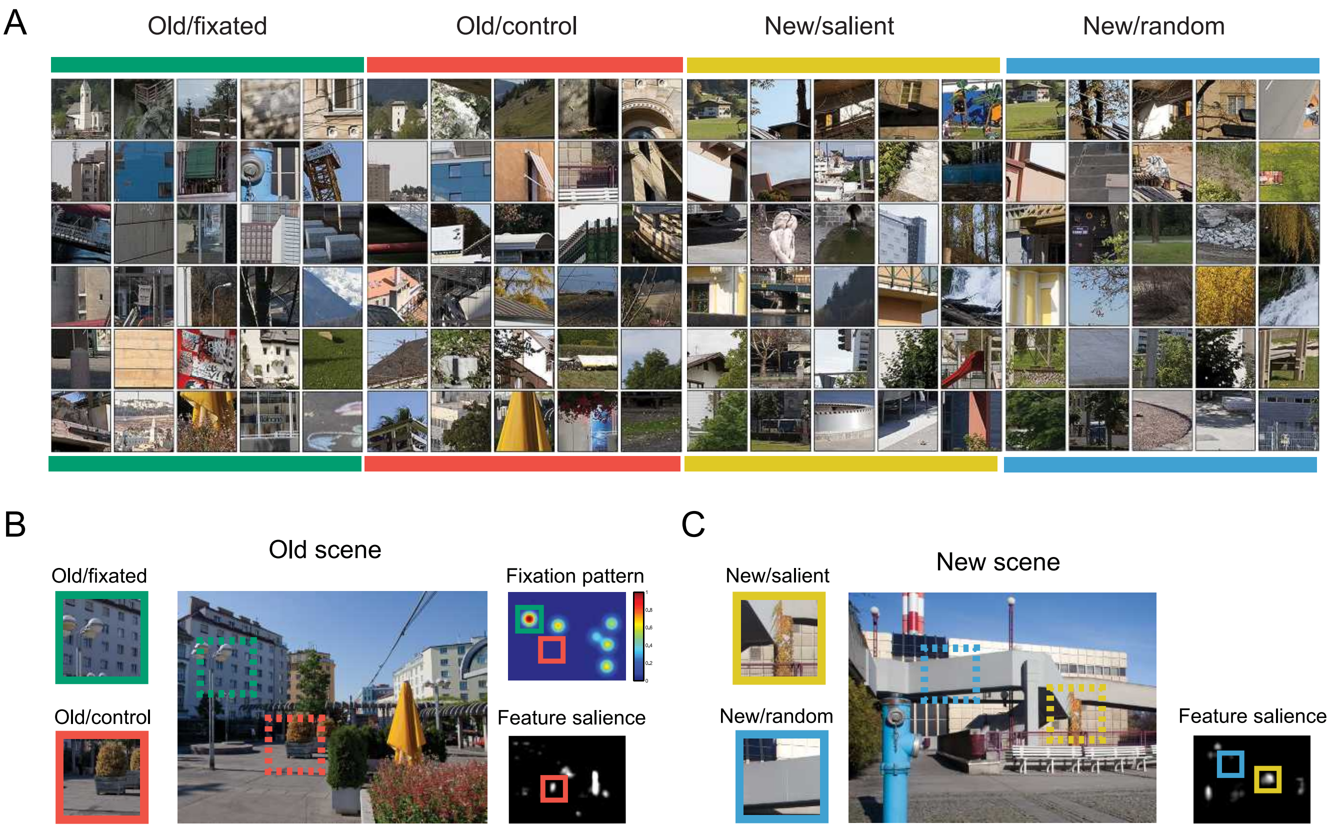

Figure 6. Stimuli for the recognition task in the transfer block of Experiment 2. After finishing the learning block, an individual stimulus set was prepared automatically for every participant. (A) Example image cutouts for the four different experimental conditions used in the recognition task of the transfer block (based on representative fixation data from one participant). (B) Cutouts from old images were selected contingent upon the participant's gaze pattern. Old/fixated cutouts showed the location of longest fixation. Old/control cutouts showed a nonfixated but highly salient location. (C) Cutouts from new images showed highly salient scene regions or were randomly chosen.

and "J" keys (pressed with the left and the right index fingers) on a standard USB keyboard. Key mappings were again balanced across participants.

Procedure and design

Participants were informed that the experiment consisted of a learning block and a transfer/recognition block and that their task in the learning block was to view and memorize scenes in photographs. They were further informed that in the later transfer/recognition block they would be confronted with smaller cutouts from photographs that either could or could not be taken from any of the learned scenes. Participants were aware of the fact that their eye movements were recorded during the learning block; however, they were naive with regard to the purpose of this recording and the experimental manipulation in the transfer/recognition block. (The postexperimental debriefing revealed that none of the participants had become aware of our central manipulation.) In the learning block, observers viewed and memorized 30 photographs from different scenes (i.e., half of the complete stimulus set). The photographs of the remaining 30 different scenes were used for the transfer/recognition block. The assignment of the images to the learning versus transfer/recognition block was counterbalanced across participants.

Prior to each scene presentation, a circular drift-check fixation target was presented at screen center. If the measured gaze position during the drift check deviated more than 1° from the fixation target's position, a nine-point recalibration of the eye tracker was performed. Each scene was presented once for five seconds (see Figure 7A). After completing the learning block, observers saw a "pause" screen, informing them that they should now have a break of 2–3 min before continuing with the transfer/recognition block of the experiment. During this time, and unknown to the observers, their fixations during the learning block were determined and an individual stimulus set for the current participant was prepared contingent on her/his fixations in every of the learned scenes. This stimulus set encompassed four conditions with 30 smaller image cutouts per condition. The cutouts had a size of 3.5° × 3.5° roughly corresponding to an area around fixation

