Influence of a psychological perspective on scene viewing and memory for scenes
Johanna K. Kaakinen a, Jukka Hyönä a & Minna Viljanen a
a Department of Psychology, University of Turku, Turku, Finland
Available online: 03 Mar 2011

To cite this article: Johanna K. Kaakinen, Jukka Hyönä & Minna Viljanen (2011): Influence of a psychological perspective on scene viewing and memory for scenes, The Quarterly Journal of Experimental Psychology, 64:7, 1372-1387
To link to this article: http://dx.doi.org/10.1080/17470218.2010.548872

PLEASE SCROLL DOWN FOR ARTICLE

Full terms and conditions of use: http://www.tandfonline.com/page/terms-and-conditions

This article may be used for research, teaching and private study purposes. Any substantial or systematic reproduction, re-distribution, re-selling, loan, sub-licensing, systematic supply or distribution in any form to anyone is expressly forbidden.

The publisher does not give any warranty express or implied or make any representation that the contents will be complete or accurate or up to date. The accuracy of any instructions, formulae and drug doses should be independently verified with primary sources. The publisher shall not be liable for any loss, actions, claims, proceedings, demand or costs or damages whatsoever or howsoever caused arising directly or indirectly in connection with or arising out of the use of this material.
Influence of a psychological perspective on scene viewing and memory for scenes

Johanna K. Kaakinen, Jukka Hyöna, and Minna Viljanen
Department of Psychology, University of Turku, Turku, Finland

In the study, 33 participants viewed photographs from either a potential homebuyer's or a burglar's perspective, or in preparation for a memory test, while their eye movements were recorded. A free recall and a picture recognition task were performed after viewing. The results showed that perspective had rapid effects, in that the second fixation after the scene onset was more likely to land on perspective-relevant than on perspective-irrelevant areas within the scene. Perspective-relevant areas also attracted longer total fixation time, more visits, and longer first-pass dwell times than did perspective-irrelevant areas. As for the effects of visual saliency, the first fixation was more likely to land on a salient than on a nonsalient area; salient areas also attracted more visits and longer total fixation time than did nonsalient areas. Recall and recognition performance reflected the eye fixation results: Both were overall higher for perspective-relevant than for perspective-irrelevant scene objects. The relatively low error rates in the recognition task suggest that participants had gained an accurate memory for scene objects. The findings suggest that the role of bottom-up versus top-down factors varies as a function of viewing task and the time-course of scene processing.

Keywords: Scene perception; Eye tracking; Visual memory.
when reading a written description of a house, a pretend burglar spends longer time on sentences describing burglar-relevant details than on sentences describing other details of the house (e.g., Kaakinen & Hyöna, 2008). After reading, people are more likely to recall perspective-relevant than perspective-irrelevant details of the description (e.g., Kaakinen & Hyöna, 2008).

Previous studies show that the viewer's task plays an important role in eye guidance during scene viewing. For example, the eyes are directed to different locations in scenes during visual search and memorization tasks (Buswell, 1935; Castelhano, Mack, & Henderson, 2009). Eye movements also differ between different kinds of visual search tasks (e.g., count the number of people vs. paintings vs. mugs; Torralba, Oliva, Castelhano, & Henderson, 2006), such that locations where the searched objects are likely to appear (upper vs. bottom half of the picture) are fixated with a higher probability. Relevant to the influence of a psychological perspective on scene viewing, Yarbus (1967) recorded eye movements of an observer while he was viewing I. E. Repin's painting “They did not expect him” (in the painting, a man walks into a room where there are several people by a coffee table). The participant was asked to view the painting repeatedly with different perspective instructions—for example, to estimate the material circumstances of the family or to estimate how long the unexpected visitor had been away. Yarbus's results showed that the viewer looked mainly at the faces of the people when the task was to evaluate the time the visitor had been away, whereas when the task was to evaluate the economical situation of the family, the viewer also looked at the clothes of the people (for recent replications of the Yarbus study, see Benson, Piper, & Fletcher-Watson, 2009; DeAngelus & Pelz, 2009). Yarbus's results thus demonstrate that the perspective taken by the viewer has an influence on how the viewer inspects the scene. However, several questions remain unanswered: (a) What is the time-course of the perspective effects—that is, how early on is the eye gaze directed to perspective-relevant objects within the scene; (b) to what extent is the eye gaze governed also by image properties, such as visual saliency, when the viewer is inspecting a scene with a particular perspective in mind, and (c) how does the viewing perspective influence memory for scene information? In the present study, we aimed to answer these questions.

According to theories of visual attention in scene viewing, the eyes are “pulled” by image properties, such as intensity, colour, and orientation (i.e., saliency; Itti & Koch, 2001; Parkhurst, Law, & Niebur, 2002) on one hand, and on the other hand they are “pushed” by top-down factors, such as the observer’s prior knowledge and task (e.g., Henderson, 2003; Navalpakkam & Itti, 2005; Torralba et al., 2006). However, the role of bottom-up versus top-down factors under different task demands is still an unresolved question (Tatler, 2009). Previous research suggests that during free viewing of natural scenes, the bottom-up and top-down factors are likely to work together in guiding the eyes: Visual saliency of image regions, rated interestingness, and fixation time within the regions are all highly intercorrelated (Henderson, Brockmole, Castelhano, & Mack, 2007; Masciocchi, Mihalas, Parkhurst, & Niebur, 2009). In other words, visually salient regions in the scenes are also rated as more interesting, and people tend to look at these regions. However, some studies suggest that with specific tasks (visual search in particular), task demands override the influence of visual saliency (e.g., Einhäuser, Rutishauser, & Koch, 2008; Foulsham & Underwood, 2007; Henderson et al., 2007; Underwood & Foulsham, 2006; Underwood, Foulsham, van Loom, Humphreys, & Bloyce, 2006). It has been suggested that prior expectations about the spatial location of the target is used to quickly direct the eyes to task-relevant scene regions (e.g., Navalpakkam & Itti, 2005; Torralba et al., 2006). Moreover, task instructions are likely to prime task-relevant physical features so that attention is biased to favour scene areas in which these features are detected (e.g., Navalpakkam & Itti, 2005; Wolfe, 2007).

