The influence of social and symbolic cues on observers' gaze behaviour

Frouke Hermens
School of Psychology, University of Aberdeen, UK

Robin Walker
Department of Psychology, Royal Holloway University of
London, UK

Research has shown that social and symbolic cues presented in isolation and at fixation both have strong effects on observers, but it is unclear how cues compare when they are presented away from fixation and embedded in natural scenes. We here compare the effects of two types of social cue (gaze and pointing gestures) and one type of symbolic cue (arrow signs) on eye movements of observers under two viewing conditions (free viewing versus a memory task). The results suggest that social cues are looked at more quickly, for longer and more frequently than the symbolic arrow cues. An analysis of saccades departing from the cue suggests that the pointing cue leads to stronger cueing than the gaze and the arrow cue. While the task had only a weak influence on gaze orienting to the cues, stronger cue following was found for free viewing compared to the memory task.

Keywords: social cueing, symbolic cues, gestures

Several studies have suggested that social cues, such as eye-gaze or a pointing gesture, induce automatic shifts of attention in the observer (e.g., Burton, Bindemann, Langton, Schweinberger, & Jenkins, 2009; Driver et al., 1999; Friesen & Kingston, 1998). Furthermore, it has been shown that, when presented with a visual scene, observers show a strong tendency to look at the people in the scene (Birmingham, Bischof, & Kingstone, 2008, 2009a, 2009b; Fletcher-Watson, Findlay, Leekam, & Benson, 2008; Fletcher-Watson, Leekam, Benson, Frank, & Findlay, 2009). Combining these findings, it would be predicted that when observing a natural scene, the observer's automatic response would be to direct their gaze to people in the scene and, subsequently, to fixate on the 'looked-at' or 'pointed-at' object. In fact, some recent findings have suggested that such automatic following of social cues indeed takes place (Castelhano, Wieth, & Henderson, 2007; Fletcher-Watson et al., 2008, 2009; Kuhn, Tatler, & Cole, 2009; Zwickel & Võ, 2010). For example, when participants were presented with two images, left and right of fixation, of which one contained the image of a person, they often first looked at the image containing the person and, subsequently, the object looked at by the person in the photograph (Fletcher-Watson et al., 2008). Similarly, when presented with computer generated scenes, observers tended to follow the gaze direction of a person contained in the natural scene, but not the direction of a loudspeaker (Zwickel & Võ, 2010). Gaze following was also found for dynamic stimuli (video clips) when participants were watching a magic trick being performed (Kuhn et al., 2009), and when participants were watching a sequence of

photographs telling a story (Castelhano et al., 2007).

Viewing behaviour has, however, been found to depend on the cognitive task performed by the observer. Yarbus (1967), for example, (see also DeAngelus & Pelz, 2009; Tatler, Wade, Kwan, Findlay, & Velichkovsky, 2010) demonstrated that fixation patterns differ when the observer is asked to freely view a painting or to perform a particular task, such as to estimate the age of the people in the painting. Similarly, gaze-following behaviour was found during free-viewing (Fletcher-Watson et al., 2008; Zwickel & Võ, 2010), but this behaviour was less prominent when the task was to make a gender decision about the person in the image (Fletcher-Watson et al., 2008). Moreover, when asked to indicate on which side to pass an oncoming pedestrian, observers shifted their gaze away from the observed gaze direction, indicating that providing observers with a specific task can override any automatic tendencies to follow the gaze direction of another person (Nummenmaa & Hietanen, 2009a). Task effects on gaze behaviour were also found in a study in which participants were presented with photos of faces with different gaze and head directions. Analysis of the eye movements showed that the eye region was more often fixated when participants were asked to judge the direction of gaze than when they decided on the orientation of the head (Itier, Villate, & Ryan, 2007). Furthermore, when watching works of art in which the gaze direction of a centrally presented figure was modified, the observers' gaze direction was influenced by the central gaze cue, but only when performing a task with a social component (Dukewich, Klein, & Christie, 2008). These results suggest that the automatic tendency to fixate people and to follow their gaze may be overridden by task demands. Varying task demands while keeping all aspects of stimulus presentation identical, however, is difficult. In particular, matching the duration of presentation across tasks may not be feasible across all types of tasks. For example, in visual search eye movements related

to the search task itself are restricted to the moment until the target is found, after which participants are likely to switch to a viewing mode similar to free viewing. Eye movements early in the presentation interval may differ from those later, and it is therefore important to only compare tasks that have identical presentation durations.

Whereas the effects of gaze cues have been studied extensively, there are also other social cues, such as pointing gestures. Such cues have received less attention in the literature (however, see, e.g., Burton et al., 2009; Ivanoff & Saoud, 2009; Langton & Bruce, 2000; Sato, Kochiyama, Uono, & Yoshikawa, 2009). At fixation, one might expect similar cueing effects from gaze cues and pointing gestures. However, when presented (initially) in the periphery, the direction of someone's gaze may not be clearly visible, whereas a sense of direction from a pointing gesture might still be perceived. Evidence for such stronger attention shifts from peripherally presented pointing cues was indeed found (Burton et al., 2009), but the same research also showed that peripherally presented gaze cues provided by rotated heads (rather than averted eye gaze) have strong effects on the observer (see also Langton & Bruce, 2000). However, since the cues in these studies were presented in isolation, it is unclear whether similar results are found when the cues are embedded in a natural scene, where they may crowd in the background texture.