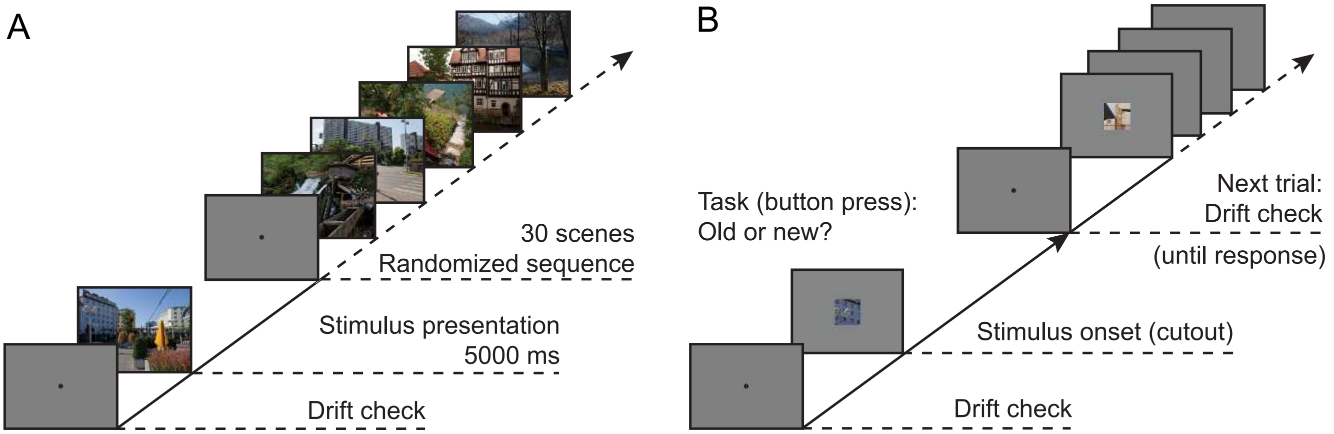

Figure 7. Procedure. (A) Scenes in the learning block were presented on full screen. (B) In the transfer block, scene cutouts in scene-corresponding size were presented at screen center.

with sensitivity for local changes during photographic scene viewing (cf. Henderson, Williams, Castelhano, & Falk, 2003). This resulted in a total of 120 image cutouts (and trials) in the recognition block. The cutouts were presented at screen center, and the depicted area within each cutout was the same as in the corresponding scene image during learning if the image had been used during learning.

Two classes of smaller cutouts were created from each of the learned scenes (see Figure 6B). An (a) *old/fixated* cutout showed the region of one participant's individual longest fixation in a particular scene image, and (b) an *old/control* cutout showed a region that was not fixated by this participant but contained salient low-level features, as determined by the Saliency Toolbox for MATLAB (Walther & Koch, 2006). For each of the images of the learning block, one old/fixated and one old/control cutout was created. In addition to this, 30 new (as yet not presented) photographs were used to prepare cutouts from hitherto unseen or unfamiliar images (see Figure 6C). For every new image, one (c) *new/salient* cutout showed a highly salient scene region, and one (d) *new/random* cutout showed a randomly selected region of the same new image. In the transfer/recognition block, all these cutouts were presented one per each of the trials at screen center and in random order (see Figure 7B). Participants were instructed to decide rapidly and accurately for every cutout whether it came from an old or a new (previously not presented) scene by pressing one of two alternative keys on the keyboard in front of them (key mappings were balanced across participants).

Data analysis

Fixation detection was based on the same criteria as in Experiment 1. Candidate locations for old/fixated cutouts were all fixated positions that were outside a circular area of 3.5° (100 pixels diameter) around the screen center, and at least 70 pixels away from the outer borders of the image (corresponding to a distance of 2.5° from the screen border). The old/fixated cutout of an individual viewer in a particular image was centered on the longest fixation that fell within this area. Also, all locations within this area of a particular image that had less than 5% overlap with any region fixated by an individual viewer in this particular image were used as potential old/control cutouts of this image and viewer. Among these cutouts, the cutout with the highest salience (determined by the Saliency Toolbox; Walther & Koch, 2006) was used as the old/control cutout for this image and viewer. (Additionally, we applied the same constraints to the position of the old/control cutout as with the old/fixated cutouts, so that it was outside the center region and not too close at the border of the image.) We again set α at 0.05 for statistical tests and applied a Bonferroni-corrected α for post-hoc comparisons.

(For the correlation between salience and fixations [see below], the first fixation per image and the fixations below 100 ms and above 2000 ms were again excluded from the analysis, just as in Experiment 1.)

Results

Out of all 2,880 trials, 1.9% were eliminated from analyses because RT differed from the individual mean RT per condition by more than 2.5 *SD*s.

Recognition performance

For the results, see Figure 8. A repeated-measures ANOVA, with the single four-step variable cutout type during recognition (old/fixated; old/control; new/sa-

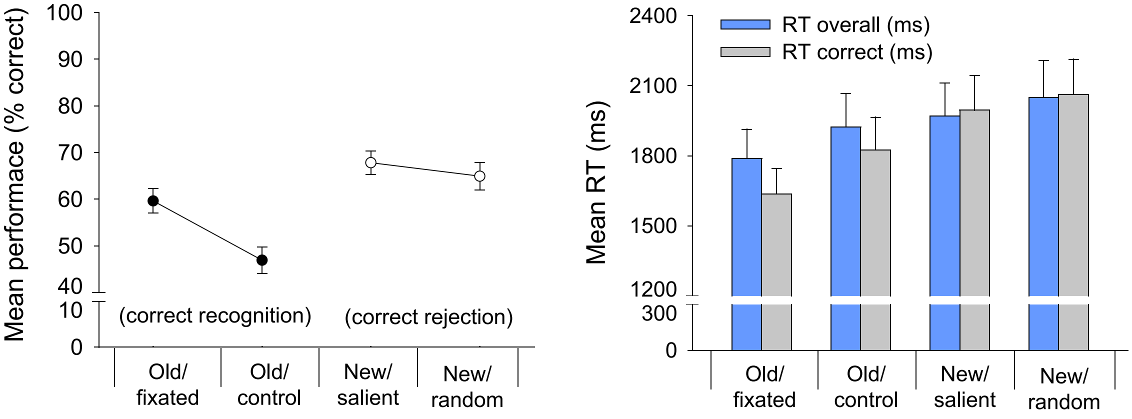

Figure 8. Mean performance (rate of correct responses in percent, left graph) and reaction times for all trials (right graph, blue bars), and only for trials in which participants responded correctly (right graph, gray bars) as a function of cutout type in the transfer block of Experiment 2. Error bars represent 1 *SEM*.

lient; new/random) run on the individual mean RTs revealed a main effect of cutout type, $F(3, 69) = 6.2$, $p < 0.01$, $\eta_p^2 = 0.21$. Post-hoc pairwise comparisons showed that participants responded significantly faster to old/fixated cutouts than to new/salient ($p < 0.05$) and new/random cutouts ($p < 0.01$), yet there was no significant difference in RTs between old/fixated and old/control cutouts ($p = 0.081$). No other differences reached significance.

