Journal of Eye Movement Research
1(1):1, 1-6

Why do we look at people's eyes?

Elina Birmingham
University of British Columbia

Walter Bischof
University of Alberta

Alan Kingstone
University of British Columbia

We have previously shown that when observers are presented with complex natural scenes that contain a number of objects and people, observers look mostly at the eyes of the people. Why is this? It cannot be because eyes are merely the most salient area in a scene, as relative to other objects they are fairly inconspicuous. We hypothesized that people look at the eyes because they consider the eyes to be a rich source of information. To test this idea, we tested two groups of participants. One set of participants, called the Told Group, was informed that there would be a recognition test after they were shown the natural scenes. The second set, the Not Told Group, was not informed that there would be a subsequent recognition test. Our data showed that during the initial and test viewings, the Told Group fixated the eyes more frequently than the Not Told group, supporting the idea that the eyes are considered an informative region in social scenes. Converging evidence for this interpretation is that the Not Told Group fixated the eyes more frequently in the test session than in the study session.

Keywords: Eye movements, Social attention, Scene viewing, Memory

Introduction

Recent studies have found that when viewing real world social scenes, observers look preferentially at the eyes of people relative to other scene regions (Birmingham et al., in press (a); Birmingham et al., in press (b); Smilek, Birmingham, Cameron, Bischof & Kingstone, 2006). This preference occurs in scenes displaying a variety of social situations (e.g. one or three people performing various activities) and across several different tasks (e.g. freely viewing the scenes, describing the scenes, inferring the attentional states of the people in the scenes).

Why are the eyes preferentially selected? A social attention explanation posits that eye gaze is a powerful indicator as to where other people are attending (Baron-Cohen, 1994). This is evidenced by recent studies showing that infants (Hood, Willen & Driver, 1998), preschool children (Ristic, Friesen & Kingstone, 2002) and adults alike (Friesen & Kingstone, 1998; Langton & Bruce, 1999) shift attention automatically to where other people are looking. Observers also look at the eyes of people in a scene most frequently when asked where those people are attending, although eyes are still selected more than other objects even when participants are asked to describe or just look at a scene (Birmingham et al. (a, b); Smilek et al., 2006).

Collectively, these findings suggested to us that people may preferentially select the eyes because they consider the eyes to contain important social information regarding the meaning of a scene. For instance, scanning the eyes of people in a scene might help to clarify the nature of the social situation being depicted, e.g. Are the people in the scene interested in each other, and if so, is it a friendly or aggressive interaction? In this way, the eyes are informative scene regions, allowing the observer to build an in-depth understanding of the scene and its underlying meaning.

If eyes are perceived to be informative regions of a scene, then it follows that observers should scan them more frequently when they are trying to encode a scene into memory. This is consistent with non-social scene perception studies showing that observers tend to fixate informative scene regions more frequently than uninformative scene regions when asked to encode the scene into memory (Henderson et al., 1999). For instance, Henderson et al. (1999) told observers that they would have to remember the scenes in a later memory test, and

DOI 10.16910/jemr.1.1.1

ISSN 1995-8692

found that the semantically informative scene items (i.e., those items that are inconsistent with the basic context of a scene, such as a microscope being in a kitchen) were fixated more frequently than semantically uninformative scene items (those that are consistent with scene context, such as a glass being in a kitchen). Although no memory test was actually given to the observers in Henderson et al.'s study, their results suggest that observers strategically fixate informative scene regions when asked to encode scenes into memory.

The objective of the present study was to determine whether observers perceive the eyes to be informative scene regions. Of course, the way in which the eyes are informative must differ profoundly from the definition used by Henderson et al. (1999) for non-social scenes, because the eyes are not semantically inconsistent within a scene containing people. Rather, we suggest that the eyes are informative because they provide social-communicative information that adds meaning to a scene. However, we adopted the logic from Henderson et al. (1999) and reasoned that if the eyes are perceived as informative, they will be fixated more when observers are trying to encode scenes into memory than when simply free viewing the scenes. Thus, we gave a set of scenes (social and non social) to observers in a study session. One group was told that they would later be asked to recognize the scenes in a test session (Told group); another group was not informed of the later memory test and simply asked to freely view the images (Not Told group). Both groups were subsequently given a memory test session, in which scenes from the study session were presented along with scenes never seen before. We hypothesized that observers in the Told group would fixate the eyes more frequently than observers in the Not Told group. This effect of group was expected for the study session, when observers were encoding the scenes into memory. However, we also expected that observers would use the eyes to recognize the scenes in the later memory test, again resulting in a higher fixation frequency on eyes for the Told group than for the Not Told group.

