
Rapid detection of person information in a naturalistic scene

Sue Fletcher-Watson[¶], John M Findlay, Susan R Leekam, Valerie Benson

Department of Psychology, University of Durham, Science Site, South Road, Durham DH1 3LE, UK
Received 17 October 2006, in revised form 4 May 2007; published online 26 March 2008

Abstract. A preferential-looking paradigm was used to investigate how gaze is distributed in naturalistic scenes. Two scenes were presented side by side: one contained a single person (person-present) and one did not (person-absent). Eye movements were recorded, the principal measures being the time spent looking at each region of the scenes, and the latency and location of the first fixation within each trial. We studied gaze patterns during free viewing, and also in a task requiring gender discrimination of the human figure depicted. Results indicated a strong bias towards looking to the person-present scene. This bias was present on the first fixation after image presentation, confirming previous findings of ultra-rapid processing of complex information. Faces attracted disproportionately many fixations, the preference emerging in the first fixation and becoming stronger in the following ones. These biases were exaggerated in the gender-discrimination task. A tendency to look at the object being fixated by the person in the scene was shown to be strongest at a slightly later point in the gaze sequence. We conclude that human bodies and faces are subject to special perceptual processing when presented as part of a naturalistic scene.

1 Introduction

When viewing a natural scene or a pictorial depiction of such a scene, individuals make a series of saccadic eye movements to direct attention successively at different details (Henderson 2003). In an early series of studies, Buswell (1935) noted that, in scenes containing human figures, these figures were disproportionately likely to be fixated. This finding has been replicated consistently, although the exact proportion may be modulated by cognitive factors such as task instructions (Yarbus 1967). It has not yet been established at what point in the viewing sequence the preference for human figures emerges. The traditional view (eg Henderson and Hollingworth 1999) suggests that very early fixations on a scene are determined by low-level visual factors, such as contrast, and are not influenced by semantic characteristics of the scene content. Prior to the direction of gaze to a specific location, some aspect of the stimulation from this location must have been analysed in peripheral vision to allow the selection processes to operate. It is well known that simple visual features, such as high-contrast borders, can be detected rapidly in peripheral vision so that, for example, they can be used to direct the first saccade in a visual-search task (Findlay 1997).

However, there is now considerable evidence suggesting that more complex visual stimuli may be processed rapidly, even in the visual periphery (eg Thorpe et al 2001). In a recent study (Kirchner et al 2003; Kirchner and Thorpe 2006) participants were presented with two images of natural scenes simultaneously, one containing a non-human animal and one not. Their task required participants to make a saccade as rapidly as possible to the scene containing an animal. Participants were not only capable of achieving around 90% accuracy, but also produced correct responses significantly above chance when the saccade was made as little as 120 ms after stimulus onset. This research indicates the capacity for attentional capture by animal stimuli, independent of lower-level visual cues. The authors conclude that “object related visual information can be extracted extremely rapidly from complex natural scenes” (Kirchner and Thorpe 2006, page 1775).

[¶]Current address: Child and Adolescent Mental Health, Sir James Spence Institute, Royal Victoria Infirmary, Newcastle upon Tyne NE1 4LP, UK; e-mail: sue.fletcher-watson@ncl.ac.uk.

In contrast to Kirchner and Thorpe's results, an earlier study (Brown et al 1997) failed to find evidence of rapid processing of human-face stimuli. Participants were asked to make an eye movement to one intact face, presented among three or seven inverted-face or jumbled-face distractors. When first given this task, no participant succeeded in generating more first saccades to the face than the proportion expected by chance. Improvement occurred following training, but this improvement was accompanied by a substantial delay in the initiation of the first saccade. Our present study was motivated in part by the difference between this finding and that of Kirchner and Thorpe. A number of design distinctions can be noted between the studies. Kirchner and Thorpe used a two-element search task, whereas the Brown et al displays had at least four elements. The stimuli in the Kirchner and Thorpe study were coherent scenes and contained whole animals whereas the Brown et al study used isolated exemplars showing a face only.

In the current study we used stimuli similar to those of Kirchner and Thorpe (2006), presenting two natural scenes, one of which contained a human figure. One goal was to examine spontaneous-looking patterns—to determine whether a complex visual object, such as a person, would capture gaze without being a search target. Of particular interest was the rapidity with which participants might direct their gaze to this complex stimulus. We aimed to discover whether the short latencies for eye movements to an animal stimulus found in a search task (Kirchner and Thorpe 2006) could be replicated for a human stimulus under free-viewing conditions.

A second goal of this study was to investigate looking patterns when participants were given viewing instructions; in this case, to decide on the gender of a person in the picture. This second condition is more similar to the methods of previous studies (Brown et al 1997; Kirchner and Thorpe 2006). We reasoned that our instruction would provide a naturalistic viewing goal, while avoiding any ad hoc strategy that might be employed if specific search instructions were used.

We additionally took the opportunity to investigate a further question of interest. It has been shown in simple tasks that the eye gaze and head direction of others form a powerful cue to direct the attentional system towards objects in the environment. This 'gaze-following' process operates from early on in human development and has also been found when adults and children view static images (Chawarska et al 2003; Friesen and Kingstone 1998; Langton and Bruce 2000). We developed a way to test whether gaze-following would occur automatically, both in the spontaneous condition and also when participants are given instructions, even though for the instruction condition this response would not contribute to the task in hand (making a gender judgment).