In a second step, we confined the analysis to trials in which the participants responded correctly (disregarding all RTs from incorrect trials). Again, the analysis yielded a significant main effect of cutout type $F(3, 69) = 12.3$, $p < 0.001$, $\eta_p^2 = 0.35$. Post-hoc pairwise comparisons showed that participants were significantly faster in responding to old/fixated cutouts than to old/control ($p < 0.05$), new/salient ($p < 0.01$), and new/random cutouts ($p < 0.001$). Like in the previous analysis, RTs in old/control cutouts did not differ significantly from RTs in new/salient or new/random cutouts and no other differences between the conditions reached significance.

Next we analyzed accuracy. Hit rates were computed as the fraction of correct trials after eliminating trials with outlier RTs (see above). We ran another repeated-measures ANOVA on mean discrimination performance (i.e., hit rates as % correct in old/fixated and old/control trials, and correct rejections as % correct in new/salient and new/random trials) with the single variable cutout type during recognition which again revealed a significant effect, $F(3, 69) = 11.9$, $p < 0.001$, $\eta_p^2 = 0.34$. Here, post-hoc pairwise comparisons showed that the rate of correct responses was significantly lower with old/control cutouts than with any other cutout type (all $p < 0.01$). No other pairwise comparison reached significance. We further tested the rate of correct responses against the chance probability of 0.5 for a correct response. Performance was significantly higher than chance with old/fixated ($t[23] = 3.7$, $p < 0.01$), new/salient ($t[23] = 5.1$, $p < 0.001$), and new/random ($t[23] = 7.2$, $p < 0.001$) cutouts. However, hit rates for old/control cutouts were not significantly different from chance, $t(23) = -1.1$, *ns*.

Signal detection analysis of recognition performance

To ensure that higher recognition performance of old/fixated than old/control images reflected perceptual sensitivity rather than response biases, we employed Signal Detection Theory's indices (Green & Swets, 1966). The hit rate (i.e., the probability of "old" responses in old/fixated or old/control trials) and the false-alarm rate (i.e., the probability of responding with "old" for new/salient or new/random cutouts) was calculated for each participant. Together, these measures can be used to calculate an individual's perceptual recognition performance—that is, the sensitivity (d'), which reflects correct discrimination of old from new cutouts. The hit and false alarm rates can also be used to test whether observers were selectively biased towards one or the other of the response options in either of the conditions. We computed c as a measure for response bias. A c of zero would indicate that neither of the responses was favored. If c is significantly different from zero in the negative direction, this points to a bias towards responding with "old"; positive values reflect a bias towards "new" responses (Stanislaw & Todorov, 1999). The resulting values for d' and c are depicted in Figure 9.

For obtaining d' we computed the false alarm rate as the probability of an "old" response in either new/salient or new/random trials. Hit rate was computed separately for old/fixated and old/control trials in order to enable a comparison of sensitivity between these two

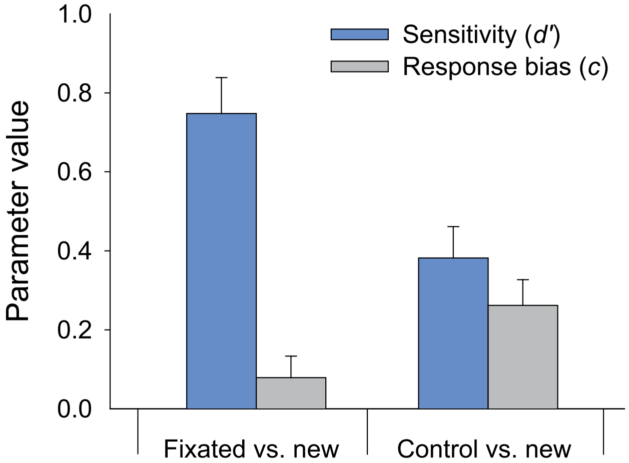

Figure 9. Signal detection theory indices for sensitivity (in blue) and response bias (in gray) in the transfer block of Experiment 2. False alarms were computed based on the probability of an "old" response in either new/salient or new/random trials, whereas hit rates were computed separately for old/fixated and old/control trials to enable a comparison of perceptual sensitivity between these two conditions. Error bars represent 1 *SEM*.

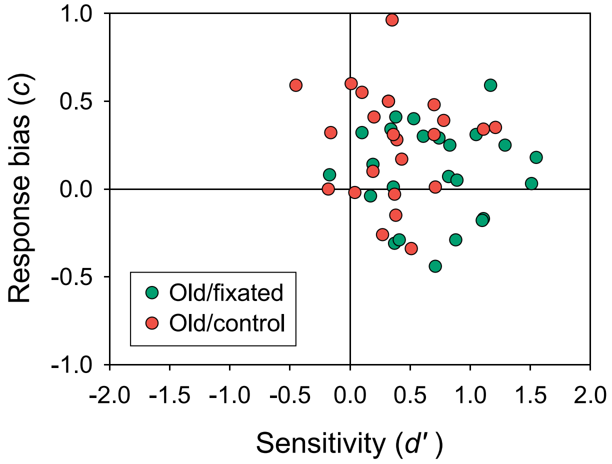

Figure 10. Individual *d*′ values on the ordinate as a function of individual *c* values on the abscissa for old/fixated cutouts (green disks) and old/control cutouts (red disks) in Experiment 2.

conditions. *d*′ was significantly different from zero in both old/fixated, $t(23) = 8.2$, $p < 0.001$, as well as old/control trials, $t(23) = 4.8$, $p < 0.001$. However, a paired *t* test indicated that sensitivity was significantly higher in old/fixated than in old/control cutouts, $t(23) = 6.0$, $p < 0.001$ (see also Figure 10).