A further question is whether the seemingly automatic following of cues is restricted to social cues, such as eye-gaze or pointing gestures, or whether common directional cues, such as arrows, also result in eye movements towards the cued object. Although earlier investigations with foveally presented and isolated gaze and arrow cues have suggested that social cues have a privileged status in their ability to shift attention and gaze in the observer (e.g., Driver et al., 1999; Friesen, Ristic, & Kingstone, 2004), more recent work has shown that similar effects can be observed with symbolic cues (Kuhn & Kingstone, 2009; Tipples, 2008). The influence of symbolic cues may be different, however, when cues appear in a natural scene rather than in a simple array -more typical of studies on covert attention- (Posner, 1980; Posner & Cohen, 1984) both in their ability to attract and to direct attention. For example, a recent study by Birmingham et al. (2009a) found that arrows attracted the observer's gaze to a much smaller extent than did people located in a scene. Additionally, objects 'pointed-at' by an uncommon directional cue (i.e., a large standing loudspeaker) were fixated less often than those gazed-at by a person in the scene, even though both cues provided similar direction information and were of similar saliency (Zwickel & Võ, 2010). It therefore appears that observers seek out people in a natural scene, but do not automatically show the same gaze behaviour for symbolic cues of direction.

In the light of the above considerations, the present study has the following aims. Past studies have shown that observers fixate people in a scene (e.g., Birmingham et al., 2009a, 2009b). However, it is unclear whether cueing of direction by these people depends on the type of cue they provide. Our first aim is therefore to determine whether gaze following in natural scenes is similar for gaze cues (where the person in the scene looks at an object) and pointing gestures (with a person pointing to an object). Second, while studies have suggested that observers do not fixate (Birmingham et al., 2009a) or follow (Zwickel & Võ, 2010) cues not provided by another human being, this lack of attention to such cues may have been due to the cues being strongly embedded in the surrounding (as in traffic signs painted on the road or arrows placed on a sign with adjacent text, Birmingham et al., 2009a) or to the cues not being typical for directing attention (e.g., by loudspeakers, Zwickel & Võ, 2010). Our second aim is to compare cueing between social cues and arrows signs specifically placed in the scenes that may be more obvious to the observers. Finally, our third aim is to compare different tasks, but keep the presentation of the stimuli constant. To achieve this aim, we compared free viewing (for a fixed interval) with a memory task where participants received the same amount of time to memorize the scene for a subsequent memory test (determine whether a shown section was part of the previously seen image). Because the memory task requires to take in details about the image (particularly, because the test stimulus was selected at random from the original scene), observers may focus less on the cues and the cued objects than when freely viewing the images.

Participants were presented with photographic images of natural scenes, some of which contained a directional cue (a person looking or pointing at an object or an arrow directed towards an object). Half of the participants were asked to freely view the images for two seconds. The other half viewed the images for the same amount of time with the knowledge that a memory test would follow immediately after stimulus presentation. Four conditions were compared within the same set of scenes across the cues (Figure 1a). The images either contained (1) a gaze cue, consisting of a person directing eyes, head and body towards the cued object, (2) a 'pointing' cue, showing a person directing eyes, head and body towards the cued object, as well as pointing towards the object with their arm and hand, (3) an arrow cue, which was specifically placed in the scene, or (4) no cue at all.

On the basis of earlier results (Birmingham et al., 2008, 2009a, 2009b; Zwickel & Võ, 2010), we predict that observers will fixate the people in the scene, as well as the objects that were looked or pointed at by these people. On the basis of cues presented in isolation in the periphery (Burton et al., 2009; Langton & Bruce, 2000), we may expect that pointing gestures lead to stronger cueing than gaze cues. Second, on the basis of earlier findings showing that observers look at people, but not arrows (Birmingham et al., 2009a) and that people do not follow cues provided by non-humans (Zwickel & Võ, 2010), we predict that people in the scene attract the observers' eye gaze more often than arrows, and that gaze and pointing cues lead to more successive fixations on the cued object than the arrows do. Finally, based on past results (Dukewich et al., 2008; Yarbus, 1967), we predict that the task will influence gaze behaviour, with possibly fewer fixations on the cue and the cued object in the memory task.

Methods

Participants

Forty-eight participants, naive to the purpose of the study, took part in the experiment. Twenty-four of them performed the memory-based task and twenty-four the free viewing task. From the 24 participants in the memory-based task, data of three participants had to be removed, because of missing data occurring during eye tracking due to reflections from their glasses that interfered with accurate corneal reflection detection. In the free-viewing task, eye tracking in one participant failed due to issues with tracking the pupil centre. Participants were psychology students from the University of Aberdeen, taking part in return of course credit (the majority of the participants) or participating without receiving reimbursement. All participants reported normal or corrected-to-normal vision. They all signed informed consent for their participation in the study that was approved by the local ethics committee.

Apparatus

Data across the two tasks were collected on similar, but not identical setups (due to lab availability), that would not be expected to change the pattern of results. Eye movements in the memory-based task were collected using an Eyelink 1000 (SR Research) desk-mounted eye tracker (at 1000 Hz, $0.25°$ - $0.5°$ average accuracy, $0.01°$ RMS resolution), whereas for the free viewing task an Eyelink II system (at 250 Hz $0.5°$ average accuracy, $0.01°$ RMS resolution) was used. Both sampling rates are considered to suffice for the accurate detection of fixations, which are the units of analysis in the present study. Stimulus presentation in both setups was controlled by a PC running using the same software (Experiment Builder) under the same operating system (Windows XP). Stimuli were presented on 19 inch flat screens (different makes across setups, but with the same spatial resolution and refresh rates). Eye movements were tracked using the combined pupil and corneal reflection modes of the system, except for two participants in the free viewing task, where the pupil-only mode was used. The images used in the experiment were taken with a digital point-and-shoot camera (Canon Powershot A430).