2 Method

2.1 Participants

Participants were twelve university undergraduate and postgraduate students, half of them male, with normal uncorrected vision. Ages ranged from 18 to 22 years.

2.2 Materials and apparatus

From 80 colour photographs of rooms and gardens 40 stimuli were developed, either containing one person (person-present), or containing no one (person-absent). Each setting was used twice, photographed once with and once without a person, to provide matching across all trials for scene complexity and content. Person-present scenes were composed to provide a naturalistic example of an individual in a setting. Therefore the people depicted had different body positions and were engaged in different types of activities. However, all were fixating on an inanimate object in the room (not looking at the camera) and had a neutral expression when photographed.

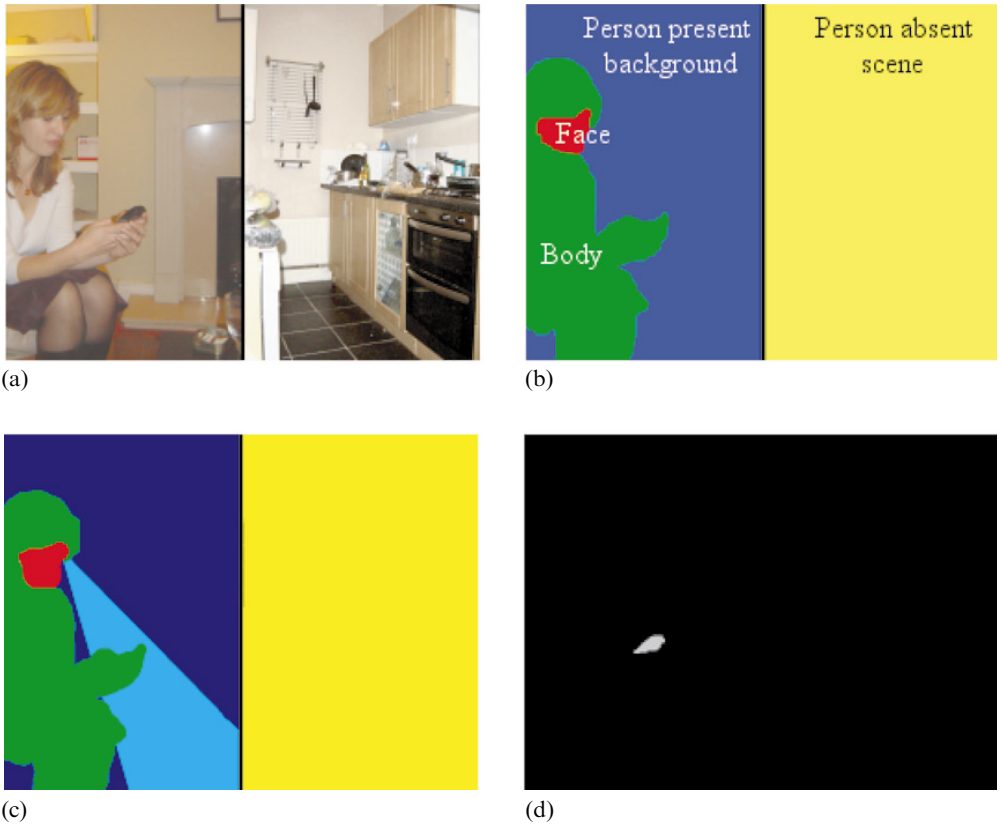


Figure 1. Examples of the stimuli used and the domains defined for analysis purposes: (a) trial stimulus; (b) analysis domain; (c) viewing cone; (d) object domain.

Person-present and person-absent scenes were randomly paired (with the exception that a particular setting was never paired with itself) to create combined stimuli with a scene on either side of a central, bisecting black line (see figure 1a). These stimuli were then processed in a graphics package (Paint Shop Pro), such that overall luminance and contrast range were equated across the image pair to create a stimulus whose two halves were matched for these low-level characteristics.

Stimuli were divided into two blocks of 20 combined scenes, and these blocks were matched for content of the person-present scene, by using various content categories. These categories included the setting of the image (garden, office, living room, kitchen), the location of the person-present scene (on the left or right of the stimulus), the direction of gaze of the person depicted (into or away from the person-absent scene), the posture of the person depicted (sitting or standing), the fixation of the person depicted (on a visible object or off-camera), the amount of face visible, the head angle of the person depicted (level, looking up, or looking down), and finally the gender of the person depicted. Participants saw all the stimuli, either seeing block one in the free-viewing condition, followed by block two in the gender-discrimination condition, or vice versa. Within each block, stimuli were presented in a random order, different for each participant.

Stimuli were presented on a colour monitor and filled the entire screen (1024×766 pixels). Eye movements were recorded with a Dual Purkinje Image eye tracker as participants viewed the stimuli binocularly from a distance of 1 m. The stimuli therefore subtended a visual angle of approximately $22 \text{ deg} \times 15 \text{ deg}$. Participants' eye movements

were monitored at a rate of 200 recordings per second. Analysis software gave an output including the path of each saccade and the location and duration of every fixation.