On average, a response bias towards rejection was observable in both conditions. However, looking at *c* independently for old/fixated and old/control cutouts, respectively, showed that this response bias was only present in old/control trials, as reflected by a significant difference between the mean *c* values and zero, $t(23) = 4.0$, $p < 0.01$. No significant difference from zero was identified in old/fixated trials. Thus, the enhanced recognition performance with old/fixated cutouts truly reflected perceptual sensitivity and not a mere response bias.

How does low-level salience relate to human fixations and recognition?

Salience was proposed to be an important factor for the control of gaze (cf. Elazary & Itti, 2008). Therefore, our old/control cutouts were individually selected with respect to a maximum salience among the scene content that was not fixated by the participant. However, is it true that salience in the old/control cutouts was at least as high as in the old/fixated cutouts? To test this, we did a post-hoc comparison of the average salience in the area that corresponded to old/fixated and old/control cutouts, respectively, based on the initial (unmodified) salience map. To that end, we counted the relative frequency of trials in which the mean salience of the old/fixated cutout exceeded the mean salience of the old/control cutout within the same image for every participant. We found that the relative probability of trials in which the old/fixated cutout was more salient than the old/control cutout was low ($M = 0.17$, $SD = 0.079$), and significantly lower than 0.5 (where salience would be on average equal in both types of cutouts), $t(23) = -20.2$, $p < 0.001$. This means that the old/control cutouts were even more salient than the old/fixated cutouts. Consequently, any advantage in recognizing old/fixated over old/control cutouts should not be due to higher low-level salience in old/fixated cutouts.

Related to this: Was the assumption that salience drives the gaze justified at all? To answer this question, we also assessed the correlation between gaze and salience during scene learning, by using the salience maps as binomial classifiers for fixations (cf. Wilming et al., 2011). We computed a salience map for every one of the 60 scenes in the current stimulus set using the Salience Toolbox for MATLAB (Walther & Koch, 2006) with standard settings. We then applied a receiver operating characteristic (ROC) analysis of classification performance to evaluate how well a particular salience map detected actual fixations of participants viewing the corresponding image. For every salience map, a ROC curve was derived by stepwise thresholding of the salience map from its minimum to its maximum values. The maximum salience values in a circular region of 2° around average fixation coordinates were compared against the threshold value at each level. For fixated image regions, salience values above the threshold constituted a hit, and salience values below the threshold constituted a miss. Image regions that were not fixated but had a value above the threshold were treated as false alarms. The area under the ROC curve (AUC) represents the classification performance of the salience map. If the mean AUC across images does not differ significantly from 0.5, the salience model would not explain human fixations in the given class of images any better than chance. If however, the mean AUC value is significantly higher than 0.5, the salience maps classify fixated image regions with above chance accuracy. As some studies reported that the correlation of fixated locations and salience is highest during the early phase of scene examination (e.g., Parkhurst et al., 2002), we performed this analysis for all fixations, and once again, separately for only the first five fixations in a trial.

Across all images that were presented in the learning block, we obtained a mean AUC of 0.53 ($SD = 0.028$), which was significantly different from chance level (0.5), $t(59) = 8.97$, $p < 0.001$, and thus reflected above chance accuracy. Including only the first five fixations in every trial resulted in an identical accuracy, with a mean AUC of 0.53 ($SD = 0.039$), again significantly above chance, $t(59) = 6.18$, $p < 0.001$. Classification accuracy was not improved by including only the first five fixations in the analysis, $t(59) = -0.43$, *ns*. The result suggested that salience indeed correlated with the fixation behavior of our participants, and thus choosing the old/control cutouts based on highest salience in non-fixated scene regions was justified.

Discussion

In agreement with a supportive role of fixations for scene memory, we found that participants were much better at recognizing old/fixated cutouts than old/control cutouts (that they never fixated during learning). In fact, recognition performance in the old/control cutouts was not statistically different from chance performance, although the salience of these cutouts was at least as high as that of the old/fixated cutouts. This drastic drop in performance with the old/control cutouts was probably due to the fact that most of a scene's gist was not repeated during recognition: In Experiment 2, in an attempt to prevent scene recognition on the basis of gist alone (Oliva & Torralba, 2006), we have used image cutouts in all conditions. We may have expected better scene recognition even with nonfixated parts of the old images if we would have used larger cutouts of the learned images during recognition with the old/control cutouts. In fact, even with the old/fixated cutouts that we used recognition performance was not optimal: Correct recognition of old/fixated cutouts was numerically below correct rejection of new cutouts. This means that even with the old/fixated cutouts, recognition performance was not at ceiling, suggesting that scene gist could have fostered the recognition of all old images.

Also, we observed that in the new images, performance was good and that it was not significantly affected by salience either. This indicates that salient and less salient image positions were approximately equally informative for the classification of the new images. Of course, in an eye-tracking experiment, recognition of salient regions would not only be facilitated because of the informativeness of salient regions (cf. van der Linde et al., 2009), but also via attraction of the gaze in a stimulus-driven way (cf. Itti & Koch, 2000, 2001). To note, the bias to gaze at salient regions in the image could account for (part of) the correlation between salience and fixations that we found in the learning blocks of the present experiment. Yet, this low-level influence of salience's gaze attraction was ruled out in the recognition blocks because cutouts were presented at fixation (at the center of the screen).

In relating our findings from Experiment 2 to salience, however, one should note that locations of low-level salience and of objects are often confounded in images of natural scenes. For example, Einhäuser et al. (2008) reported that interesting objects in scenes are stronger predictors of fixations than visual salience alone, and that visual salience adds only little explanatory power on top of the locations of objects. In line with this object-based view of fixation selection, Nuthmann and Henderson (2010) reported that viewers preferentially saccade close to the center of objects in a scene. These results point to a potential limitation of the present findings, namely the possibility that our (nonfixated) old/control cutouts, although they were (on average) more salient than the old/fixated cutouts, exhibited fewer interesting objects and thus, might have been less informative than the old/fixated cutouts. Consequently, the performance benefit we found for old/fixated cutouts as compared to old/control cutouts might have been partly due to interesting objects being more frequently present in old/fixated cutouts.