Stimuli

A total of 88 photographic images was used, showing scenes of the Royal Holloway University campus and local residential areas. Images were unedited except for an occasional mirror reverse of the image (along the vertical midline) to balance out the number of images with the cue on the right and the left. The target subset of these images (examples in Figure 1a, note that these only show images with the author as the 'actor' providing the cues, whereas two more actors were included in the experiment) were constructed such that each of 12 scenes was presented with (1) a gaze cue, in which an actor was standing and looking at an object in the scene, (2) a pointing cue, in which an actor was standing and pointing at the same object in the scene, (3) a specifically placed arrow sign that was pointing at the same object in the scene, (4) the same scene without a cue. Five additional scenes had gaze, arrow and no-cue versions, but pointing cue versions of these stimuli were missing. The remaining scenes served as fillers, and were a mixture of pictures with arrow cues (mostly traffic signs), gaze cues, and scenes without cues. Three different female actors (author FH, whom participants had only met just before the experiment, and two people unknown to the participants; actors randomly distributed across the images) provided the social cues in the images. Note that in our gaze and pointing cues, the direction of eyes, head, body and pointing arm were all in the direction of the cued object, avoiding any incongruencies between the different aspects of cues. Pictures of the scenes were scaled down to a size of 1024x768 pixels and presented in a display of 1280x1024 pixels. For the memory task, sections of each of the images were created using a custom-built Matlab script, selecting a randomly placed area of 301x301 pixels from the scene.

Design

Half of the participants performed the free viewing task, whereas the other half of the participants performed the memory task, making task a between-subject factor, ensuring that each individual image was seen only once by each participant. Type of cue was presented as a within-subject factor, but by analyzing only the first presentation of each scene, it could also be treated as a between-subject factor (more about this in the data analysis section). The total list of items was divided in four blocks (unknown to the participants). Within each block, there were similar numbers of arrow, pointing, gaze, and no-cue images. The order of the trials within each block was randomized for each participant, and the order of the blocks was counter-balanced across participants by means of a Latin square. In the memory task, half of the images were paired with a section from the same image, and the other half with a randomly selected section from a different scene.

Procedure

Participants were tested individually. Before taking part, they signed an informed consent and received written and verbal instructions. They then took place in front of the computer screen, placing their chin on a chin rest to avoid head movements that may interfere with eye tracking. A nine-point calibration was performed, which was repeated until fixations were clearly aligned with the three by three grid on which the fixation targets were presented, after which experiment was started. Each trial started with a drift correction target placed outside the image at one of four positions, randomly selected on each trial (Figure 1c). The placement of the drift correction target ensured that participants fixated outside the image before stimulus onset, avoiding fixations on the centre of the image before at trial onset. Trials were initiated by the experimenter, confirming fixation on the drift correction target. The target image was presented for 2000

a) Stimuli

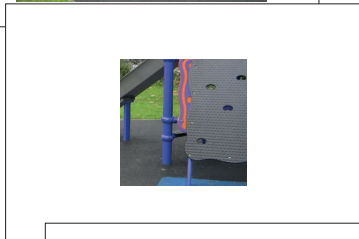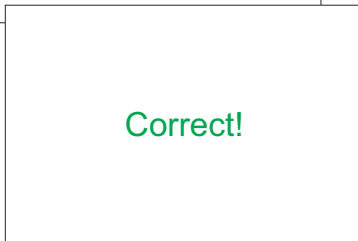Gaze cue

Pointing cue

Arrow cue
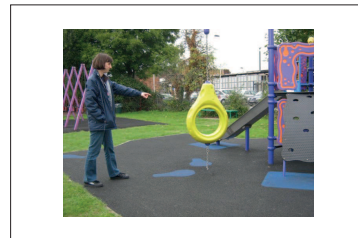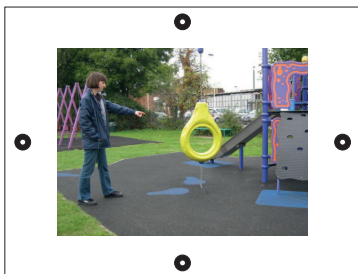
No cue

b) Tasks

Memory task

2000ms

Until keypress

Correct!

Free viewing

2000ms

d) Examples ROIs

Gaze

Pointing

Arrow

c) Possible fixation positions

Figure 1. (a) Examples of stimuli used in the experiment, with the same scene including a gaze cue, pointing cue, arrow, or no cue. The order of presentation of these different versions of the scenes was counter-balanced across participants. (b) Stimulus sequence for the two tasks. Each task was completed by a different group of participants. In both tasks the target image was presented for 2000ms. (c) Possible locations of the fixation stimulus (randomly chosen on each trial), which participants had to fixate in order to initiate the trial. (d) Examples of regions of interest for the eye movement analysis.

ms and followed, in the case of the memory task, by a section of an image taken from the target image or a other image (Figure 1b). This section was shown until participants pressed the key to indicate their response ('q' for not in the target image, 'p' for in the target image). In the memory task, feedback about the accuracy ('Correct!' in green or 'Incorrect' in red) was provided, before the next trial started. After the experiment, participants were debriefed about the purpose of the experiment and were given the chance to ask further questions.