2.3 Procedure

Participants were given an information sheet outlining the study and gave their consent to be involved. They were introduced to the eye-tracking apparatus and a nine-point calibration was taken before each half of the experiment.

The first half of the experiment was the free-viewing condition, in which no instructions were given beyond: "You are going to see some colour pictures. Just have a look." Each trial consisted of a 1 s blank screen followed by a 1 s central fixation cross and then presentation of the experimental stimulus for 3 s. Participants viewed 21 stimuli, of which the first was a practice stimulus with no data recorded.

The second half of the experiment was the gender-discrimination condition, in which participants were told: "You're going to see some more pictures just like before, but this time I want you to decide on the gender of the person in the picture." Trials proceeded as before but with the addition of a response screen, saying "Respond Now", which appeared after the experimental stimulus. This response screen was presented until the participant made a button-press response to the gender. Participants viewed 23 stimuli, of which the first 3 were practice stimuli with no data recorded.

Participants were finally debriefed and paid £5 for their involvement.

3 Results

3.1 Data preparation

Each data set was run through an automatic saccade-detection programme, which detected the beginning and end of every saccade, and these selections were then checked manually. Each stimulus presented during the experiment was divided into five domains (see figure 1b) demarcating the face and body of the person in the scene, the background of each scene, and the black central bar separating the images. The tracker output was then combined with these domains to produce a data set recording how fixations were distributed across domains.

Each trial lasted 3000 ms; however, owing to tracker loss, only a fraction of this time was recorded for some trials. Since a partial recording may not be representative, all trials in which less than 500 ms of data were recorded were excluded. This resulted in the exclusion of 4% of all trials.

Three major measures were used in the analysis. First, the percentage of total viewing time spent in each domain. It was also possible to analyse the number of fixations made in each domain, but since this measure always produced the same results as the analysis of viewing time, it does not feature in this report. Second, the location of the first fixation in each trial was identified and, in addition, the time spent programming the saccade which led to this fixation (referred to hereafter as the 'first saccade latency'). This measure gave an insight into which domains were prioritised and how quickly gaze was directed to them. Finally, it was also possible to map the distribution of fixations over time, in order to reveal the order of precedence given to each domain.

3.2 Viewing time in each domain

Initial comparisons were made between the person-present scene (summing background, body and face domains) and the person-absent scene. Table 1 clearly shows a bias for participants to look at the person-present scene and table 2 shows how viewing time was distributed between parts of this scene only. Under each condition, participants spent a significantly greater proportion of their time looking at the person-present scene (free-viewing condition, $t_{11} = 4.0$, $p = 0.002$; gender-discrimination condition, $t_{11} = 12.6$, $p < 0.001$).

Table 1. Proportion of total viewing time and proportion of first fixations spent in each scene—unadjusted and domain-relative scores. The table shows means with standard deviations in parentheses.

	Person-absent scene		Person-present scene	
Unadjusted scores/%				
Total viewing time				
free-viewing	35.7	(7.6)	58.6	(9.4)
gender-discrimination	16.8	(6.7)	78.7	(9.9)
First fixations				
free-viewing	26.2	(13.8)	67.0	(17.7)
gender-discrimination	12.7	(9.5)	81.7	(13.5)
Domain-relative scores				
Total viewing time				
free-viewing	0.72	(0.16)	1.07	(0.27)
gender-discrimination	0.37	(0.17)	1.51	(0.3)
First fixations				
free-viewing	0.43	(0.52)	1.20	(0.4)
gender-discrimination	0.24	(0.42)	1.52	(0.39)

Note: Percentages for the person-absent and person-present scenes do not sum to 100% because a small proportion of fixations were made on the central black dividing line.

Table 2. Proportion of total viewing time and proportion of first fixations spent in sub-domains of the person-present scene. The table shows means with standard deviations in parentheses.

	Total	Background (minus object)	Body	Face	Object
Unadjusted scores/%					
Total viewing time					
free-viewing	58.6 (9.4)	18.3 (5.7)	12.3 (4.4)	17.6 (6.6)	10.4 (12.2)
gender-discrimination	78.7 (9.9)	15.9 (12.1)	22.7 (7.9)	32.6 (14)	7.5 (11.8)
First fixations					
free-viewing	67.0 (17.7)	24.1 (12.5)	27.1 (13.7)	14.9 (11.5)	0.8 (3.0)
gender-discrimination	81.7 (13.5)	26.6 (11.2)	25.7 (10.4)	28.2 (16.0)	1.28 (5.3)
Domain-relative scores					
Total viewing time					
free-viewing	1.07 (0.27)	0.59 (0.13)	2.4 (1.1)	72.6 (37.6)	11.73 (35.8)
gender-discrimination	1.51 (0.3)	0.46 (0.2)	4.86 (1.91)	150.47 (71.3)	7.62 (19.1)
First fixations					
free-viewing	1.2 (0.4)	0.51 (0.73)	5.57 (8.2)	51.53 (152.23)	0.39 (1.42)
gender-discrimination	1.52 (0.39)	0.60 (0.74)	5.71 (9.13)	87.1 (153.9)	0.47 (1.9)

Note: Standard deviations are enlarged by the practice of converting scores to a domain-relative measure. This is because domain-relative scores must take account of the variance between individual viewing times and between domain sizes.