However, it is difficult to say whether the results obtained with simple visual search tasks generalize to other task environments. For
example, visual search poses different demands to the viewer than viewing a scene with a psychological perspective in mind. First of all, in visual search the viewer scans the scene for certain physical features (e.g., a red ball) and responds whether the searched object is present or not. When viewing a picture from a certain perspective (e.g., a potential homebuyer looking at photographs of house interiors), the objects depicted in the picture are evaluated in the light of the knowledge activated by the perspective: For example, a potential homebuyer would be interested in anything that indicates repair needs. These potential regions of interest could be located almost anywhere within the scene. It is thus possible that because the perspective does not (necessarily) define a restricted set of visual features or certain locations in the scene as task relevant, as is the case in simple visual search tasks, visual saliency might play a more prominent role in perspective-driven scene viewing than in visual search. Second, psychological perspective emphasizes memory encoding of perspective-relevant details. If perspective-relevant objects or regions are detected in the scene, information about them has to be encoded to memory so that it can be subsequently used as a basis for future actions (e.g., in order to make a decision whether one is willing to buy the house). In simple visual search, it is enough to be able to indicate whether the search object was present or not—remembering visual details of the scene is of less importance.

Memory for scenes builds up across time, so that the more time the viewer is allowed to look at the scene, the more visual details will be remembered (e.g., Melcher, 2006; Tatler, Gilchrist, & Rusted, 2003). Memory for visual details accumulates across the fixations made on the scene (e.g., Hollingworth & Henderson, 2002); memory for object details is the better the longer the fixation time is on the object (Tatler, Gilchrist, & Land, 2005). Based on the research examining the relationship between viewing time and memory performance (e.g., Hollingworth & Henderson, 2002; Melcher, 2006; Tatler et al., 2005; Tatler et al., 2003), it can be assumed that objects that are fixated for a longer time are also remembered better than objects that attract shorter fixation times. Thus, if viewing perspective guides viewers’ eyes to perspective-relevant objects within the scene, viewers should also have better memory for perspective-relevant than for perspective-irrelevant objects. However, previous research suggests that relatively detailed scene memory is acquired even when the task does not emphasize memory encoding (Castelhano & Henderson, 2005): Visual details are encoded for fixated objects also during simple visual search. In other words, good memory for the objects presented in the scene does not necessarily require intentional memory encoding. It is thus possible that the viewing perspective does not actually restrict memory for the scenes but viewers have a relatively good memory also of the perspective-irrelevant objects.

In the present study, we examined the role of visual saliency and perspective relevance in eye guidance during scene viewing. We were also interested in whether the psychological perspective adopted by the viewer restricts memory for the scenes. In the present experiment, participants viewed natural photographs from a given perspective (a potential burglar, a homebuyer, or in preparation for a memory test), while their eye movements were tracked. Areas of interest to burglars and homebuyers were identified in the pictures, and saliency values for the images were computed. Relevance ratings and saliency values were used to define salient and nonsalient burglar- and homebuyer-relevant areas within the scenes. Eye fixation patterns were registered during scene viewing. After the viewing session, participants produced a free recall of the scenes and completed an object recognition task.

We used a recognition task akin to the false memory paradigm for pictures (e.g., Koutstaal & Schacter, 1997) to study the nature of the memory representation for scene information. In the recognition task, pictures of the objects presented in the scenes (“old” objects), pictures of objects belonging to the same category as the original objects (lures), and pictures of objects not related to the presented scenes (“new” objects) were presented. For example, if a scene included a CD-player (which was a burglar-relevant object),
a picture of the CD-player was presented in the recognition task as an old object, and a picture of another CD-player was presented as a lure.

We expected that the viewing perspective guides the eyes to perspective-relevant details within the scene, so that perspective-relevant regions (as defined in a pretest) would attract more and longer eye fixation times than perspective-irrelevant regions (Benson et al., 2009; DeAngelus & Pelz, 2009; Yarbus, 1967). As regards effects of visual saliency, previous research suggests that during visual search, task instructions may completely override the influence of saliency (e.g., Einhäuser et al., 2008; Foulsham & Underwood, 2007; Henderson et al., 2007; Underwood & Foulsham, 2006; Underwood et al., 2006). However, if saliency does play a role in where the eyes are likely to land within a picture when a viewer has a specific perspective in mind, we should see more eye fixations on salient than on nonsalient targets.

We also expected participants to show superior recall of the scene contents that receive the most fixation time (e.g., Hollingworth & Henderson, 2002; Tatler et al., 2005). If visual details of scene objects are encoded to memory, then in the recognition task participants should make relatively few errors of accepting the lures as “old”. However, if information about scene objects is represented on a more abstract, gist-based level, then participants may accept the lures as “old” with a relatively high probability. Moreover, if perspective-relevance improves memory for visual details, then the participants adopting a specific viewing perspective should make fewer recognition errors for relevant than irrelevant lures. On the other hand, if the viewing perspective encourages gist-based memory processing, then participants may actually show increased recognition errors for relevant in comparison to irrelevant objects.

Method

Participants
Thirty-three University of Turku students (13 males) participated in the experiment.

Apparatus
Eye movements were recorded with a desktop-mounted Eyelink® 1000 (SR Research Ltd., Ontario, Canada) system. Sampling frequency was 1,000 Hz. The stimuli were presented on a 21” CRT screen with a screen resolution of 1,024 × 768 pixels and a 150-Hz refresh rate. Participants were seated 70 cm from the screen, and a chin rest was used to stabilize the head.