To conclude, our findings from Experiment 2 also suggested that fixations were valid reflections of attention. This is an important point to note because, as it was first described by Helmholtz (1867), and later reported by Posner (1980), the direction of covert attention and the direction of fixation do not have to co-align in all instances. While the eyes remain fixated, attention can be covertly allocated to different regions in the image, and it is only immediately prior to an upcoming saccade that attention is shifted to the landing location of this saccade (Deubel & Schneider, 1996; Kowler, Anderson, Dosher, & Blaser, 1995). Yet, in the present experiment we found clear evidence that the information from fixated locations was also more strongly represented in the memory of our participants:

Fixating a detail in the present experiment's learning block improved recognition as compared to not fixating a detail. In conclusion, after Experiment 2, it seems clear that fixations on scene details can have a supportive influence on scene memory.

General discussion

Our experiments showed that (a) particular objects or locations that a participant fixated during learning were repeatedly fixated during recognition by the same participant (Experiment 1), and (b) that recognition was facilitated by looking at a previously fixated location (Experiment 2). These supportive effects of fixations on scene memory are in general agreement with visual memory theory (Hollingworth & Henderson, 2002), and they also align with some results of research conducted to explore scanpath theory (cf. Didday & Arbib, 1975; Foulsham & Kingstone, 2013; Foulsham & Underwood, 2008; Stark & Ellis, 1981; Underwood et al., 2009). In two respects, however, the present results go beyond these past findings. Here, we showed that the behavioral fixation pattern does also hold in conditions in which the view of a scene undergoes a change of perspective between learning and recognition and that the fixation-repetition pattern critically depended on the task of the participants: Repeated fixations during the two blocks were more frequently found with a recognition task than with a free-viewing task. Thus, we also ruled out that stimulus-driven salience alone provided an explanation for the behavioral fixation-repetition effect (see also Foulsham & Underwood, 2008).

What we cannot say with certainty after the present experiments is what kind of memory was responsible for the effects. On the one hand, it is possible that scene memory was supported by visual input taken up during the fixations proper (cf. Sanocki, 2003). For example, maybe the representation of some view-dependent layout properties of the scenes (e.g., Wood, 2010) benefited from directing fixations to these properties. On the other hand, it is also possible that visual object information itself (e.g., Melcher & Kowler, 2001) or local scene context (e.g., Hollingworth & Rasmussen, 2010) accounted for the recognition effect and the facilitation of repeated fixations during learning and recognition. Because we neither changed the objects nor the local scene properties from learning to recognition, the encoding of objects and/or local scene properties was probably also helpful for later scene recognition. It is even possible that the participants encoded some of the material that they took up during the fixations in a semantic or verbal way and that the corresponding memory effects rested upon retrieval of scene-specific semantic knowledge.

We have also repeatedly argued that many of our findings are in line with scanpath theory (Noton & Stark, 1971; Stark & Ellis, 1981). In particular, it is possible that some of our findings in Experiment 1 reflected similar scanpaths during learning and recognition of a scene. However, scanpath theory would have predicted not only a repetition of fixations during learning and recognition; it would have also predicted a particular order in which the fixations during recognition would have been made. Relatedly, scanpath theory emphasizes that visual recognition is partly brought about by the pattern of sensorimotor signals, which couple specific visual input with specific eye movements. These aspects of scanpath theory do not correspond (exactly) to our findings. In particular, the results in the present Experiment 2 clearly showed that scene recognition is possible by the viewer looking at the content that is picked up during the longest scene-learning fixation alone. Also, for this beneficial effect of fixated content on memory, the viewer did not have to move her/his eyes and it was not necessary to present the recognized input during recognition at the same position as in the learning image. This finding does not leave much room for a beneficial influence of sensorimotor correspondences between learning and later recognition. In line with this negative verdict, we observed a large number of repeated fixations in the old/shifted images in Experiment 1, which required ignoring learnt sensorimotor correspondences and looking at a shifted location in the recognized image relative to the learned image. Admittedly, however, there are versions of the scanpath theory that would allow recognition when an image undergoes a change of the perspective, for instance, with rigid position shifts for all fixated locations as it was the case in the old/shifted images of the present Experiment 1, for example.

The relation between recognition and visual search

Another important question concerns the relation of our findings to standard observations of repetition priming and contextual cueing in more controlled laboratory search experiments. To start with, the repetition of features is a strong attractor of attention in visual search experiments. Maljkovic and Nakayama (1994), for example, presented their participants one color singleton (e.g., in green) as a pop-out stimulus among two equally colored distractors of a different color (e.g., two red distractors), and the participants had to search for the color singleton and report its shape. In this situation, Maljkovic and Nakayama observed facilitated search when the color of the singleton repeated from trial to trial. Maljkovic and Nakayama called their finding *priming of pop-out*. Such priming of attention capture by repeated features is a very robust finding and has been shown numerous times during visual search, in the form of faster search times for stimuli with repeated features (Fecteau, 2000; Hillstrom, 2000; Maljkovic & Martini, 2005; for a review see Kristjánsson & Campana, 2010), and as faster saccades to stimuli with repeated features and earlier fixations on them (Becker, 2008; Becker, Ansorge, & Horstmann, 2009; McPeek, Maljkovic, & Nakayama, 1999). Related observations showed that holding a feature in working memory alone can be sufficient for facilitated search (Ansorge & Becker, 2012; Olivers, 2009; Olivers, Meijer, & Theeuwes, 2006) and that the repetition of stimuli can lead to tacit knowledge of the displays that can facilitate searching even with a large number of interleaving trials without feature repetition and when the target is not a pop-out stimulus (Chun & Jiang, 1998, 1999; Jiang & Wagner, 2004). Is there a connection between the repetition effects found during typical visual search experiments and the repetition effects during recognition of images, as observed in the present study?