This result was very consistent between individual participants. For the free-viewing condition, only one participant did not show this bias with fixation time roughly equally distributed between both scenes. The remaining eleven participants all showed a difference where the fixation time to the person-present scene was at least 7 percentage points above that to person-absent scenes, and five participants showed a difference of over 30 percentage points. For the gender-discrimination condition, eleven participants showed a difference of over 50 percentage points and the remaining participant showed a difference of approximately 25 points. This was not the same individual who showed no bias in the free-viewing condition.

The domains varied in size between images, and, in particular, face domains were always much smaller than any other domain. Therefore scores were adjusted to give domain-relative measures. Viewing times as percentages were divided by the size of the relevant domain, also expressed as a percentage of the total stimulus size. If fixations were spread randomly across the stimulus, one would expect a score close to one (eg 50% of fixation time/50% of stimulus size). If fixations were being directed to a domain more than randomly predicted, this score would be greater than one, and vice versa.

Tables 1 and 2 illustrate how viewing time was distributed across domains in each condition. In the free-viewing condition, the normalised viewing times to the person-absent scene and to the person-present scene background were significantly less than 1 (both $ps < 0.001$). In fact, there was no significant difference between domain-relative viewing time in the person-absent scene and the person-present scene background ($p = 0.12$), indicating that the difference in viewing time between these two scenes is entirely due to looking to the body and face domains. There is also a disproportionately large normalised viewing time for fixations to the body and particularly the face domain (both $ps < 0.001$). The same pattern is found for the gender-discrimination condition; all scores are significantly different from 1 (all $ps < 0.001$), in the same direction as for free-viewing, and once more there is no significant difference between viewing time in the person-absent scene and the person-present scene background ($p = 0.11$).

3.3 First fixations

The first fixation is defined as the fixation following the first saccade made after trial onset. This measure shows which items received attentional priority in the scene, rather than simply which items received the greatest amount of viewing time. Data on the distribution of first fixations are shown in tables 1 and 2.

In the free-viewing condition, significantly more first fixations fell on the person-present scene than the person-absent scene ($t_{11} = 5.1$, $p < 0.001$), indicating that the attentional bias found in the viewing-time measure is present from the very first fixation. Once more, this bias was very consistent among participants. Only one participant showed the opposite bias, making two more first fixations in the person-absent than in the person-present scene under free-viewing conditions. This was a different individual from the two who showed lesser or absent biases in percentage fixation times (see above). Seven of the twelve participants made over three times as many first fixations in the person-present as in the person-absent scene.

In addition, in the free-viewing condition, 27% of first fixations fell on the body and 15% on the face domains. When converted to domain-relative scores, the proportion of first fixations falling on the body domain is 5.57 while the face domain score is 51.5. The number of first fixations falling on the body was significantly negatively correlated with the distance of the body from the central fixation cross ($r = -0.479$, $p < 0.001$, one-tailed). Likewise, there was a significant negative correlation between the number of first fixations on the face for each stimulus, and the distance from the central fixation cross to the centre of the face ($r = -0.29$, $p = 0.007$, one-tailed).

In the gender-discrimination condition there was again a significant difference between first fixations in the person-present and person-absent scenes ($t_{11} = 10.1$, $p < 0.001$). All participants showed the person-present bias in their distribution of first fixations.

In this condition, 26% of first fixations fall on the body and 28% fall on the face. As domain-relative scores, these proportions are 5.7 for the body domains and 87.1 for the face domain. Again there was a significant correlation between the number of first fixations on the body and distance of the body from the fixation cross ($r = -0.279$, $p = 0.009$, one-tailed) and between first fixations on the face and distance of the face from fixation cross ($r = -0.21$, $p = 0.037$, one-tailed).

3.4 First-saccade latencies

Brown et al (1997) found that improved discrimination occurred when the first saccade was delayed. Thus, it may be that first fixations on the face and body are only achieved by extending the programming time—ie increasing the latency of the saccade before the first fixation. To investigate this, all first fixations were allotted to a 50 ms latency ‘bin’. We could then examine the distribution of possible locations of the first fixation within each latency bin. Figure 2 illustrates the distribution of first fixations in the person-present scene and person-absent scene, within each bin under free-viewing and gender-discrimination conditions. This clearly shows that the preference for first fixations to go into the person-present scene is manifest for saccades with short latencies.

Overall, there was no significant difference between mean first-saccade latencies between the two conditions ($p = 0.27$, free-viewing mean = 276 ms, gender-discrimination mean = 229 ms). In both conditions the majority of first saccades had latencies between 100 ms and 249 ms; free-viewing 57% and gender-discrimination 71%. In addition, the majority of first saccades to every scene domain fell into the same range of latencies—person-absent scene 50%; person-present scene: background 73%, body 75%, face 80%. This indicates that no domain entailed longer first-saccade latencies than any other.