Materials
Twenty-two colour photographs presenting three different houses and their interiors were used as materials (an example is presented in Figure 1; all materials are available at http://users.utu.fi/johkaa/Materials/picture_set.zip; Kaakinen, 2010a). A separate rating study was conducted to identify areas of interest to be used in the analyses. A panel of 6 participants (who did not participate in the actual experiment) were shown the pictures on a computer
screen; the pictures were also printed on a leaflet that was given to the participants. Participants were asked to imagine that they were going to buy a house and to encircle the pictures areas relevant to the homebuyer. After the first viewing, participants were instructed to change the perspective and to imagine that they were burglars. They were then shown the pictures again and were asked to encircle areas that would be interesting to a burglar. Only areas that were uniformly identified as interesting to either perspective by most of the raters were selected to be used in the analyses. In some cases, the same area was identified as interesting to both burglars and homebuyers—however, only interest areas that most of the raters found relevant to only one perspective were selected for the analyses. In other words, none of the selected interest areas was relevant to both the burglar’s and homebuyer’s perspective. There were 23 burglar-relevant and 26 homebuyer-relevant targets. Burglar-relevant target areas contained mainly valuable and easily portable objects, such as a painting, valuable decorative objects, a cell phone, a handbag, a TV, and jewellery. Homebuyer-relevant areas included details of the décor—for example, wallpaper, wooden doors, discoloured toilet seat, cupboards, a fireplace, and wooden support structures in the ceiling. Homebuyer-relevant areas thus included more background-type areas whereas burglar-relevant areas contained mainly objects. As a result, homebuyer-relevant target areas were larger (mean size 131,274 pixels) than burglar-relevant target areas (39,656 pixels).

Next, we computed the five most salient locations in each picture using the iLab C++ Neuromorphic Vision Toolkit (University of Southern California, iLab, & Itti, 2002–2004; see http://ilab.usc.edu/toolkit/home.shtml). If one or more of the salient locations fell within the predefined homebuyer- or burglar-relevant target areas, these targets were labelled as salient. There were 8 salient homebuyer-relevant and 12 salient burglar-relevant targets. In sum, there were 49 target areas: 18 homebuyer-relevant nonsalient targets, 8 homebuyer-relevant salient targets, 11 burglar-relevant nonsalient targets, and 12 burglar-relevant salient targets. Examples of target areas are presented in Figure 1.

The recognition memory task consisted of 60 pictures: 10 burglar-relevant and 10 homebuyer-relevant objects taken from the original scenes (i.e., “old” objects), 10 burglar-relevant and 10 homebuyer-relevant lures, and 20 new objects depicting objects not presented in the pictures (examples are presented in Figure 1; pictures are available at http://users.utu.fi/johkkaa/Materials/recognition_set.zip; Kaakinen, 2010b). Six of the 10 burglar-relevant old objects and 4 of the 10 homebuyer-relevant old objects were salient (however, due to the small number of items, data were collapsed across salient and nonsalient items in the analyses). The lures were created by selecting for each old object a picture of another object that belonged to the same category as the old object.

Procedure
Participants were randomly assigned to three perspective groups: burglars ($n = 11$), homebuyers ($n = 11$), and memorizers ($n = 11$). Participants in the burglar and homebuyer groups were instructed to imagine that they were either burglars or homebuyers and told to view the pictures so that they could decide which house they would break into (burglar perspective) or buy (homebuyer perspective; Anderson & Pichert, 1978). Participants in the memorizer group were told to view the pictures in preparation for a memory test.

The photographs were presented in a fixed sequence. Pictures from the same house were presented together so that the first picture in the “house set” depicted the exterior of the house, followed by pictures of the interiors. Before each picture, a drift correction dot was presented in the middle of the screen, and participants were instructed to fixate it. Each picture was presented for 6 s, followed by a black screen presented for 500 ms.

After picture viewing, participants completed a visuospatial working memory task as a memory distractor task and then provided a free recall of the scene contents. The participants were instructed to recall everything they could recall of the pictures and not to restrict themselves to the perspective they had in mind during viewing.
On the recall sheet there were three columns labelled “what?”, “where?”, and “what kind?”. Participants were told to describe what they had seen and, if they could, where (i.e., in which one of the three houses) it had been located and to describe it in more detail (colour and/or size).

Finally, a recognition test was administered. In the test, pictures of objects were presented one at a time on a computer screen. The participant’s task was to identify whether she or he had seen the object in the scenes or not. The presentation order of pictures was randomized.

The experimental session lasted for about 45 min.

**Results**

**Data trimming and statistical analyses**

Only fixations that occurred during the presentation of the stimulus pictures were analysed; fixations spanning across the screen change were removed from the data. Fixations shorter than 50 ms were either merged with a nearby fixation (if the distance between the fixations was < 1 degrees) or removed from the data.

Total fixation time, number of visits, first-run dwell time, and the probability of fixating the target area were computed to obtain a comprehensive picture of the time-course of picture processing. Total fixation time reflects the summed duration of fixations landing on a target area. Number of visits to the target area indicates how many times viewers entered the target area—that is, it breaks down the number of visits that the total fixation time consists of. The first-run dwell time is the summed duration of the fixations that viewers made during the first visit to the target area, and it reflects the initial processing of targets. The time-course of picture viewing was further examined by computing the probability of fixating a target area as a function of fixation order (10 first fixations of each trial were analysed). For the sake of clarity, in this analysis the data were collapsed across the two viewing perspectives, and separate measures were computed for perspective-relevant and perspective-irrelevant target areas.

**Total fixation time**

The means and standard errors for total fixation time\(^1\) are presented in Table 1. The data were analysed with a 3 (perspective) × 2 (saliency) × 2 (target area) mixed analysis of variance (ANOVA). A significant main effect of saliency indicates that salient targets attracted longer total fixation times than nonsalient targets, \(F(1, 30) = 85.33, \ p < .001, \ \eta_p^2 = .74\). A Perspective × Target Area interaction indicates that the total fixation time on burglar-relevant and homebuyer-relevant target areas varied as a function of viewing perspective, \(F(2, 30) = 16.10, \ p < .001, \ \eta_p^2 = .52\). In order to examine the interaction in more detail, we analysed the data with repeated measures ANOVAs separately for each viewing perspective group.