Data analysis

The raw data was parsed into fixations and saccades using the automated algorithm of the Eyelink systems, using the default 30 deg/sec velocity and 8,000 deg/sec^2 acceleration criteria for saccade detection. Fixations were then assigned to their appropriate regions of interest. Regions of interest were defined as the area around the face and upper body for gaze cues, the face, the upper body and pointing arm for the pointing cue, and the arrow sign for the arrow cue (see Figure 1d for examples). Note that by this definition, the two social cues had very similar regions of interest, except for the area around the pointing arm in the pointing cue. The cues vary in the size of the region of interest, but analyses will be reported that compensate for these differences. Also illustrated are the regions of interest around the cued objects, where the same region of the scene was used in each of the four cueing conditions (meaning that the no-cue condition will show how often the cued object is fixated by itself). These cued object regions were placed around the object that the actor was instructed to point or look at (or the arrow sign set to point at). Results will be presented across all relevant data for all participants, treating cue as a within-subject (repeated measures) factor, and for all images with either all four cue conditions (12 scenes) or three of the four cue conditions (5 scenes, pointing cue not included, meaning that the average for the gaze, arrow and no-cue condition were based on 17 scenes, whereas those for the pointing cue were based on 12 scenes). Very similar results were obtained if only the twelve scenes with all four cues were used, or when only the first presentation of a scene for each participant was included (treating cue as a between-subject factor). Eye movement statistics were collected for each participant separately, and then compared across participants using univariate repeated measures ANOVAs using participants as a random factor. These tests were corrected using Greenhouse-Geisser corrections where appropriate, and followed up by pairwise t-tests (Bonferroni corrected p-values will be shown).

Results

Figure 2 plots the percentage of time spent fixating (Figures 2a and 2b) and the percentage of fixations (Figures 2c and 2d) on the different cues and the associated cued objects. The size of the differences in viewing times and number of fixations between cues depended on the task (interaction: (F(2,84)=7.49, p<0.001, η_p=0.15). However, for both tasks, the two social cues were fixated for longer than the arrow cue

(memory task: F(2,44)=59.9, p<0.001, η_p=0.73; free viewing: F(2,44)=57.5, p<0.001, η_p=0.74). For viewing times on the cued objects, the interaction between the cue and the task just fell short of significance (F(3,126)=2.62, p=0.054, η_p=0.059), but the effect of the cue on the total viewing time on the cued object was significant (F(3,126)=26.7, p<0.001, η_p=0.39). Paired comparisons across participants in both tasks showed that this was due to significant differences between each of the cues (all p-values <0.01; Bonferroni corrected for 6 comparisons), except for between the pointing and arrow cues (p=0.99).

Also for the percentage of fixations (Figures 2c and 2d), task interacted significantly with cue type (F(2,84)=9.46, p<0.001, η_p=0.18). However, for both tasks, the percentage of fixations was less for the arrow cue than for the social cues (memory task: F(2,44)=74.8, p<0.001, η_p=0.79; free viewing: F(2,40)=96.5, p<0.001, η_p=0.81; paired comparisons: p<0.001 for comparisons between the arrow and the social cues, p=0.072; Bonferroni corrected for 3 comparisons). The effect of the cue on the percentage of fixations on the cued object differed across tasks (interaction: F(3,126)=5.50, p<0.001, η_p=0.12), but the difference between cues was significant for both tasks (memory task: F(3,60)= 10.7, p<0.001, η_p=0.35; free viewing: F(3,66)=24.26, p<0.001, η_p=0.52). Paired comparisons for both tasks showed significant differences between all cue conditions, except between the pointing and the arrow cue, and between the pointing and the absent cue (Bonferroni corrected p-values for the remaining comparisons <0.018, based on the total of 6 comparisons).

These differences in fixations on the cue could be due to the difference in size of the cues. Figure 2e examines how cues compare when the viewing time is normalized for the area occupied by the cue relative to the size of the screen (Zwickel & Võ, 2010). This plot suggests that, given their small size, arrow cues were fixated for a relatively long time (F(1.27, 53.3)=40.3, p<0.001, η_p=0.49). This difference was not significantly influenced by the task (interaction: F(2,84)=2.46, p=0.091, η_p=0.055). Paired comparisons across data of both tasks showed significant differences between each of the cues (p-values <0.006; corrected for 6 comparisons). The same correction for object size can be used for cued objects, resulting in the relative viewing times shown in Figure 2f. The pattern of results in this figure deviates slightly from those in Figures 2a and 2b despite the same size of the cued objects across cues, because of differing relative contributions of objects of different sizes to the weighted viewing time. Relative viewing times differed significantly across cues (F(2.38,99.8)=34.5, p<0.001, η_p=0.45) and across tasks (F(1,42)=6.79, p=0.013, η_p=0.14), without an interaction between the two factors (F(1,42)=0.002, p=0.97). Comparisons between cues (with task as a between-subject factor) showed significantly longer viewing times for pointing cues compared to the other cues (both p-values <0.001), while there was no significant difference between the gaze and the arrow cue (F(1,42)=0.731, p=0.40).