There was a small number of cases where the first saccades had latencies less than 100 ms, and for these saccades participants appeared equally likely to make their first fixation in either scene, and in fact often made their first fixation within the central dividing bar. These saccades may be presumed to be anticipatory.

The bias to fixate in the person-present scene appears in the 100–149 ms bin to which six participants contributed a total of 9 fixations. Here 83% of participants show a bias for the person-present scene (in fact, two-thirds of participants make every first fixation in the person-present scene) with a mean latency of 142 ms for the preceding saccades. The one participant who does not show this bias made only one first fixation in this bin which was to the person-absent scene. Ten of the participants contributed to the 150 ms latency bin, which contains a total of 56 fixations, and all showed a bias for the person-present scene.

In the free-viewing condition, the highest latency bins indicate a return to near-random distribution of first fixations between the two scenes. This may be because longer latencies are produced when there is unusually large conflict between the two scenes in terms of the interest they present. This occurs in a minority of cases (only 13% of first fixations occurred after saccade latencies greater than 400 ms).

The gender-discrimination data reveal a similar pattern (see figure 2b) in terms of the near-random distribution of fixations with the shortest saccade latencies, followed by a very strong preference for fixating the person-present scene from the 100–149 ms latency bin upwards. In this condition, 75% of participants showed the person-present

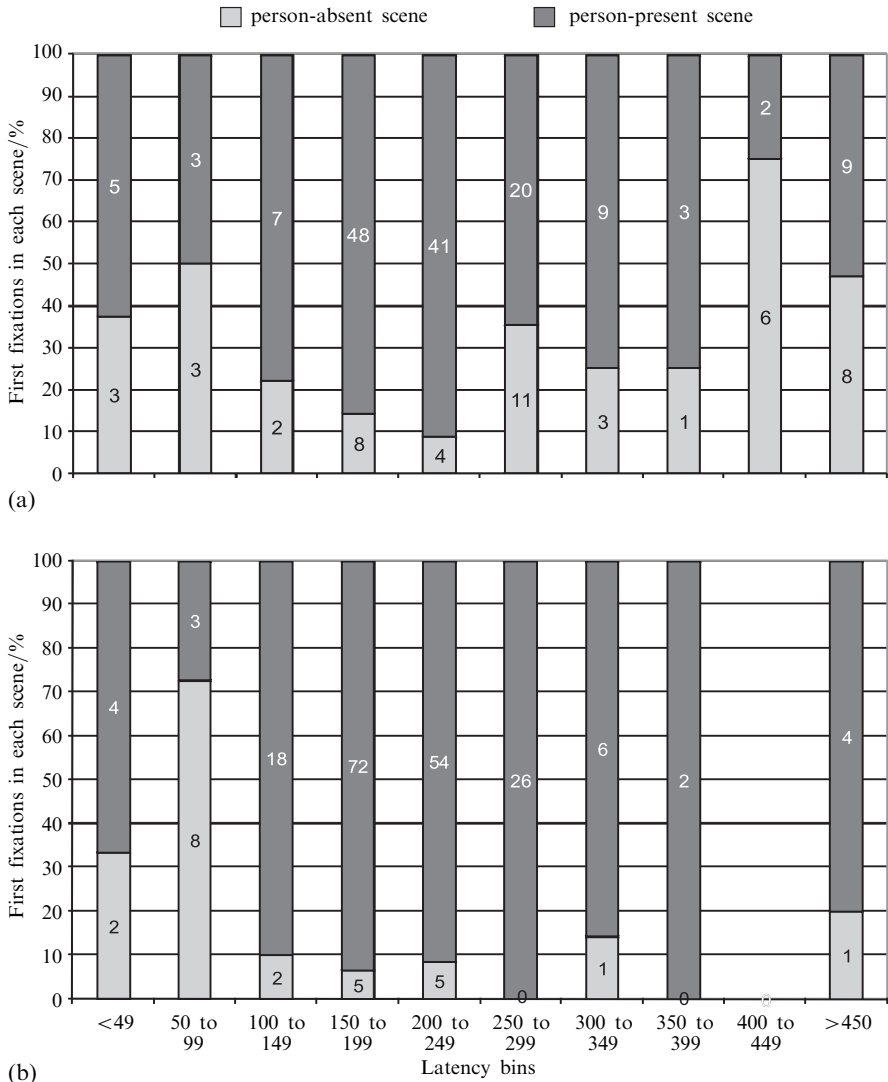


Figure 2. Distribution of fixations in person-present and person-absent scenes, sorted into 50 ms latency bins, under (a) free-viewing and (b) gender-discrimination conditions. The data labels show the actual number of fixations for each scene in each bin. Percentages for the two scenes have been calculated excluding fixations made on the central black dividing line, so that they always sum to 100%.

bias in the 100–149 ms bin, to which four participants contributed a total of 20 fixations, with a mean latency of 136 ms for the preceding saccades. Again, the single participant who did not show this bias made only one fixation in this bin. In the 150–199 ms bin, to which eleven participants contributed a total of 77 fixations, the proportion of participants showing a bias was 100%.

In contrast to the free-viewing data, the gender-discrimination data do not indicate a return to evenly distributed fixations at the longest latencies. Presumably, the task requirement of locating the scene character overrode any conflicting interest coming from the person-absent scene.