For participants who adopted the memory encoding perspective, only the main effect of saliency was significant, \(F(1, 10) = 110.20, \ p < .001, \ \eta_p^2 = .92\). In other words, when the goal was to memorize the picture contents, salient targets received more attention than nonsalient objects. As for the participants with the burglar perspective, a significant main effect of saliency was obtained, \(F(1, 10) = 19.87, \ p = .001, \ \eta_p^2 = .67\), indicating that participants who had adopted the burglar perspective spent longer time on salient than on nonsalient targets. A significant main effect of target area, \(F(1, 10) = 13.47, \ p = .004, \ \eta_p^2 = .57\), showed that they also viewed burglar-relevant targets for longer time than homebuyer-relevant targets. Also homebuyers showed significant effects of saliency, \(F(1, 10) = 13.74, \ p = .004, \ \eta_p^2 = .58, \) and target area, \(F(1, 10) = 14.36, \ p = .004, \ \eta_p^2 = .59\). Thus, participants with the homebuyer perspective spent a longer time on salient than on nonsalient targets and also a longer time on homebuyer-relevant than on burglar-relevant targets.

\(^1\) We also analysed the total number of fixations. The analyses replicated the results for total fixation time; for the sake of brevity, they are not reported here.
Saliency \times Target Area interaction, $F(1, 10) = 10.10, p = .01, \eta_p^2 = .50$, indicating that for the homebuyer group, the effect of saliency was greater for perspective-irrelevant than for perspective-relevant target areas (i.e., for burglar-relevant targets).

But does a specific viewing perspective (burglar or homebuyer) increase the viewing time on perspective-relevant targets, decrease the viewing time on perspective-irrelevant targets, or both? In order to answer this question, we computed total fixation times for perspective-relevant and perspective-irrelevant target areas across the burglar and homebuyer groups and compared these means to the mean of the memory encoding group (baseline). The viewing perspective increased total viewing time on perspective-relevant target areas and decreased the total viewing time on perspective-irrelevant target areas: greatest $p$ value, $t(26) = -2.43, p = .022$, for the difference between the means for the salient relevant target areas and the baseline.

### Number of visits
The mean number of visits and standard errors of the means are presented in Table 1. The data were analysed with a $3 \times 2 \times 2$ mixed ANOVA. Salient target areas received more visits than nonsalient target areas, as indicated by a significant main effect of saliency, $F(1, 30) = 67.47, p < .001, \eta_p^2 = .69$. A two-way interaction between target area and saliency indicates that the effect of saliency was greater for the burglar-relevant than for the homebuyer-relevant target areas, $F(1, 30) = 9.49, p = .004, \eta_p^2 = .24$. A Target Area $\times$ Perspective interaction indicates that the number of visits on different target areas was dependent on the viewing perspective. Moreover, there was a significant three-way interaction between target area, saliency, and viewing perspective, $F(2, 30) = 5.88, p = .007, \eta_p^2 = .28$. The interactions were further examined with separate ANOVAs for each perspective group.

Participants who had adopted the memory encoding perspective showed a main effect of saliency, $F(1, 10) = 71.11, p < .001, \eta_p^2 = .88$. In other words, the memorizers visited salient targets more often than nonsalient targets. There was a two-way interaction between target area and saliency, $F(1, 10) = 17.43, p = .002, \eta_p^2 = .64$, indicating that the effect of saliency was more pronounced for burglar-relevant than for homebuyer-relevant target areas. Burglars visited more often salient than nonsalient targets, $F(1, 10) = 14.88, p = .003, \eta_p^2 = .60$. Moreover, they

<table>
<thead>
<tr>
<th>Measure</th>
<th>Perspective</th>
<th>Salient</th>
<th>Nonsalient</th>
<th>Salient</th>
<th>Nonsalient</th>
<th>Salient</th>
<th>Nonsalient</th>
<th>Salient</th>
<th>Nonsalient</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total fixation time (ms)</td>
<td>Homebuyer</td>
<td>894</td>
<td>607</td>
<td>778</td>
<td>314</td>
<td>607</td>
<td>85</td>
<td>778</td>
<td>314</td>
</tr>
<tr>
<td></td>
<td>Burglar</td>
<td>632</td>
<td>1,199</td>
<td>333</td>
<td>910</td>
<td>632</td>
<td>1,199</td>
<td>333</td>
<td>910</td>
</tr>
<tr>
<td></td>
<td>Memory</td>
<td>776</td>
<td>937</td>
<td>454</td>
<td>548</td>
<td>776</td>
<td>937</td>
<td>454</td>
<td>548</td>
</tr>
<tr>
<td>Number of visits</td>
<td>Homebuyer</td>
<td>1.83</td>
<td>1.47</td>
<td>1.68</td>
<td>1.40</td>
<td>1.83</td>
<td>1.47</td>
<td>1.68</td>
<td>1.40</td>
</tr>
<tr>
<td></td>
<td>Burglar</td>
<td>1.42</td>
<td>1.71</td>
<td>0.94</td>
<td>1.39</td>
<td>1.42</td>
<td>1.71</td>
<td>0.94</td>
<td>1.39</td>
</tr>
<tr>
<td></td>
<td>Memory</td>
<td>1.37</td>
<td>1.60</td>
<td>1.14</td>
<td>0.95</td>
<td>1.37</td>
<td>1.60</td>
<td>1.14</td>
<td>0.95</td>
</tr>
<tr>
<td>First-run dwell time (ms)</td>
<td>Homebuyer</td>
<td>476</td>
<td>361</td>
<td>491</td>
<td>436</td>
<td>476</td>
<td>361</td>
<td>491</td>
<td>436</td>
</tr>
<tr>
<td></td>
<td>Burglar</td>
<td>440</td>
<td>661</td>
<td>344</td>
<td>627</td>
<td>440</td>
<td>661</td>
<td>344</td>
<td>627</td>
</tr>
<tr>
<td></td>
<td>Memory</td>
<td>419</td>
<td>379</td>
<td>326</td>
<td>379</td>
<td>419</td>
<td>379</td>
<td>326</td>
<td>379</td>
</tr>
</tbody>
</table>