Methods

Participants

Ten undergraduate students from the University of British Columbia participated. Participants were randomly assigned to the Told group (n=5, 3 male, 2 female, mean age = 19) or the Not Told group (n=5, 1 male, 4 female, mean age = 20). All had normal or corrected to normal vision, and were naïve to the purpose of the experiment. Each participant received course credit for participation in a one-hour session.

Apparatus

Eye movements were monitored using an Eyelink II, which is a head mounted video-based eye tracking system . The on-line saccade detector of the eye tracker was set to detect saccades with an amplitude of at least 0.5°, using an acceleration threshold of 9500°/s2 and a velocity threshold of 30°/s. Sampling frequency was set to 250Hz, and we used pupil+corneal reflection (CR) tracking or pupil-only tracking in cases where CR was unreliable. Eyelink II has a resolution of 0.01° and an average gaze position accuracy of 0.5°.

Stimuli

Full color digital photos were taken of different rooms in the UBC Psychology building. Image size was 36.5 x 27.5 (cm) corresponding to 40.1° x 30.8° at the viewing distance of 50 cm, and image resolution was 800 x 600 pixels.

Study session (15 scenes: 3 rooms, 5 scene types). Twelve of the fifteen study scenes were "People scenes". These scenes contained a variety of social situations containing 1 or 3 persons. All scenes were comparable in terms of their basic layout: each room had a table, chairs, objects, and background items (e.g. see Figure 2). Three of the fifteen study scenes were "No people scenes", containing a single object resting on the table. Examples of these scene-types are presented in Figure 1.

Test session (56 scenes: 8 rooms, 7 scene types). Thirty-two of the test scenes were "People scenes" as above (12 old, 20 new). Sixteen of the test scenes were "No people scenes", containing one or three objects resting on the table (3 old, 13 new). Eight additional (new) scenes contained one person doing something unusual, such as sitting with a Frisbee on his head. These scenes were included to keep the participants interested, but were not included in the analysis.

Due to differences in the number of people (1 or 3) and variation in distance between the people and the camera, the eye region in the People scenes varied in area from 1.69 deg2 to 9.47 deg2, with an average area of 4.57 deg2. Specific experimental details are presented below. The No People scenes were included so that we could determine whether scene regions that were informative to

DOI 10.16910/jemr.1.1.1

ISSN 1995-8692

No People scenes would be informative once people were added to the scene.

Procedure

Participants were seated in a brightly lit room, and were placed in a chin rest so that they sat approximately 50 cm from the display computer screen. Participants were told that they would be shown several images, each one appearing for 10 seconds.

Before the experiment, a calibration procedure was conducted. Participants were instructed to fixate a central black dot, and to follow this dot as it appeared randomly at nine different places on the screen. This calibration was then validated with a procedure that calculates the difference between the calibrated gaze position and target position and corrects for this error in future gaze position computations. After successful calibration and validation, the scene trials began. At the beginning of each trial, a fixation point was displayed in the centre of the screen in order to correct for drift in gaze position. Participants were instructed to fixate this point and then press the spacebar to start a trial. A picture was then shown, filling the entire screen. Each picture was chosen at random and without replacement. The picture remained visible until 10 seconds had passed, after which the picture was replaced with the drift correction screen in the Study session, or a response screen in the Test session. In the Test session, after the participant entered a response, the drift correction screen appeared in preparation for the next trial. This process repeated until all pictures had been viewed.

Study session: Each participant was randomly assigned to one of two instruction groups. The Told group was told that they would be shown 15 images, and that they would be asked to recognize each image in a later memory test. The Not Told group was told to simply "look at" each image, and was not informed of the later memory test. After the study session, a brief questionnaire was given to participants asking them about their impressions of the experiment.

Test session: Both groups (Told, Not Told) were told that they would be shown 56 images, and that they were to view each one and then decide if the image was OLD (i.e. they had seen it in the Study session), or NEW (i.e. they had never seen it before). After an image was presented, a response screen appeared asking them to respond with '1' on the keyboard if they thought the image was OLD, and '2' on the keyboard if they thought the image was NEW. Participants had an unlimited amount of time to respond.