It seems so, as indicated by findings showing that contextual cueing effects can also be found when participants have to search for visual objects under conditions with more variable and natural scene contexts (Brockmole, Castelhano, & Henderson, 2006; Brooks, Rasmussen, & Hollingworth, 2010). Although findings of Brockmole et al. (2006) suggested that global scene input was responsible for contextual cueing with more natural images, Brooks et al. (2010) found evidence for both local and global contextual cueing of search with natural images. The later finding is particularly interesting in the present context because at least with the cutouts of the present Experiment 2, it is more likely that local than global context provided the necessary input during later recognition. On a more general level, one can conceive of visual search tasks as recognition tasks, too (Nakayama & Martini, 2011; Zelinsky, 2008). Imagine that your task would be to find a singleton target that changes its position and its color randomly from trial to trial. This is the standard condition in the visual search experiments of Maljkovic and Nakayama (1994). This situation is not so different from scene recognition under conditions of varying perspective. In general, to find a target during search, one could draw on a template of target-defining features (its color, shape, luminance, and so on) that discriminate between target and distractors and that one holds in memory (Duncan & Humphreys, 1989; Wolfe, 1994). However, because most of the candidate features (such as shape and luminance) would be the same for target and distractors and other features (such as color) would be changing from trial to trial, one would not be able to find the target by any of these features, without first noticing which stimulus is the actual target. In this situation, the memory of the target's last appearance (its last color, its last position, and so on) might be used as a default search template because this template would noticeably help on some of the trials to find the target or because this search template would passively carry over to the next trial. Probably many of these repetition priming and contextual cueing effects would be a mixture of both these processes.

A final comment concerns the exact stage of memory processes during which the helpful influence of fixations took place in the current study. We are relatively certain that fixations facilitated *encoding* of the critical features into memory. However, we cannot say much about the role of fixations during the *retrieval* of an image. Of course, in the current study, the preferred fixations during recognition that were directed at previously fixated image parts in Experiment 1 would be suggestive of a retrieval-based effect. Yet, such fixation preferences could be a mere side-effect of processes owing to encoding of the features alone. Related, from Experiment 1, nothing could be concluded about the recognition performance at all. The latter gap was filled by Experiment 2, but again, Experiment 2 lacks the critical control condition that would be needed to pinpoint a retrieval-based effect of the fixations. A role of fixations during retrieval could be shown, for example, where the material that was fixated during encoding (or learning) would be presented during recognition at a nonfixated or at a fixated position. In this situation, a retrieval-based recognition effect of the fixations should be reflected in better performance with previously fixated and repeated image content that is also fixated in the recognition trials as compared to previously fixated and repeated image content that is not fixated during the recognition trials. The question whether retrieval is affected by fixations in the context of the present study, thus, has to await further testing in the future.

Conclusion

Our present article presents new evidence for a role of repeated fixations for improved memory during scene recognition. Repeated fixations under recognition conditions exceeded the influence of salience alone. In addition, we discussed links between scene recognition and visual search that emphasize the importance of repetition priming under various conditions.

Keywords: attention, eye movements, fixations, priming, recognition, salience, visual memory

Acknowledgments

The authors thank Gerhard Böck, Alexander Kudrna, Anna Lukyanova, and Helena Stockinger for assistance with data collection. This research was supported by grants from the Wiener Wissenschafts-, Forschungs-, und Technologiefonds (WWTF, Vienna Science and Technology Fund) grant no. CS 11-009 awarded to Ulrich Ansorge, Shelley Buchinger, and Otmar Scherzer, and from the Australian Research Council (ARC APD) grant no. DP110100588 awarded to Stefanie I. Becker.

Commercial relationships: none.
Corresponding author: Christian Valuch.
Email: christian.valuch@univie.ac.at.
Address: Faculty of Psychology, University of Vienna, Austria.

Footnote

[1] The new images were only presented as filler items, to allow the necessary discrimination in the transfer/recognition blocks of the recognition task, and could thus not be analyzed with regard to repeated fixations from learning to recognition.

References

Ansorge, U., & Becker, S. I. (2012). Automatic priming of attentional control by relevant colors. *Attention, Perception, & Psychophysics*, *74*, 83–104. [PubMed]

Bacon-Macé, N., Kirchner, H., Fabre-Thorpe, M., & Thorpe, S. J. (2007). Effects of task requirements on rapid natural scene processing: From common sensory encoding to distinct decisional mechanisms. *Journal of Experimental Psychology: Human Perception & Performance*, *33*, 1013–1026. [PubMed]

Ballard, D. H., Hayhoe, M. M., Pook, P. K., & Rao, R. P. N. (1997). Deictic codes for the embodiment of cognition. *Behavioral & Brain Sciences*, *20*, 723–767. [PubMed]

Becker, S. I. (2008). Can intertrial effects of features and dimensions be explained by a single theory? *Journal of Experimental Psychology: Human Perception & Performance*, *34*, 1417–1440. [PubMed]

Becker, S. I., Ansorge, U., & Horstmann, G. (2009). Can intertrial priming effects account for the similarity effect in visual search? *Vision Research*, *49*, 1738–1756. [PubMed]

Brady, T. F., Konkle, T., Alvarez, G. A., & Oliva, A. (2008). Visual long-term memory has a massive storage capacity for object details. *Proceedings of the National Academy of Sciences, USA*, *105*, 14 325–14 329. doi:10.1073/pnas.0803390105. [PubMed]

Brainard, D. H. (1997). The psychophysics toolbox. *Spatial Vision*, *10*, 433–436. [PubMed]

Bridgeman, B., Hendry, D., & Stark, L. (1975). Failure to detect displacements of the visual world during saccadic eye movements. *Vision Research*, *15*, 719–722. [PubMed]

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

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

Brooks, D. I., Rasmussen, I. P., & Hollingworth, A. (2010). The nesting of search contexts within natural scenes: Evidence from contextual cueing. *Journal of Experimental Psychology: Human Perception & Performance*, *36*, 1406–1418. [PubMed]

Buswell, G. T. (1935). *How people look at pictures.* Chicago: University of Chicago Press.