Table 1. Means and standard errors of the means for eye fixation measures

<table>
<thead>
<tr>
<th>Area of interest</th>
<th>Salient</th>
<th>Nonsalient</th>
<th>Salient</th>
<th>Nonsalient</th>
</tr>
</thead>
<tbody>
<tr>
<td>Measure</td>
<td>Perspective</td>
<td>M</td>
<td>SE</td>
<td>M</td>
</tr>
<tr>
<td>Total fixation time (ms)</td>
<td>Homebuyer</td>
<td>894</td>
<td>60</td>
<td>607</td>
</tr>
<tr>
<td>Burglar</td>
<td>632</td>
<td>62</td>
<td>1,199</td>
<td>121</td>
</tr>
<tr>
<td>Memory</td>
<td>776</td>
<td>24</td>
<td>937</td>
<td>55</td>
</tr>
<tr>
<td>Number of visits</td>
<td>Homebuyer</td>
<td>1.83</td>
<td>0.09</td>
<td>1.47</td>
</tr>
<tr>
<td>Burglar</td>
<td>1.42</td>
<td>0.14</td>
<td>1.71</td>
<td>0.11</td>
</tr>
<tr>
<td>Memory</td>
<td>1.37</td>
<td>0.07</td>
<td>1.60</td>
<td>0.06</td>
</tr>
<tr>
<td>First-run dwell time (ms)</td>
<td>Homebuyer</td>
<td>476</td>
<td>45</td>
<td>361</td>
</tr>
<tr>
<td>Burglar</td>
<td>440</td>
<td>45</td>
<td>661</td>
<td>74</td>
</tr>
<tr>
<td>Memory</td>
<td>419</td>
<td>24</td>
<td>379</td>
<td>21</td>
</tr>
</tbody>
</table>
made more visits to burglar-relevant than to burglar-irrelevant target areas, $F(1, 10) = 8.74, p = .014, \eta_p^2 = .47$. As for the homebuyers, they made more visits to salient than to nonsalient target areas, $F(1, 10) = 18.30, p = .002, \eta_p^2 = .65$. They also visited more often homebuyer-relevant than burglar-relevant targets, $F(1, 10) = 29.93, p < .001, \eta_p^2 = .75$. A two-way interaction between saliency and target area indicates that in this group the effect of saliency was more pronounced for burglar-relevant than for homebuyer-relevant target areas, $F(1, 10) = 19.43, p = .001, \eta_p^2 = .66$.

Further analyses showed that a specific viewing perspective increased the number of visits to the relevant target areas: greatest $p$, $t(28) = -3.75, p = .001$, for the difference between the salient relevant targets and baseline. A specific viewing perspective also decreased the number of visits made to the nonsalient irrelevant targets, $t(31) = 2.23, p = .033$, but not to salient irrelevant targets, $t(25) = 0.40, p = .69$.

**First-run dwell time**

The means and standard errors of first-run dwell times are presented in Table 1. The analysis of the first-run dwell times yielded somewhat different results than the other measures. The main effect of saliency was not significant, $F < 2$. A two-way interaction between saliency and target area, $F(1, 30) = 4.70, p = .038, \eta_p^2 = .14$, suggests that only homebuyer-relevant targets showed the saliency effect—for burglar-relevant targets, there was practically no difference between the salient and nonsalient targets. A Perspective × Target Area interaction showed that the first-pass dwell time on burglar-relevant and homebuyer-relevant target areas depended on the viewing perspective, $F(2, 30) = 14.01, p < .001, \eta_p^2 = .48$. Moreover, a significant main effect of target area, $F(1, 30) = 4.66, p = .039, \eta_p^2 = .14$, was obtained, indicating that burglar-relevant target areas attracted longer first-pass dwell times than did homebuyer-relevant target areas.

In order to examine the Perspective × Target Area interaction in more detail, we analysed the data with repeated measures ANOVAs separately for each viewing perspective group. For participants who adopted the memory encoding perspective, the main effect of saliency was marginally significant, $F(1, 10) = 4.37, p = .063, \eta_p^2 = .30$. A marginal interaction between target area and saliency, $F(1, 10) = 4.57, p = .058, \eta_p^2 = .31$, indicated that the effect of saliency was only observed for homebuyer-relevant target areas—for burglar-relevant areas there was practically no difference between the first-run dwell times on salient and nonsalient targets. As for the participants with the burglar perspective, a significant main effect of target area indicated that the first-run dwell times were longer on burglar-relevant than on homebuyer-relevant target areas, $F(1, 10) = 20.64, p = .001, \eta_p^2 = .67$. The main effect of saliency did not reach significance, $F(1, 10) = 2.69, p = .13, \eta_p^2 = .21$. As for homebuyers, none of the effects reached significance, even though the mean first-pass dwell times were somewhat longer for the homebuyer-relevant than for the burglar-relevant target areas, $F(1, 10) = 2.36, p = .155, \eta_p^2 = .19$.

The further analyses comparing first-pass dwell times on relevant and irrelevant targets to the baseline showed that a specific viewing perspective increased the first-pass dwell times on perspective-relevant targets: largest $p$ observed, $t(26) = -3.38, p = .002$, for the difference between salient relevant targets and baseline.

**Probability of fixation**

Next, we analysed the time course of fixations by computing the probability of fixating a target as a function of the order of fixation during the trial. First 10 fixations were analysed. For the sake of clarity, we collapsed the data across the burglar and homebuyer target areas and the burglar and homebuyer groups to compute measures for perspective-relevant and perspective-irrelevant target areas. The data from the memorizer group (collapsed across the homebuyer and burglar target areas) served as a baseline. The means and standard errors are presented in Figure 2.

We first analysed the data of the two specific perspective groups (burglars and homebuyers)
with a 2 (relevance: relevant vs. irrelevant) × 2 (saliency: nonsalient vs. salient) × 10 (fixation order) repeated measures ANOVA. A significant main effect of relevance indicates that relevant target areas were more likely to be fixated than irrelevant target areas, $F(1, 21) = 38.66$, $p < .001$, $\eta^2_p = .65$. A two-way interaction between relevance and fixation order indicates that the magnitude of the relevance effect depended on the order of fixation, $F(9, 189) = 2.69$, $p = .006$, $\eta^2_p = .11$. Moreover, a significant main effect of saliency shows that salient targets were more likely to be fixated than nonsalient targets, $F(1, 21) = 31.74$, $p < .001$, $\eta^2_p = .60$. However, a Salience × Fixation Order interaction implies that the effect of saliency varied as a function of fixation order, $F(1, 21) = 3.49$, $p = .001$, $\eta^2_p = .14$. The interactions were further examined by conducting separate analyses for each fixation.