3.5 Comparisons between conditions

Comparing the proportion of viewing time per domain between the two conditions (free-viewing and gender-discrimination) reveals a strong effect of task requirement on looking behaviour. The proportion of viewing time spent in the person-present scene and also in the body and face domains is significantly greater under the gender-discrimination condition (person-present scene, $t_{11} = 6.3$, $p < 0.001$; body, $t_{11} = 5.2$, $p < 0.001$; face, $t_{11} = 4.4$, $p < 0.001$). Likewise, the proportion of viewing time spent in the person-absent scene is less in the gender-discrimination condition ($t_{11} = 7.0$, $p < 0.001$).

A similar pattern is revealed by the distribution of first fixations which shows a significantly larger number in the gender-discrimination condition for the person-present scene and face domain (person-present scene, $t_{11} = 4.3$, $p = 0.001$; face, $t_{11} = 3.3$, $p = 0.007$), though no difference for the body domain. Again, there are also fewer first fixations in the person-absent scene in the gender-discrimination condition ($t_{11} = 6.2$, $p < 0.001$).

3.6 Looking where others look

Two new domains were also developed (see figures 1c and 1d) in order to investigate the secondary question whether participants looked at the areas looked at by the person represented in the scene. The first 'viewing cone' domain represented the area that could conceivably be being viewed by the scene character, as defined by a 30° cone extending from the centre of the person's eye area. In addition, for the subset of scenes (26/40) in which a visible object was being fixated by the person in the scene, this object was analysed separately with its own domain, though it always fell almost entirely within the viewing cone.

Domain-relative fixation time scores were calculated as before, a score of 1 indicates randomly allocated viewing, less than 1 indicates a paucity of viewing of this area, and scores greater than 1 indicate an excess of viewing in this area. The data show that participants directed their fixation to the object being fixated by the scene character more than would be predicted by a random viewing pattern in the free-viewing condition ($t_{11} = 2.2$, $p = 0.05$). However, this was not replicated in the gender-discrimination condition ($p = 0.08$).

The viewing cone did not receive a significantly inflated amount of fixation in either condition (free-viewing proportional score = 1.21, gender-discrimination proportional score = 1.5). The slightly enhanced scores can probably be attributed to time spent looking at the object, which was wholly or mostly encompassed by the viewing cone.

3.7 Distribution of fixations over time

The pattern of looking over time was examined by plotting the percentage of first, second, third, fourth, etc fixations made within each domain in each condition (see figures 3a and 3b). The graphs illustrate that first fixations are as likely to be made in the body as in the face domain. Attention to the face increases over the first two fixations, then decreases, and becomes stable from the fourth fixation onwards in both conditions. This indicates that participants may be using the human figure as a whole to direct their first saccade. They can then use the location of their first fixation to direct a second saccade to the most interesting part of the body, usually the face. Attention to the body diminishes after the first fixation in the free-viewing condition. This trend is not visible in the gender-discrimination condition, indicating that information from the body was being used to identify gender.

Attention to the object fixated by the person in the scene does not appear until slightly later and increases as looking at the face domain decreases. This indicates that gaze-following may be occurring, whereby the participant looks first to the face of the person depicted in the scene and then to the object being fixated by that person.