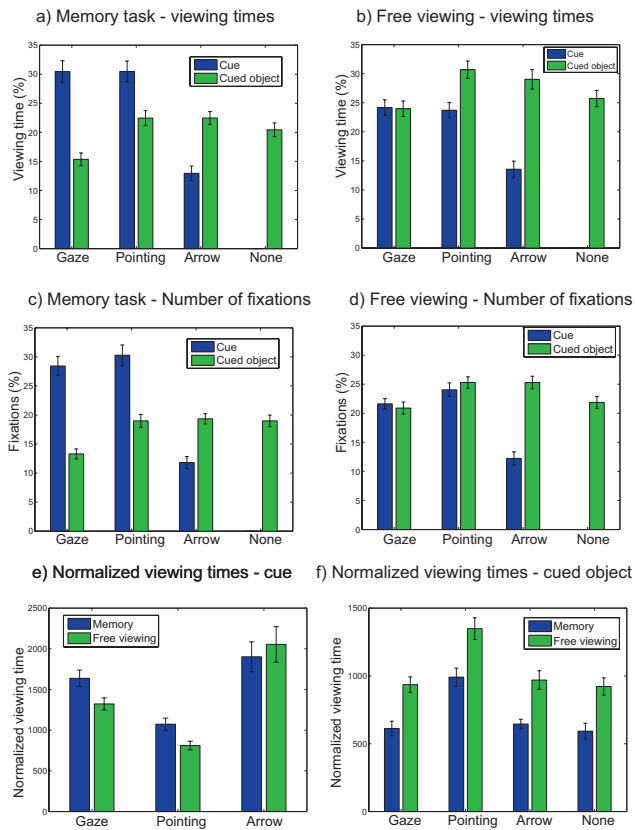These results are supported by more detailed measures of

Figure 2.   (a-b) Total viewing times and (c-d) number of fixations as a percentage of the total duration or number of fixations in the trial. Bars labelled as 'none' show the viewing times for scenes without a cue. This no-cue condition indicates how much time viewers spent on looking at the object otherwise cued in the versions of the scene with the cue. Left plots show the memory task, right plots the free viewing condition. e) Normalized viewing times on the cues, where each individual viewing time was multiplied by the area of the screen, divided by the area of the cue, giving higher weight to smaller cues that were fixated. Error bars show the standard error of the mean across participants.
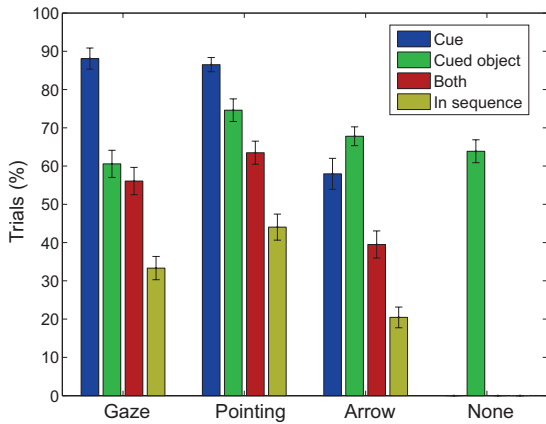
viewing behaviour, shown in Figure 3. The top row of this data plot shows the percentage of trials with (1) at least one fixation on the cue, (2) at least one fixation on the cued object, (3) at least one fixation on the cue and one fixation on the cued object, (4) at least one fixation on the cue immediately followed by a fixation on the cued object. As before, social cues were looked at more often than the arrow cues (memory task: $F(1.55, 31.0)=42.5$, $p<0.001$, $\eta_p=0.68$; free viewing: $F(1.42,31.2)=20.2$, $p<0.001$, $\eta_p=0.48$; paired samples showed significant differences between cues, $p<0.001$, except between the two social cues, $p=0.87$; 6 comparisons). Objects cued by gaze and pointing cues were fixated less often (memory task: $F(3,60)=7.24$, $p<0.001$, $\eta_p=0.27$; free viewing: $F(3,63)=11.8$, $p<0.001$, $\eta_p=0.36$). Combinations of fixations on cues and cued objects were also more frequent for social cues, either without taking the order of the fixations into account (paired samples t-tests comparing the different cues, p-values <0.001; 6 comparisons), or by only examining trials with a fixation on the cue immediately followed by a fixation on the cued object (paired samples t-tests, p-values <0.006 for all (six) comparisons).

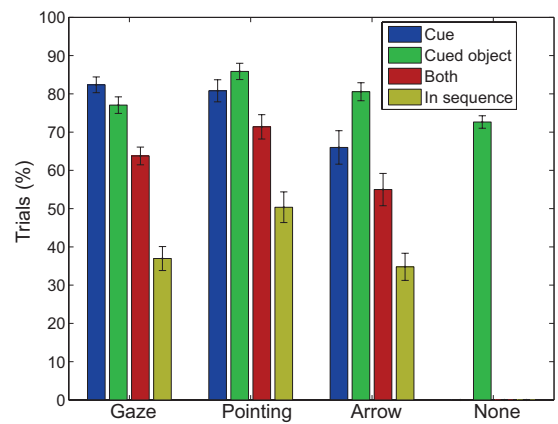The time course of fixations, shown in Figures 2c and 2d

as stacked histograms of cue, cued object and other region fixations, suggests that social cues were often fixated early in the trial, whereas for the symbolic arrow cue, participants tended to first fixate the cued object. A similar pattern of results is obtained when the time until the first fixation of the cue is considered (Figures 2e and 2f). On average, the social cues were fixated before their cued objects, but this was not the case for the arrow cues, where the object was looked at earlier than the cue (all p-values <0.001, 6 comparisons, but those for the arrow cues were associated with an opposite direction of the effect). These results confirm our assumption about the importance of matching the presentation duration across tasks.