For the probability of the first fixation made during the presentation of the picture (note that the fixations that spanned across the screen change were removed), the only significant main effect was that of saliency, $F(1, 21) = 48.20$, $p < .001$, $\eta^2_p = .70$, indicating that the first fixation was more likely to land on a salient than on a non-salient target area, irrespective of its relevance. For the second fixation, however, both the main effect of saliency, $F(1, 21) = 4.81$, $p = .04$, $\eta^2_p = .19$, and that of relevance, $F(1, 21) = 17.14$, $p < .001$, $\eta^2_p = .45$, were obtained. In other words, already the second fixation was more likely to land on a relevant than on an irrelevant target (and salient targets continued to attract fixations with a greater likelihood than nonsalient targets). The main effect of relevance was significant for all subsequent fixations: smallest $F(1, 21) = 9.99$, $p = .005$, $\eta^2_p = .32$, for the seventh fixation. The main effect of saliency failed to reach significance for the fourth, fifth, and the sixth fixation, but it was significant for all other fixations: smallest $F(1, 21) = 9.60$, $p = .005$, $\eta^2_p = .31$, for the seventh fixation.

Next, the data for the memory instruction group were examined with a 2 (target area) × 2 (saliency) × 10 (fixation order) repeated measures ANOVA. Only the main effect of saliency was significant, $F(1, 10) = 39.46$, $p < .001$, $\eta^2_p = .80$. All other effects and interactions failed to reach significance (all $F$s < 2). In other words, when the participants were viewing the pictures with the intention to remember the contents afterwards, visually salient locations were more likely to attract a fixation than nonsalient locations.

Recall

Participants’ recall protocols were coded for correctly recalled targets. If the participant mentioned the target object and placed it to the correct house, she or he was given 1 point (e.g., *there was jewellery*). If the participant mentioned the target object, its correct location (i.e., the first, second, or third house) and some details (e.g., *there was jewellery on top of the laundry basket in the bathroom*), she or he was given 2 points. Two raters first coded together 10 protocols to reach agreement on the coding criteria. Then they independently coded 10 more protocols. Because the interrater agreement was good (97%), the rest of
the protocols were coded by one rater. A mean recall score was computed as the ratio of the participant’s total recall score and the number of targets (there were 26 buyer-relevant and 23 burglar-relevant targets).

The means and their standard errors are presented in Table 2. The data were analysed with a 3 (perspective) × 2 (saliency) × 2 (target area) mixed ANOVA. A main effect of viewing perspective, \( F(2, 30) = 3.21, p = .054, \eta^2_p = .18 \), suggests that the recall rates depended on the adopted perspective. Looking at the means, it can be seen that the memory encoding group did not gain superior memory for scene contents—rather, their memory performance for the predefined targets was typically somewhere between the two perspective groups. Moreover, there was a significant main effect of target area, \( F(1, 30) = 16.94, p < .001, \eta^2_p = .36 \), indicating that burglar-relevant targets were overall better recalled than homebuyer-relevant targets. However, the two main effects were qualified by a significant interaction between target area and viewing perspective, \( F(1, 30) = 15.62, p < .001, \eta^2_p = .51 \), indicating that the recall of the burglar- and homebuyer-relevant information depended on the viewing perspective. The interaction was examined by separate ANOVAs for the perspective groups.

The memory encoding group showed a better recall of salient than nonsalient targets, \( F(1, 10) = 4.77, p = .054, \eta^2_p = .32 \). They also recalled more burglar-relevant than homebuyer-relevant details, \( F(1, 10) = 8.62, p = .015, \eta^2_p = .46 \).

Burglars recalled more burglar-relevant than homebuyer-relevant details, \( F(1, 10) = 25.73, p < .001, \eta^2_p = .72 \). Homebuyers, on the other hand, showed slightly better memory for the homebuyer-relevant than for the burglar-relevant details, but this difference failed to reach significance, \( F(1, 10) = 2.68, p = .133, \eta^2_p = .21 \).

Comparisons between the recall performance for relevant and irrelevant targets and the baseline showed that a specific reading perspective decreased the recall of irrelevant information: largest \( p \) value, \( t(31) = 2.24, p = .033 \), for the difference between nonsalient irrelevant targets and baseline. The differences between the recall performance for relevant targets and baseline failed to reach significance: smallest \( p \) value, \( t(31) = -1.78, p = .085 \), for the difference between nonsalient relevant targets and the baseline.

The recall results thus suggest that memory for scene contents is restricted by the adopted viewing perspective. When the viewers were preparing for a memory test, their memory was better for the salient than for the nonsalient scene objects—these objects were also more likely to be fixated and gazed at for a longer time during scene viewing. On the other hand, when the viewers had a specific viewing perspective in mind, the

<table>
<thead>
<tr>
<th>Measure</th>
<th>Perspective</th>
<th>Salient</th>
<th></th>
<th>Nonsalient</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Homebuyer-relevant</td>
<td>Burglar-relevant</td>
<td>Homebuyer-relevant</td>
<td>Burglar-relevant</td>
</tr>
<tr>
<td>Recall</td>
<td>Homebuyer</td>
<td>.19</td>
<td>.06</td>
<td>.11</td>
<td>.04</td>
</tr>
<tr>
<td></td>
<td>Burglar</td>
<td>.09</td>
<td>.04</td>
<td>.36</td>
<td>.07</td>
</tr>
<tr>
<td></td>
<td>Memory</td>
<td>.18</td>
<td>.04</td>
<td>.25</td>
<td>.07</td>
</tr>
<tr>
<td>Recognition accuracy</td>
<td>Homebuyer</td>
<td>.77</td>
<td>.05</td>
<td>.50</td>
<td>.07</td>
</tr>
<tr>
<td></td>
<td>Burglar</td>
<td>.68</td>
<td>.07</td>
<td>.68</td>
<td>.05</td>
</tr>
<tr>
<td></td>
<td>Memory</td>
<td>.73</td>
<td>.08</td>
<td>.52</td>
<td>.09</td>
</tr>
</tbody>
</table>

Table 2. Means and their standard errors for the recall and recognition tasks
memory bias for salient objects disappeared. Instead, memory performance was guided by the viewing perspective: Memory was overall better for perspective-relevant than for perspective-irrelevant scene objects. The bias in recall reflected decreased memory performance for perspective-irrelevant scene information.