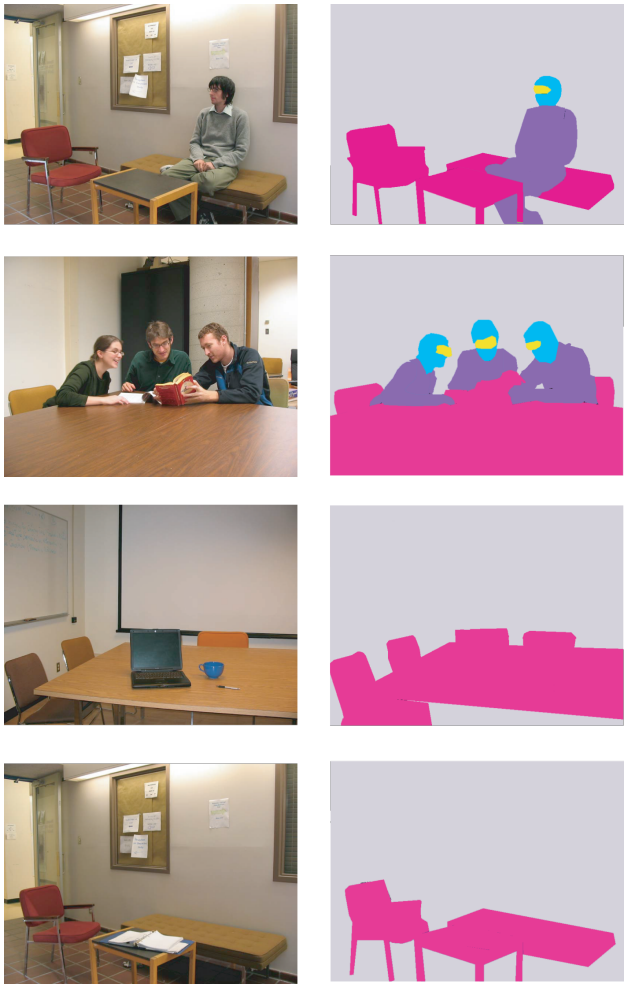

Figure 1. Examples of People scenes and No People scenes used in the experiment. (left-hand column); Corresponding regions of interest used in analysis (eyes, head, body, foreground objects, background objects) (rightward column).

Results

For each image, an outline was drawn around each region of interest (e.g. "eyes") and each region's coordinates and area were recorded. We defined the following regions in this manner: eyes, heads (excluding eyes), bodies (including arms, torso and legs), foreground objects (e.g., tables, chairs, objects on the table) and background objects (e.g., walls, shelves, items on the walls). Figure 1 illustrates these regions.

DOI 10.16910/jemr.1.1.1

ISSN 1995-8692

To determine what regions were of most interest to observers we computed fixation proportions by dividing the number of fixations for a region by the total number of fixations over the whole display. These data were area-normalized by dividing the proportion score for each region by its area (Smilek, et al., 2006; Birmingham et al. (a),(b)).

Figure 2

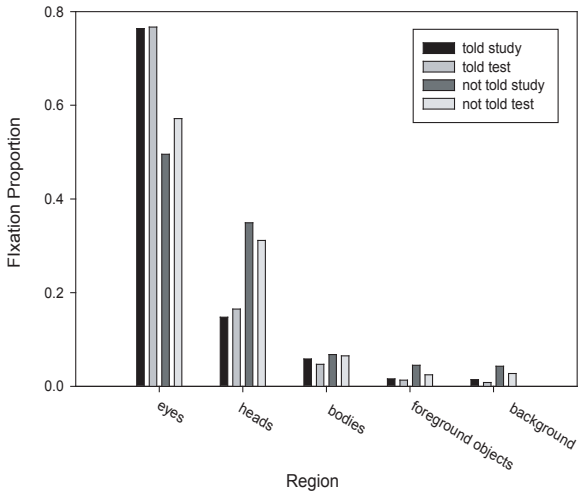

Figure 2. Area-normalized fixation proportions for the People scenes as a function of Instruction (Told, Not Told), Session (Study, Test), and Region (eyes, heads, bodies, foreground objects, background)

Figure 3

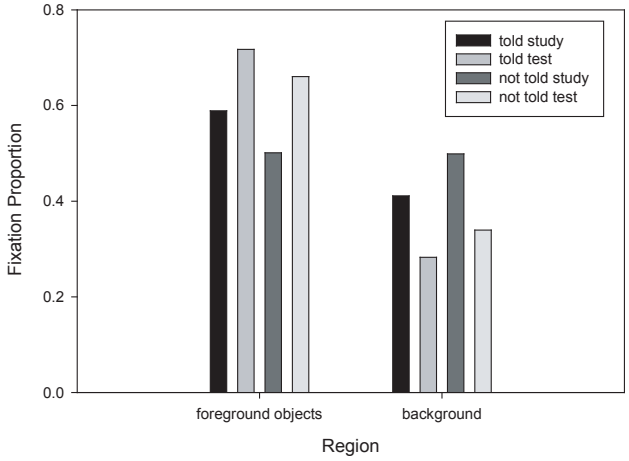

Figure 3. Area-normalized fixation proportions for the No People scenes as a function of Instruction (Told, Not Told), Session (Study, Test), and Region (foreground objects, background).