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

Castelhano, M. S., Mack, M., & Henderson, J. M. (2009). Viewing task influences eye movements during active scene perception. *Journal of Vision*, *9*(3):6, 1–15, http://www.journalofvision.org/content/9/3/6, doi:10.1167/9.3.6. [PubMed] [Article]

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

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

Deubel, H., & Schneider, W. X. (1996). Saccade target selection and object recognition: Evidence for a common attentional mechanism. *Vision Research*, *36*, 1827–1837. [PubMed]

Didday, R. L., & Arbib, M. A. (1975). Eye movements and visual perception: A "Two Visual System" model. *International Journal of Man-Machine Studies*, *7*, 547–569.

Droll, J. A., Hayhoe, M. M., Triesch, J., & Sullivan, B.

T. (2005). Task-demands control acquisition and storage of visual information. *Journal of Experimental Psychology: Human Perception & Performance*, *31*, 1416–1431. [PubMed]

Duncan, J., & Humphreys, G. W. (1989). Visual search and stimulus similarity. *Psychological Review*, *96*, 433–458. [PubMed]

Einhäuser, W., Rutishauser, U., & Koch, C. (2008). Task-demands can immediately reverse the effects of sensory-driven saliency in complex visual stimuli. *Journal of Vision*, *8*(2):2, 1–19, http://www.journalofvision.org/content/8/2/2, doi:10.1167/8.2.2. [PubMed] [Article]

Einhäuser, W., Spain, M., & Perona, P. (2008). Objects predict fixations better than early saliency. *Journal of Vision*, *8*(14):18, 1–26, http://www.journalofvision.org/content/8/14/18, doi:10.1167/8.14.18. [PubMed] [Article]

Elazary, L., & Itti, L. (2008). Interesting objects are visually salient. *Journal of Vision*, *8*(3):3, 1–15, http://www.journalofvision.org/content/8/3/3, doi:10.1167/8.3.3. [PubMed] [Article]

Fecteau, J. H. (2007). Priming of pop-out depends upon the current goals of observers. *Journal of Vision*, *7*(6):1, 1–11, http://www.journalofvision.org/content/7/6/1, doi:10.1167/7.6.1. [PubMed] [Article]

Foulsham, T., & Kingstone, A. (2013). Fixation-dependent memory for natural scenes: An experimental test of scanpath theory. *Journal of Experimental Psychology: General*, *142*, 41–56. [PubMed]

Foulsham, T., & Underwood, G. (2008). What can saliency models predict about eye movements? Spatial and sequential aspects of fixations during encoding and recognition. *Journal of Vision*, *8*(2):6, 1–17, http://www.journalofvision.org/content/8/2/6, doi:10.1167/8.2.6. [PubMed] [Article]

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

Goferman, S., Zelnik-Manor, L., & Tal, A. (2010). Context-aware saliency detection. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, art. no. *5539929*, 2376–2383.

Green, D. M., & Swets, J. A. (1966). *Signal detection theory and psychophysics*. New York: Wiley.

Helmholtz, H. (1867). *Handbuch der physiologischen optik* [*Handbook of physiological optics*]. Leipzig, Germany: Voss.

Henderson, J. M. (2003). Human gaze control during real-world scene perception. *Trends in Cognitive Sciences*, *7*, 498–504. [PubMed]

Henderson, J. M., Williams, C. C., Castelhano, M. S., & Falk, R. J. (2003). Eye movements and picture processing during recognition. *Perception & Psychophysics*, *65*, 725–734. [PubMed]

Hillstrom, A. P. (2000). Repetition effects in visual search. *Perception & Psychophysics*, *62*, 800–817. [PubMed]

Hirose, Y., Kennedy, A., & Tatler, B. W. (2010). Perception and memory across viewpoint changes in moving images. *Journal of Vision*, *10*(4):2, 1–20, http://www.journalofvision.org/content/10/4/2, doi:10.1167/10.4.2. [PubMed] [Article]

Hollingworth, A., & Henderson, J. M. (2002). Accurate visual memory for previously attended objects in natural scenes. *Journal of Experimental Psychology: Human Perception & Performance*, *28*, 113–136.

Hollingworth, A., & Rasmussen, I. P. (2010). Binding objects to locations: The relationship between object files and visual working memory. *Journal of Experimental Psychology: Human Perception & Performance*, *36*, 543–564. [PubMed]

Hollingworth, A., Williams, C. C., & Henderson, J. M. (2001). To see and remember: Visually specific information is retained in memory from previously attended objects in natural scenes. *Psychonomic Bulletin & Review*, *8*, 761–768. [PubMed]

Intraub, H. (1981). Rapid conceptual identification of sequentially presented pictures. *Journal of Experimental Psychology: Human Perception & Performance*, *7*, 604–610.

Intraub, H. (2012). Rethinking visual scene perception. *Wiley Interdisciplinary Reviews: Cognitive Science*, *3*, 117–127.

Irwin, D. E., & Zelinsky, G. J. (2002). Eye movements and scene perception: Memory for things observed. *Perception & Psychophysics*, *64*, 882–895. [PubMed]

Itti, L., & Baldi, P. (2009). Bayesian surprise attracts human attention. *Vision Research*, *49*, 1295–1306. [PubMed]

Itti, L., & Koch, C. (2000). A saliency-based search mechanism for overt and covert shifts of visual attention. *Vision Research*, *40*, 1489–1506. [PubMed]

Itti, L., & Koch, C. (2001). Computational modeling of visual attention. *Nature Reviews Neuroscience*, *2*, 194–203. [PubMed]

Itti, L., Koch, C., & Niebur, E. (1998). A model of saliency-based visual attention for rapid scene analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *20*, 1254–1259.