**Target recognition**

The recognition rates (percentage correct) are presented in Table 2. Most importantly, there were significant interactions between target and viewing perspective, $F(2, 30) = 10.36, p < .001$, $\eta^2_p = .41$, and target, saliency, and viewing perspective, $F(2, 30) = 3.46, p = .044$, $\eta^2_p = .19$, indicating that the recognition rates for different targets depended on the participant’s viewing perspective. Moreover, there was a significant main effect of target, indicating that homebuyer-relevant targets were recognized overall more accurately than burglar-relevant targets, $F(1, 30) = 6.03, p = .02$, $\eta^2_p = .17$, and a two-way interaction between target and saliency, $F(1, 30) = 4.36, p = .05$, $\eta^2_p = .13$, suggesting that saliency played a different role for homebuyer-relevant and burglar-relevant targets. The interactions involving the viewing perspective were further examined with separate ANOVAs for each perspective group.

For the memory encoding group, there were no significant main effects or interactions. One-sample $t$ tests indicated that the memorizers recognized burglar-relevant targets only at the chance level, $t < 2$, while homebuyer-relevant targets were recognized above chance, smallest $t(10) = 2.47, p = .033$, for nonsalient targets. For the homebuyers, there was a significant main effect of target, $F(1, 10) = 15.41, p = .003$, $\eta^2_p = .61$, indicating that for this group the recognition rates were higher for homebuyer-relevant than for burglar-relevant targets. The recognition rates for burglar-relevant targets were at the chance level, $t < 2$, whereas homebuyer-relevant targets were recognized significantly better than chance, smallest $t(10) = 4.95, p = .001$, for nonsalient targets. As for the burglars, the main effect of target was marginally significant, $F(1, 10) = 3.75, p = .082$, $\eta^2_p = .27$, indicating that the pretend burglars showed somewhat better recognition rates for burglar-relevant than for homebuyer-relevant targets. A significant interaction between target and saliency was obtained, $F(1, 10) = 10.76, p = .008$, $\eta^2_p = .52$. This interaction reflects the fact that there was a difference between the recognition rates for burglar-relevant and homebuyer-relevant targets only for the nonsalient targets. For salient targets, the mean recognition rates were practically identical. In this group, recognition rates were at the chance level for nonsalient homebuyer-relevant targets, $t < 1$, and above chance for other types of targets, smallest $t(10) = 2.67, p = .024$, for salient homebuyer-relevant targets.

In sum, the viewing perspective had an influence on the recognition of burglar-related and homebuyer-related targets: The data for the perspective groups showed that the recognition performance was better for perspective-relevant than for perspective-irrelevant targets. Recognition performance was only at the chance level for perspective-irrelevant targets (except for the salient irrelevant targets for the burglar group).

**Recognition errors for lures**

One participant had extremely high error rates (100% for both types of lures). Thus, this participant was left out from the recognition error analyses. For the remaining participants ($N = 32$), the mean false-recognition rates were .40 ($SD = .23$) for the homebuyer-relevant and .28 ($SD = .18$) for the burglar-relevant lures, respectively. There was a lot of individual variability in the error rates; the rates varied between .00 and .80 for the homebuyer-relevant and .00 to .60 for the burglar-relevant lures, respectively. The mean error rates and their standard errors are presented in Table 3.

The data were analysed with a 3 (perspective) $\times$ 2 (picture type) ANOVA. Only the main effect of picture type was significant, $F(1, 29) = 10.42, p = .003$, $\eta^2_p = .26$, indicating that homebuyer-relevant lures produced more recognition errors than burglar-relevant lures. One-sample $t$ tests indicated that participants made fewer
recognition errors for lures than would be expected by chance, $t(31) = -2.56$, $p = .016$, for homebuyer-relevant, and $t(31) = -6.86$, $p < .001$, for burglar-relevant lures, respectively. In other words, participants had overall relatively accurate memory for the targets.

**Discussion**

The results of the present experiment showed that both visual saliency and viewing perspective play a role in eye movement guidance during scene viewing: The initial fixation on the scene was more likely to land on a salient than on a nonsalient area, but already the second fixation after the scene onset was more likely to land on a perspective-relevant than on a perspective-irrelevant area. The effect of viewing perspective was observed in several different eye movement measures: Perspective-relevant areas were gazed at more often and for a longer time than perspective-irrelevant areas, and first-pass dwell times were longer on relevant than on irrelevant target areas. Memory for scenes was restricted by the viewing perspective: Participants recalled more perspective-relevant than perspective-irrelevant scene information.

Comparisons between the perspective groups (burglars and homebuyers) and the baseline group (memorizers) showed that a specific viewing perspective increased the first-pass dwell times, the number of visits, and the total time spent on perspective-relevant targets and decreased the number of visits and the total viewing time devoted to perspective-irrelevant targets. Also recall performance for the perspective-relevant scene information was enhanced by the specific viewing perspective.

The recognition data showed that participants are overall better at recognizing perspective-relevant than perspective-irrelevant targets. Moreover, the analysis of the recognition errors for lures showed that participants had a relatively good memory for the visual details of the scene objects. In other words, increased memory for perspective-relevant scene information does not seem to be a consequence of gist-based memory encoding. However, it should be noted that there were individual differences in the error rates, which may, for example, be a result of individual differences in memory encoding strategies.

As noted above, perspective relevance increased the likelihood of looking at scene areas containing information relevant to the assigned perspective, leaving less time to look at other areas within the scene, as indexed by decreased total fixation time on perspective-irrelevant scene areas. When a perspective-relevant area within the scene was fixated for the first time, the eyes stayed there longer than they did in the baseline condition in which no target area was specifically “relevant” or “irrelevant”. These findings are in line with the classical study of Yarbus (1967; see also Benson et al., 2009; DeAngelus & Pelz, 2009), who showed that the viewing perspective guides the eye gaze to scene areas that are informative with respect to the viewing perspective. What our study adds to these previous demonstrations of the importance of a viewing perspective is that the perspective effects can be observed very early on after the scene onset. Already the second fixation after scene onset is more likely to land on a perspective-relevant than on a perspective-irrelevant area within the scene.