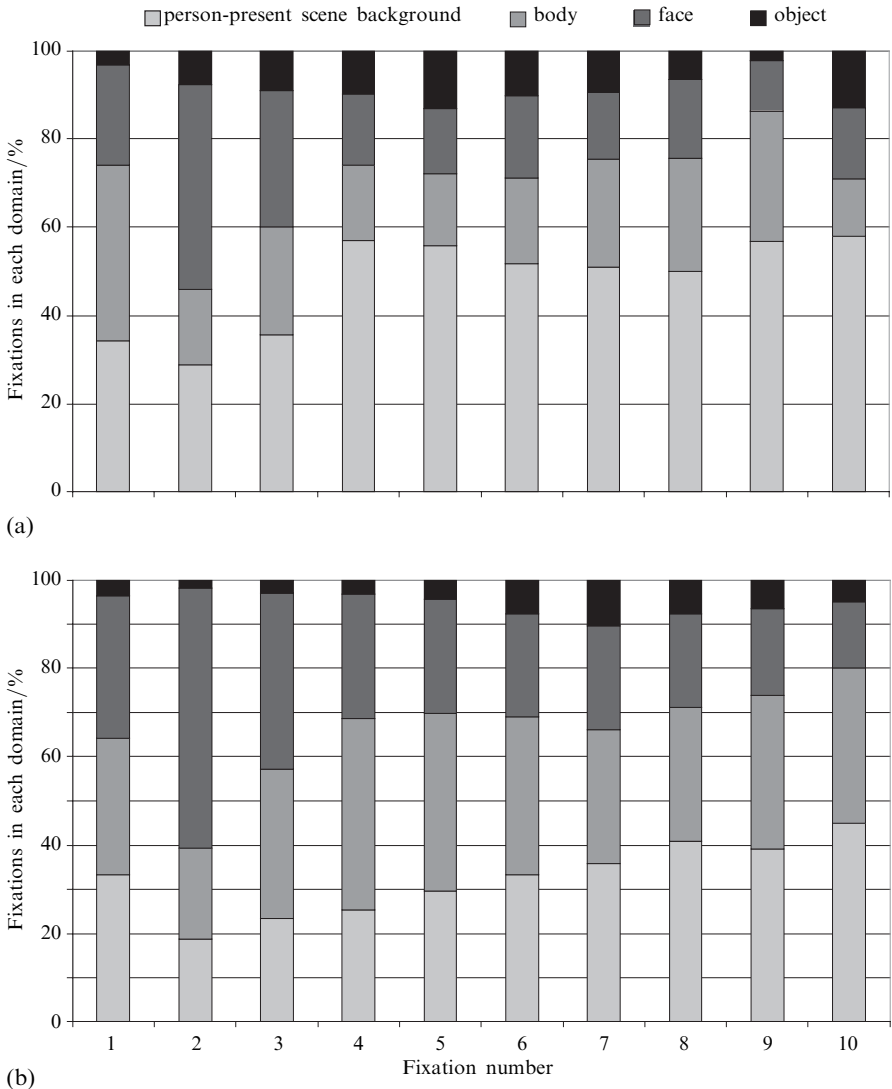


Figure 3. Fixations on the person-present scene only. Distributions of fixations to different domains over time: (a) free-viewing, (b) gender-discrimination conditions. In this figure, fixation 1 is the fixation made after the first saccade after trial onset.

4 Discussion

This study clearly shows a strong bias towards attending to a scene containing a person, and particularly towards that person. Even when no task instructions are given, this bias leads people to spend well over half their time looking at the person-present scene and, when looking at this scene, to spend most of their time looking at the person. In fact, there are no differences in looking to the background of the person-present and person-absent scenes when domain size is taken into account. This indicates that the difference between looking to the person-absent and person-present scenes is entirely due to the time spent looking at the person.

The bias affects the landing position of the very first saccade, occurring as early as 100 ms after the scene presentation. This finding replicates the work of Kirchner and Thorpe (2006) and other recent work (Herschler and Hochstein 2005, 2006; Rousselet et al 2003; Smilek et al 2006) that has demonstrated rapid attentional

capture by high-level stimuli. Furthermore, we believe our demonstration is the first that shows this tendency is present in a spontaneous-looking condition (although see Langton et al 2006). As in the Kirchner and Thorpe study, ultra-rapid visual processing of complex socially relevant visual information is demonstrated. This process is clearly very effective in directing attention, although some limits to its power are shown by the negative correlation between proportion of person-directed first fixations and the distance of the person from the initial site of foveation.

We do not feel that it is likely that any account of our results could be given in terms of low-level visual factors; for example the person in figure 1a is composed of several different small areas of different colour, none in themselves distinctive. However, given the multiplicity of possible low-level factors (size, contrast, curvature, complexity) and their combinations, we note that this possibility cannot be completely excluded. An alternative high-level explanation has been suggested by an anonymous referee in that the person provides an item of 'central interest' within the scene (Rensink et al 1997) and it is this, rather than any social aspect of the figure, which is important. Further experimental work will be necessary to examine this interesting suggestion.

Our data would seem at first sight to contradict the results of Brown et al (1997) who failed to find an ability to generate first saccades to faces in a search task. In the current study, participants managed to land 15% of their first fixations on the face even with no task instructions and these fixations took no longer to prepare than those to other domains. As discussed in section 1, various reasons might be suggested for the discrepancy, such as the presence of multiple options in Brown et al's search task versus our simpler two-choice preferential-looking design. A second possibility is that the inverted and jumbled faces used as distractors in Brown et al's study themselves partially activated the peripheral face perception processes and were thus not readily discriminable from the upright-face target. Our present study contained not only a single face, but no distractor which could be easily confused with a face.

Another important difference between our study and that of Brown et al is that our face stimuli were presented as part of a whole person, in a real-world scene. It is plausible that one should find it easier to detect a face when its whereabouts is indicated by other cues, such as the presence of a body. Although the face is a socially powerful stimulus, it is also relatively small and thus covers a smaller angle in visual periphery than the body. In our task, first fixations were more likely to be directed to the body than face in the free-viewing condition. Figure 3 shows that the likelihood of fixating the face increases, in both conditions, between fixations one and two. Domain-relative scores emphasise the importance of the face area, illustrating a highly non-random distribution of fixations with a powerful focus on fixating the face. Therefore, although the pattern of fixations may have the aim of bringing gaze to the face, this often occurs in stages and our data suggest that it is the entire human figure, rather than the face, that is subject to ultra-rapid processing.

When a task requirement was introduced, namely to identify the gender of the person in the scene, looking time to the person predictably rose although the differences were most marked after about the third fixation. Fixations were still rapidly and accurately directed to the person-present scene, and to face and body domains specifically. As in the free-viewing condition, 40% of the time looking at the person was spent on the body. In addition, in the gender-discrimination condition, fixation on the body was spread consistently over time. This indicates that not only may the body attract initial attention in a spontaneous viewing situation, but that it contributes significantly to making a decision about a person's gender.