The results so far indicate that social cues are fixated more often, but that the cued object is fixated in similar rates across cueing conditions (Figures 2a and 2b). In fact, it seems that the gaze cue may actually lead to less cueing than the other types of cues (Figures 2a and 2b). However, before such a conclusion can be drawn, one has to take into account that while observers spend time fixating the gaze cue itself, they cannot spend this same time fixating the cued object, which could explain the reduced cueing by the gaze cues
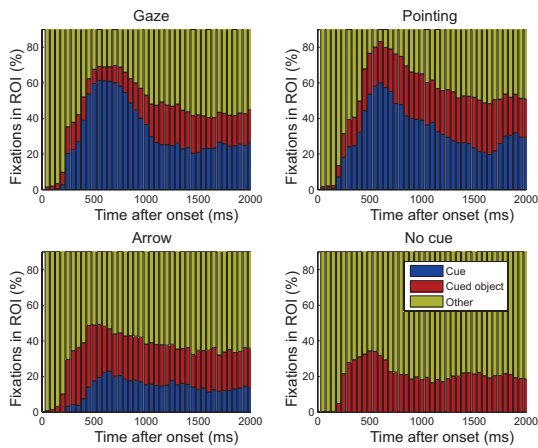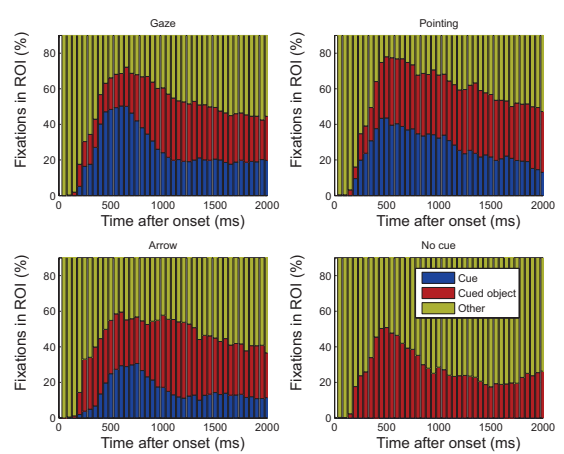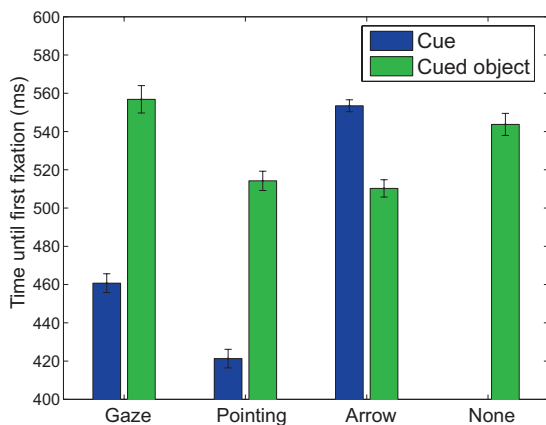
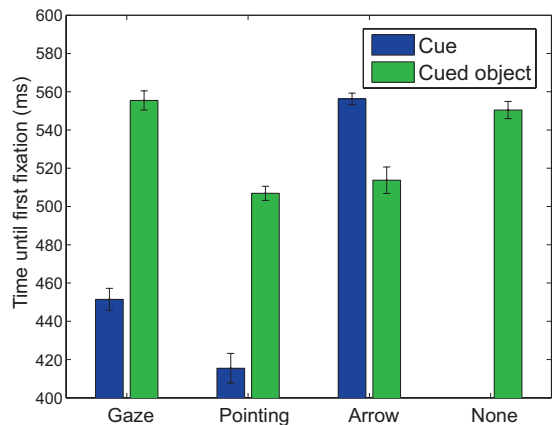*Figure 3.* (a-b) Percentage of trials with fixations on the cue, the cued object, on both the cue and the cued object, and these objects in sequence. (c-d) Histograms showing the percentage of participants fixating a certain region of interest in the scene across time. (e-f) Times until the first fixation of the cue and the cued object. Error bars show the standard error of the mean across participants.

simply from the fact that observers fixate the cue more often. Differences in viewing time of the different cues can occur for a range of reasons: due to a difference in the size of the cues in the image, due to differences in saliency, or due to differences in relevance for the goal of the observer. When using natural scenes (rather than relying on computer-generated images, e.g., Zwickel & Võ, 2010), it will be virtually impossible to fully control for such aspects, and therefore it is likely that viewing times on the cues will be different. To examine cue following independent of how often the cue is fixated, a measure is needed that provides the conditional probability of viewing the cued object, given that the cue is fixated first. For this, we use a measure similar to that used in Castelhano et al. (2007) and Zwickel and Võ (2010), illustrated in Figure 4a. In a first comparison, frequencies of refixations on the cue, saccades to the cued object and saccades to elsewhere are compared (Figure 4c and 4d). The percentage of saccades staying within the cue depends on the cue ($F(1.73, 72.2)=142.1$, $p<0.001$, $\eta_p=0.77$) and the task ($F(1,84)=1062.5$, $p<0.001$, $\eta_p=0.96$) independently (interaction: $F(2,84)=2.64$, $p=0.086$, $\eta_p=0.059$). Across both tasks, fewer refixations were found for arrows than for gaze ($t(43)=12.0$, $p<0.0001$) and pointing ($t(43)=14.5$, $p<0.0001$) cues. The difference between the two social cues is not significant ($t(43)=2.30$, $p=0.078$, Bonferroni corrected for three comparisons). Frequencies of saccades from the cue to the cued object also depend on the cue (main effect: $F(1.56, 65.3)=15.5$, $p<0.001$, $\eta_p=0.27$) and the task (main effect: $F(1,42)=17.8$, $p<0.001$, $\eta_p=0.30$; no interaction: $F(2,84)=1.02$, $p=0.37$, $\eta_p=0.024$). Across the two tasks, fewer object-directed saccades were found for gaze cues compared to pointing ($t(43)=4.00$, $p<0.001$) and arrow cues ($t(43)=5.93$, $p<0.001$). Between pointing and arrow cues there were no significant differences ($t(43)=2.05$, $p=0.14$, Bonferroni corrected).