Moreover, our results showed that even though the viewing perspective plays an important role in scene viewing, also visual saliency (e.g., Navalpakkam & Itti, 2005) predicts where the viewers are likely to look at: Salient scene areas attract more visits and longer total fixation time than nonsalient areas. Especially the initial fixation...
on a scene is directed to a visually salient area. The impact of saliency, however, remains significant also during the subsequent fixations over and above the pervasive effects of the adopted viewing perspective. These results are in line with theoretical views that incorporate both bottom-up and top-down processes in models of eye guidance during scene viewing (e.g., Navalpakkam & Itti, 2005). It is likely that the task the viewer has in mind defines to what degree bottom-up versus top-down information is useful in performing the task, and the role of these factors in scene viewing varies accordingly. For example, visual search may be more efficient if the viewer makes use of top-down information rather than processes the scene in a bottom-up manner. When a scene is viewed from a certain psychological perspective, however, low-level visual saliency information may be useful in efficiently guiding the eyes to relevant locations. Previous research suggests that informative areas within the scene tend to have high saliency values and also attract more fixation time (Henderson et al., 2007; Masciocchi et al., 2009). A study by Nyström and Holmqvist (2008) suggests that the influence of low-level factors (i.e., contrast, which is one aspect of visual saliency) on the distribution of eye fixations on the scene depends on the semantic content of the picture. Applied to our study, this means that in order to efficiently evaluate the house scenes in the light of the adopted perspective, it pays to direct the eyes to salient scene areas that are likely to carry the most semantic content. This may explain why we found an effect of saliency throughout the scene viewing, in addition to the pervasive perspective effects.

The present results also showed that the memory for scenes is restricted by the viewing perspective: Recall rate and recognition performance were better for perspective-relevant than for perspective-irrelevant targets. The increased memory for perspective-relevant targets was gained at the expense of perspective-irrelevant targets: A viewing perspective decreased the memory for irrelevant targets in comparison to the baseline condition. Also visual saliency had an influence on memory performance: Salient scene objects were remembered better than non-salient objects. These memory effects are easily accounted for by the corresponding effects obtained for scene viewing: Targets that attracted more fixations and longer fixation times were remembered better than targets that attracted fewer fixations and shorter fixation times. These results are in line with views assuming that memory for scenes accumulates across time and fixations (e.g., Hollingworth & Henderson, 2002; Melcher, 2006; Tatler et al., 2005; Tatler et al., 2003). According to these views, scene information is encoded to long-term memory (LTM)—the longer the viewing time, the more information is consolidated into LTM. For example, according to the model proposed by Hollingworth and Henderson (2002), information about the scene objects are coded into object files, which retain abstracted visual and conceptual information of the scene objects and their spatial locations.

Our results suggest that a viewing perspective does not increase memory for perspective-relevant information due to gist-based memory encoding. Instead, viewers retain in LTM relatively accurate visual representations of the scene objects. In our picture recognition task akin to the false memory paradigm (Koutstaal & Schacter, 1997), participants were relatively good (better than by chance) at rejecting both perspective-relevant and perspective-irrelevant lures (pictures of objects belonging to the same category as the original targets). These results are in line with previous research showing that visual memory is relatively robust (e.g., Hollingworth, 2005). According to Hollingworth (2005), there are three stages of encoding visual scenes to memory: (a) sensory persistence, (b) visual short term memory (VSTM), and (c) visual LTM (VLTM). Information in VLTM is stored in an abstracted format (not exact perceptual experiences); however, this format is supposedly sufficient to retain detailed visual information about the scene objects. The VLTM is stable (not transient), and forgetting is gradual. Moreover, there are no capacity limitations in VLTM. Thus, memory for scene objects should remain relatively good even after
longer periods of time, even when there is a wealth of information to be remembered. Our findings are generally in line with this view.

However, exactly how accurate the visual representation constructed during perspective-driven scene viewing is remains to be solved in future experiments. For example, an interesting question is whether viewers accurately encode information about the exact or approximate spatial location and orientation of the perspective-relevant versus perspective-irrelevant objects (Tatler et al., 2005). Moreover, as noted above, there was substantial individual variability in the susceptibility to false recognition of lures. It is possible that different people use different memory encoding strategies when given a viewing perspective: Perhaps some viewers use a gist-based memory encoding, whereas others try to store scene information as perceptually accurately as possible. Future experiments should examine individual differences in scene memory in more detail.

In sum, the present experiment demonstrates that the perspective the person adopts plays a significant role in how scenes are inspected (see also Benson et al., 2009; DeAngelus & Pelz, 2009; Yarbus, 1967). The novel aspect of the present study is that in addition to examining the perspective effects, we also examined the role of bottom-up saliency. In particular, we examined how these effects evolved across the initial fixations on the scene. Our results indicate that even though the viewing perspective plays an important role in where the viewers look, also visual saliency guides the eyes, especially during the initial fixations on the scene. Our findings are in line with models that incorporate both bottom-up and top-down effects. Navalpakkam and Itti (2005) presented a general framework for how a saliency map and a task-based relevance map could jointly guide visual attention within a scene. More recently, Peters and Itti (2007) presented a computational model that incorporates both bottom-up saliency and task-based importance to predict the allocation of eye gaze within a scene. Their results showed that a model combining saliency and task-based importance did a better job in predicting the allocation of eye gaze during a task than did a saliency model alone (but see Vincent, Troschianko, & Gilchrist, 2007, for a critique). Our results indicate that the importance of saliency versus task relevance may change across time: Initially saliency is more important, but as the viewer makes more fixations on the scene, also perspective relevance gains a significant role in guiding the eyes. These results support the suggestion of van Zoest, Hunt, and Kingstone (2010), who recently argued that examining the detailed time-course of visual processing is crucial in understanding the role of different stimulus properties and participant goals in visual cognition.

REFERENCES


Current Directions in Psychological Science, 19, 116–120.