Jiang, Y., & Wagner, L. C. (2004). What is learned in spatial contextual cuing—Configuration or individual locations? *Perception & Psychophysics*, *66*, 454–463. [PubMed]

Kaspar, K., & König, P. (2011). Viewing behavior and the impact of low-level image properties across repeated presentations of complex scenes. *Journal of Vision*, *11*(13):26, 1–29, http://www.journalofvision.org/content/11/13/26, doi:10.1167/11.13.26. [PubMed] [Article]

Konkle, T., Brady, T. F., Alvarez, G. A., & Oliva, A. (2010). Scene memory is more detailed than you think: The role of categories in visual long-term memory. *Psychological Science*, *21*, 1551–1556. [PubMed]

Kowler, E. (2011). Eye movements: The past 25 years. *Vision Research*, *51*, 1457–1483. [PubMed]

Kowler, E., Anderson, E., Dosher, B., & Blaser, E. (1995). The role of attention in the programming of saccades. *Vision Research*, *35*, 1897–1916. [PubMed]

Kristjánsson, Á., & Campana, G. (2010). Where perception meets memory: A review of repetition priming in visual search tasks. *Attention, Perception & Psychophysics*, *72*, 5–18. [PubMed]

Maljkovic, V., & Martini, P. (2005). Implicit short-term memory and event frequency effects in visual search. *Vision Research*, *45*, 2831–2846. [PubMed]

Maljkovic, V., & Nakayama, K. (1994). Priming of pop-out: I. Role of features. *Memory & Cognition*, *22*, 657–672. [PubMed]

McConkie, G. W., & Currie, C. B. (1996). Visual stability across saccades while viewing complex pictures. *Journal of Experimental Psychology: General*, *22*, 563–581. [PubMed]

McPeek, R. M., Maljkovic, V., & Nakayama, K. (1999). Saccades require focal attention and are facilitated by a short-term memory system. *Vision Research*, *39*, 1555–1566. [PubMed]

Melcher, D., & Kowler, E. (2001). Visual scene memory and the guidance of saccadic eye movements. *Vision Research*, *41*, 3597–3611. [PubMed]

Nakayama, K., & Martini, P. (2011). Situating visual search. *Vision Research*, *51*, 1526–1537. [PubMed]

Noton, D., & Stark, L. (1971). Scanpaths in saccadic eye movements while viewing and recognizing patterns. *Vision Research*, *11*, 929–942. [PubMed]

Nuthmann, A., & Henderson, J. M. (2010). Object-based attentional selection in scene viewing. *Journal of Vision*, *10*(8):20, 1–19, http://www.journalofvision.org/content/10/8/20, doi:10.1167/10.8.20. [PubMed] [Article]

Oliva, A., & Torralba, A. (2006). Building the gist of a scene: The role of global image features in recognition. *Progress in Brain Research*, *155*, 23–36. [PubMed]

Olivers, C. N. L. (2009). What drives memory-driven attentional capture? The effects of memory type, display type, and search type. *Journal of Experimental Psychology: Human Perception & Performance*, *35*, 1275–1291. [PubMed]

Olivers, C. N. L., Meijer, F., & Theeuwes, J. (2006). Feature-based memory-driven attentional capture: visual working memory content affects visual attention. *Journal of Experimental Psychology: Human Perception & Performance*, *32*, 1243–1265. [PubMed]

Parkhurst, D., Law, K., & Niebur, E. (2002). Modeling the role of salience in the allocation of overt visual attention. *Vision Research*, *42*, 107–123. [PubMed]

Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: Transforming numbers into movies. *Spatial Vision*, *10*, 437–442. [PubMed]

Pertzov, Y., Zohari, E., & Avidan, G. (2009). Implicitly perceived objects attract gaze during later free viewing. *Journal of Vision*, *9*(6):6, 1–12, http://www.journalofvision.org/content/9/6/6, doi:10.1167/9.6.6. [PubMed] [Article]

Posner, M. I. (1980). Orienting of attention. *Quarterly Journal of Experimental Psychology*, *32*(1), 3–25. [PubMed]

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

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

Sanocki, T. (2003). Representation and perception of scenic layout. *Cognitive Psychology*, *47*, 63–86. [PubMed]

Sanocki, T., & Epstein, P. (1997). Priming spatial layout of scenes. *Psychological Science*, *8*, 374–378.

Schütz, A. C., Braun, D. I., & Gegenfurtner, K. R. (2011). Eye movements and perception: A selective review. *Journal of Vision*, *11*(5):9, 1–30, http://www.journalofvision.org/content/11/5/9, doi:10.1167/11.5.9. [PubMed] [Article]

Schyns, P., & Oliva, A. (1994). From blobs to boundary edges: Evidence for time- and spatial-scale-dependent scene recognition. *Psychological Science*, *5*, 195–200.

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

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

Stanislaw, H., & Todorov, N. (1999). Calculation of signal detection theory measures. *Behavior Research Methods, Instruments, & Computers*, *31*, 137–149. [PubMed]

Stark, L., & Ellis, S. R. (1981). Scanpaths revisited: Cognitive models direct active looking. In D. F. Fisher, R. A. Monty, & J. W. Senders (Eds.), *Eye movements: Cognition and visual perception* (pp. 193–227). Hillsdale, NJ: Lawrence Erlbaum.

Thorpe, S. J., Fize, D., & Malot, C. (1996). Speed of processing in the human visual system. *Nature*, *381*, 520–522. [PubMed]

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. [PubMed]

Underwood, T., Foulsham, G., & Humphrey, K. (2009). Saliency and scan patterns in the inspection of real-world scenes: Eye movements during encoding and recognition. *Visual Cognition*, *17*, 812–834.

Valenti, R., Sebe, N., & Gevers, T. (2009). Image saliency by isocentric curvedness and color. *Proceedings of the International Conference on Image Processing*, art. no. *5413810*, 993–996.

van der Linde, I., Rajashekar, U., Bovik, A. C., & Cormack, L. K. (2009). Visual memory for fixated regions of natural images dissociates attraction and recognition. *Perception*, *38*, 1152–1171. [PubMed]

VanRullen, R. (2007). The power of the feed-forward sweep. *Advances in Cognitive Psychology*, *3*, 167–176. [PubMed]

Walther, D., & Koch, C. (2006). Modeling attention to salient proto-objects. *Neural Networks*, *19*, 1395–1407. [PubMed]

Wilming, N., Betz, T., Kietzmann, T. C., & König, P. (2011). Measures and limits of models of fixation selection. *PLoS ONE*, *6*(9), e24038. [PubMed]

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

Wood, J. N. (2010). A core knowledge architecture of visual working memory. *Journal of Experimental Psychology: Human Perception & Performance*, *37*, 357–381. [PubMed]

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

Zelinsky, G. J. (2008). A theory of eye movements during target acquisition. *Psychological Review*, *115*, 787–835. [PubMed]