Figures 2 and 3 show the fixation proportions for People (Figure 2) and No People (Figure 3) scenes, as a function of Instruction (Told, Not Told), Session (Study, Test), and Region (eyes, heads, bodies, foreground objects, background).

In the analyses below, Huynh-Feldt adjusted probabilities are reported in cases where sphericity is violated.

People scenes

The data for the People scenes were submitted to a mixed ANOVA with Instruction (Told, Not Told) as a between-subjects factor and Session (Study, Test) and Region (eyes, heads, bodies, foreground objects, background) as within-subjects factors. This analysis revealed a highly significant effect of Region ($F(4.32)=175.68$, $p<0.0001$), reflecting that overall participants preferred to scan the eyes than any other region. An Instruction x Region interaction ($F(4,32)=13.12$, $p<0.005$) indicated that the eyes were fixated more frequently by the Told group than the Not Told group (Tukey-Kramer multiple-comparison test, $p<.05$), and the heads were fixated more frequently by the Not Told group than by the Told group (Tukey-Kramer, $p<.05$). This result indicates that when asked to encode and remember scenes with people, participants look to the eyes for information. Finally, a Instruction x Session x Region interaction reflected the fact that the Not Told group fixated the eyes more frequently in the test session than in the study session, whereas in the Told group the fixation patterns were the same across sessions. However, this interaction did not remain significant after the Huynh-Feldt correction ($F(4,32)=2.72$, p-0.11). Post-hoc comparisons confirmed that the Not Told group fixated the eyes more frequently in the test session than in the session (Tukey-Kramer, $p<.05$), and that this was not the case for the Told group ($p>.05$). This latter finding suggests that scanning the eyes to recognize the scenes is a natural strategy, because when participants were given a surprise memory test they looked at the eyes more than when they had been simply free-viewing the scenes (in the study session).

No People scenes

No People scenes. The data for the No People scenes were submitted to an Instruction (Told, Not Told) x Session (Study, Test) x Region (foreground objects, background) mixed ANOVA. This analysis revealed a main effect of Region ($F(1,8)=15.74$, $p<0.01$, suggesting that overall, participants preferred to scan the foreground objects in scenes without people (Tukey-Kramer, $p<.05$). However, unlike with the People scenes, Instruction had

DOI 10.16910/jemr.1.1.1

ISSN 1995-8692

no significant effect on scanning patterns (Instruction x Region ($F(1,8)=1.50$, $p>0.25$). Thus, the preference for foreground objects may reflect a more general scanning strategy than one specific to encoding and recognizing scenes. In addition, a Session x Region interaction ($F(1,8)=31.34$, $p<0.001$) reflected that foreground objects were fixated more frequently in the test session than in the study session, whereas background was fixated more frequently in the study session than in the test session (Tukey-Kramer, $p<.05$). This result suggests that participants relied more on the foreground objects to retrieve the scenes from memory in the test session, and used a more distributed scanning pattern (examining foreground and background) both when encoding (Told group) or free-viewing (Not Told group).

Recognition Accuracy

There were no group differences in accuracy on the test session, with both groups performing very well (Told group mean accuracy: 94.6%; Not Told group mean accuracy: 95.4%). We were expecting very high accuracy scores given the evidence that people are excellent at recognizing even very large numbers of scenes (e.g. Standing, 1973). Indeed, the fact that both groups performed equally well confirms that our eye movement effects in the test session were due to differences in strategy and not to differences in task difficulty.

Discussion

The goal of the present study was to determine whether the eyes are preferentially selected because they are perceived to be highly informative to scene meaning. Based on previous work with non-social scenes (Henderson et al., 1999), we predicted that observers would fixate highly informative regions more frequently when asked to encode scenes into memory than when simply free-viewing the scenes. In addition, we included a test session in which observers were asked to recognize the scenes from the previous study session. We reasoned that the informative scene regions would also be fixated frequently when observers were trying to remember the scenes.

The results were clear. Observers fixated the eyes within People scenes more frequently when asked to encode the scenes than when asked to freely view them. In addition, this bias carried over from study to test: the Told group fixated the eyes just as strongly in the test session as in the study session, and again more frequently than the Not Told group in the test session. These results suggest that the eyes are scanned strategically by observers who are aware that they will have to encode and remember the scenes. Interestingly, observers in the Not Told group fixated the eyes more strongly in the (surprise) test session than in the free-viewing study session. Thus, the eyes appear to be informative for both deliberately encoding scenes and for spontaneously trying to recognize them.