These results suggest that saccades towards the cued object were more frequent for pointing and arrow cues, compared to gaze cues. However, gaze cues also showed more refixations, particularly when compared to arrow cues, and it may therefore be the case that object-directed saccades from such cues were prevented by additional refixations. Figure 4d therefore focuses on saccades made away from the cue, and plots which percentage of these saccades were directed towards the cued object (instead of to a location elsewhere in the scene). This plots shows that object directed saccades were more frequent when freely viewing the image ($F(1,42)=778.5$, $p<0.001$, $\eta_p=0.95$) compared to the memory task. It also shows significant differences between the three cues ($F(1.72,72.2)=9.95$, $p<0.001$, $\eta_p=0.19$). Paired comparisons across both tasks (corrected for 3 comparisons) shows stronger cueing for pointing cues compared to gaze cues ($t(43)=4.83$, $p<0.001$), stronger cueing for arrows compared to gaze cues ($t(43)=2.63$, $p=0.036$), but no difference in cueing between pointing and arrow cues ($t(43)=1.92$, $p=0.19$).

So far, analyses have focused on the entire cue regions (Figure 1d), pooling fixations across the eyes, head, upper body and arm (for pointing cues). In a final analysis, regions of interest were redefined for the gaze and pointing cues into an eye, a head (excluding the eyes), an upper trunk and - for the pointing cue- an arm region (Figure 5a). Figure 5b shows that for both types of cues (gaze and pointing), the majority of fixations were directed towards the head region, not including the eyes (gaze: difference between regions, across tasks: $F(1.6, 67.7)=56.3$, $p<0.001$, $\eta_p=0.57$, paired comparisons across tasks showed significant differences between all regions, $p<0.001$ for three comparisons; pointing: $F(2.2,90.3)=32.0$, $p<0.001$, $\eta_p=0.43$, paired comparisons across tasks showed significant differences between all regions, $p<0.001$ except between arm and upper trunk, $p=0.79$ for six comparisons). The distribution of fixations was unaffected by the task (gaze: main effect of task, $F(1,48)=0.89$, $p=0.35$; interaction with region, $F(1.6,67.7)=0.17$, $p=0.80$, pointing: main effect of task, $F(1,48)=0.001$, $p=0.98$, interaction with region, $F(2.2,90.3)=0.78$, $p=0.47$). The analysis of saccades leaving the cue (Figure 5c; similar logic as for Figure 4d) indicates that for gaze cues, saccades leaving the trunk area are significantly less often directed towards the cued object than saccades leaving the other areas ($F(1.7,72.5)=3.56$, $p=0.040$, $\eta_p=0.078$, no interaction with task, $F(1.7,72.5)=2.11$, $p=0.14$, $\eta_p=0.048$). No such difference between regions was found for the pointing cues ($F(3,126)=0.18$, $p=0.91$, no interaction with task, $F(3,126)=1.94$, $p=0.13$, and no effect of task, $F(1,42)=2.27$, $p=0.14$). Note that percentages of saccades directed to the cued object are relatively low across all types of cues (in the order of 50%), suggesting that cues are not strongly directing the observers' gaze.

For our analyses, we defined fairly tight regions of interest (ROIs) around the people, objects and parts of the scene serving as cues and cued objects. Recent work has suggested that the size of the ROIs may influence the results (Orquin, Ashby, & Clarke, 2015), and for this reason, we repeated our analyses to examine whether using wider ROIs influenced our pattern of results in any way. Figure 6a shows two examples of the original, narrow regions of interest, and the newly defined larger ROIs, which allow for the inclusion of fixations into the wider ROIs that may have occurred outside the narrow ROIs due to measurement error. Figure 6b shows that although percentages of fixations, trials, and saccades (data for three analyses shown) increase with wider regions of interest, the pattern of results in the mean data remains virtually identical.