Finally, the data provide evidence of gaze-following behaviour in people's viewing patterns. Participants fixated the object being looked at by the scene character more than would be randomly predicted, though this increased viewing of the area fixated

by the scene character did not extend to the whole viewing cone. The distribution of fixations over time shows that looking to the object domain seemed to occur after fixation on the face, indicating that looking at a static scene can incorporate a genuine gaze-following pattern. The evidence reveals a certain amount of control over this process, since looking to the object domain was not enhanced in the gender-discrimination task (when gaze-following was irrelevant to the task in hand). Conversely, the fact that gaze-following occurred while observing a static photo stimulus at all, indicates a certain amount of automaticity, since participants could not have had a successful interaction with the person depicted. These data also illustrate the precision of our gaze-following system in that participants did not increase looking to the entire visual field of the scene character (the viewing cone) but only to those specific items being looked at by the character.

These results strongly suggest that human figures and their faces are subject to special perceptual attention when presented as part of a scene. We are able to direct our gaze to a scene containing a person, and even directly to his/her face, incredibly rapidly and still do so when no task constraints are in place. It would be an interesting next step to investigate how much first fixations falling on the face might increase if participants were explicitly instructed to move their eyes to the face immediately upon stimulus presentation. The study also extends the finding of Kirchner and Thorpe (2006) that people can identify the scene containing an animal from two natural scenes, to include the ability to identify the scene containing a human and to do so even without specific instructions. This bias towards fixating the social scene, and specifically the person therein, argues strongly for a stimulus-driven visual system that is tuned to high-level properties.

Acknowledgments. We would like to thank Bob Metcalfe for his expertise in developing the analysis software. We are indebted to Simon Liversedge, Sarah White, Holly Joseph, and Hazel Blythe for their support in the eye-tracking laboratory. Also, our thanks go to two anonymous reviewers for their constructive comments.

References

- Brown V, Huey D, Findlay J M, 1997 "Face detection in peripheral vision: do faces pop out?" *Perception* **26** 1555–1570
- Buswell G T, 1935 *How People Look at Pictures* (Chicago, IL: University of Chicago Press)
- Chawarska K, Klin A, Volkmar F, 2003 "Automatic attention cueing through eye-movement in 2-year-old children with autism" *Child Development* **74** 1108–1122
- Findlay J M, 1997 "Saccade target selection in visual search" *Vision Research* **37** 617–631
- Friesen C K, Kingstone A, 1998 "The eyes have it: reflexive orienting is triggered by nonpredictive gaze" *Psychonomic Bulletin & Review* **5** 490–495
- Henderson J M, 2003 "Human gaze control during real world scene perception" *Trends in Cognitive Sciences* **7** 498–504
- Henderson J M, Hollingworth A, 1999 "High-level scene perception" *Annual Review of Psychology* **50** 243–271
- Herschler O, Hochstein S, 2005 "At first sight: a high-level pop-out effect for faces" *Vision Research* **45** 1707–1724
- Herschler O, Hochstein S, 2006 "With a careful look: Still no low-level confound to face pop-out" *Vision Research* **46** 3028–3035
- Kirchner H, Bacon N, Thorpe S J, 2003 "In which of two scenes is the animal? Ultra-rapid visual processing demonstrated with saccadic eye movements" *Perception* **32** Supplement, 170
- Kirchner H, Thorpe S J, 2006 "Ultra-rapid object detection with saccadic eye movements: Visual processing speed revisited" *Vision Research* **46** 1762–1776
- Langton S R, Bruce V, 2000 "You must see the point: automatic processing of cues to the direction of social attention" *Journal of Experimental Psychology: Human Perception and Performance* **26** 747–757
- Langton S R H, O'Donnell C, Riby D M, Ballantyne C J, 2006 "Gaze cues influence the allocation of attention in natural scene viewing" *Quarterly Journal of Experimental Psychology* **59** 2056–2064

-
- Rensink R, O'Regan K, Clark J J, 1997 "To see or not to see: The need for attention to perceive changes in scenes" *Psychological Science* **8** 368–373
- Rousselet G A, Mace M J, Fabre-Thorpe M, 2003 "Is it an animal? Is it a human face? Fast processing in upright and inverted natural scenes" *Journal of Vision* **3** 440–455
- Smilek D, Dixon M J, Merikle P M, 2006 "Revisiting the category effect: The influence of meaning and search strategy on the efficiency of visual search" *Brain Research* **1080** 73–90
- Thorpe S J, Gegenfurtner K R, Fabre-Thorpe M, Bülthoff H H, 2001 "Detection of animals in natural images using far peripheral vision" *European Journal of Neuroscience* **14** 869–876
- Yarbus A L, 1967 *Eye Movements and Vision* Translated into English by L A Riggs (New York: Plenum)

ISSN 0301-0066 (print)

ISSN 1468-4233 (electronic)

PERCEPTION

VOLUME 37 2008

www.perceptionweb.com

Conditions of use. This article may be downloaded from the Perception website for personal research by members of subscribing organisations. Authors are entitled to distribute their own article (in printed form or by e-mail) to up to 50 people. This PDF may not be placed on any website (or other online distribution system) without permission of the publisher.