The No People scenes, on the other hand, showed a different pattern. The Told group and the Not Told group had very similar fixation patterns: overall, they both preferentially fixated the foreground objects, and there were no group differences. Thus, it appears as though within the No People scenes, the foreground objects were preferentially fixated as part of a general viewing strategy unrelated to scene encoding. Importantly, both the Told and the Not Told group fixated the foreground objects more, and the background less, in the test session than in the study session. This latter result suggests that the foreground objects were indeed used for retrieving the scenes from memory.

An interesting finding was that unlike the No People scenes, for the People scenes observers rarely fixated the foreground objects, relative to the eyes and heads. This supports the idea that humans have a special sensitivity to people and their eyes as sources of social communicative information (Baron-Cohen, 1994). One interesting future application of this finding would be to ask individuals with autism spectrum disorder (ASD) to perform the same task as the Told group in the present study. Individuals with ASD have been found to have an aversion to social stimuli, particularly to people and their eyes (e.g. Dalton et al., 2005; Pelphrey et al., 2002). Thus, one might expect these individuals to rely more on the foreground objects when encoding and retrieving social scenes from memory.

In conclusion, the findings from the present study suggest that the eyes are perceived to be highly informative scene regions. We speculate that this is because the eyes are socially communicative, providing meaning about the nature of the social situation being depicted in the scene. In particular, the eyes convey key information about where other people are directing their attention, as well reveal information about emotional and mental states (e.g. Baron-Cohen, Wheelwright & Jolliffe, 1997). Future studies will be required to uncover which, if any, of the types of information conveyed by the eyes is most important to constructing scene meaning. While previous

DOI 10.16910/jemr.1.1.1

ISSN 1995-8692

work suggested that the eyes are perceived to be meaningful for indicating the attentional states of other people (e.g. Birmingham et al., (b); Friesen & Kingstone, 1998; Smilek et al., 2006), to our knowledge this is the first demonstration that observers strategically use the eyes of others to encode social scenes. Future studies will be required to determine whether the eyes are more informative within highly social situations (e.g. three people interacting) relative to asocial situations (e.g. one person reading on their own).

References

Baron-Cohen, S. (1994). How to build a baby that can read minds:cognitive mechanisms in mindreading. *Cahiers de Psychologie Cognitive 13,* 513–552.

Baron-Cohen, S., Wheelwright, S., and Jolliffe, T. (1997). Is there a "language of the eyes"? Evidence from normal adults, and adults with autism or Asperger syndrome. *Visual Cognition, 4*(3), 311-331.

Birmingham, E., Bischof, W.F., & Kinsgtone, A. (a). Social attention and real world scenes: The roles of action, competition and social content. Accepted to *Quarterly Journal of Experimental Psychology* pending minor revisions.

Birmingham, E., Bischof, W.F., & Kinsgtone, A. (b). Gaze selection in complex social scenes. Accepted to *Visual Cognition* pending minor revisions.

Dalton, K.M., Nacewicz, B.M., Johnstone, T., Schaefer H.S., Gernsbacher, M.A., Goldsmith, H.H., Alexander, A.L., & Davidson, R.J. (2005). *Nature, 8*, 519-526

Driver, J. et al. (1999). Shared attention and the social brain: gaze perception triggers automatic visuospatial orienting in adults. *Visual Cognition, 6*, 509–540

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

Hood, B.M., Willen, J.D., & Driver, J. (1998). Adult's eyes trigger shifts of visual attention in human infants. *Psychological Science, 9*(2), 131-134.

Langton, S.R.H. and Bruce, V. (1999). Reflexive visual orienting in response to the social attention of others. *Visual Cognition, 6*, 541–568.

Pelphrey, K.A., Sasson, N.J., Reznick, S., Paul, G., Goldman, B.D., & Piven, J. (2002). Visual scanning of faces in autism. *Journal of Autism and Developmental Disorders, 32*(4), 249-261

Ristic, J., Friesen, C. K., & Kingstone, A. (2002). Are eyes special? It depends on how you look at it. *Psychonomic Bulletin & Review, 9*, 507-513.

Standing L. (1973). Learning 10000 pictures. *The Quarterly Journal of Experimental Psychology, 25*, 207-222.

DOI 10.16910/jemr.1.1.1

ISSN 1995-8692