Discussion

Studies of social attention have largely focused on cues presented in isolation and at the centre of fixation (e.g., Driver et al., 1999; Friesen & Kingston, 1998) (for a review, see Frischen, Bayliss, & Tipper, 2007). In the present study, we have tried to move away from this classic paradigm by presenting stimuli in different locations in the visual field and in a more natural context (see also Birmingham et al., 2009a; Dukewich et al., 2008; Kingstone, Smilek, & Eastwood, 2008; Kingstone, 2009). Cues presented in such a setting differ in at least four important aspects with respect
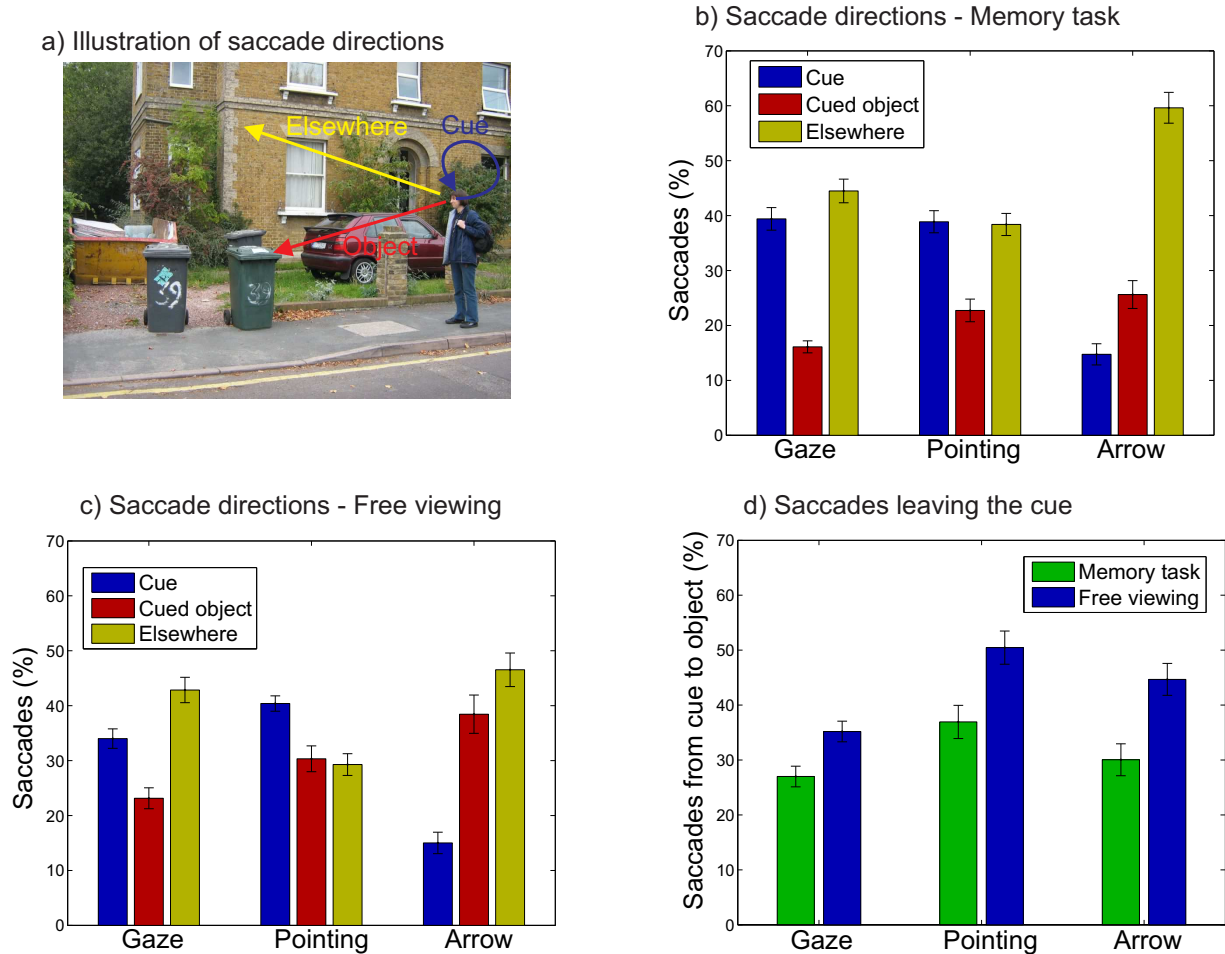
Figure 4. (a) Illustration of the analysis. All saccades leaving the cue the direction are examined, and the proportions of saccades landing on the cue, the cued object, or elsewhere are computed. (b-c) Percentage of saccades from the cue towards the cue, the cued object or to a location elsewhere in the scene. (d) From the saccades that leave the cue (i.e., do not return to the cue), the percentage of saccades that land on the cued object, for each cue and per task. Error bars show the standard error of the mean across participants.

to those presented in isolation and at fixation. First, cues embedded in natural scenes are often presented away from (initial) fixation (however, see Langton, O'Donnell, Riby, & Ballantyne, 2006). Research has suggested that cues that are effective at fixation may not be as influential when in the periphery (Burton et al., 2009; Langton & Bruce, 2000) (however, see Nummenmaa & Hietanen, 2009a). This effect was independent of whether participants were allowed to saccade towards the cue (initially presented away from fixation, Hermens & Bindemann, 2014). Second, cues in natural scenes are restricted in the location where they can occur (for example, a person is most likely to be positioned vertically, along the horizontal mid-line), and as a consequence, top-down influences may occur that are absent for isolated cues. Third, cues in natural scenes are shown against an often textured background, which may influence their effectiveness. For example, the cues in Burton et al. (2009) that led to strongest cueing (the pointing hand and rotated head cues) had distinct shape outlines when presented against an empty back-

ground. Shown against a complex background (e.g., a fence, or a tree) the cues' shapes may not be equally prominent, possibly affecting cueing. Fourth, cues provided by people in natural scenes are typically provided by the person's entire body (direction of the legs, the torso, the head and the eyes), whereas centrally presented cues often vary only in the direction of the eyes. For these reasons, it is important to determine whether cueing effects found for cues in isolation and at fixation extend to cueing in a natural context.

A number of studies have examined overt (eye movements) and covert attention for cues embedded in natural scenes. Birmingham and colleagues (Birmingham et al., 2008, 2009a, 2009b) found that observers look at people more than arrows embedded in the same scene (Birmingham et al., 2009a) and that fixation of people depends on whether these people are engaged in an interaction (Birmingham et al., 2008), while visual saliency was not a strong prediction of observers' eye movements (Birmingham et al., 2009b). Langton et al. (2006) examined the influence of cen-

(a) Examples of detailed regions of interest

(b) Fixations on the cue, split by region of interest
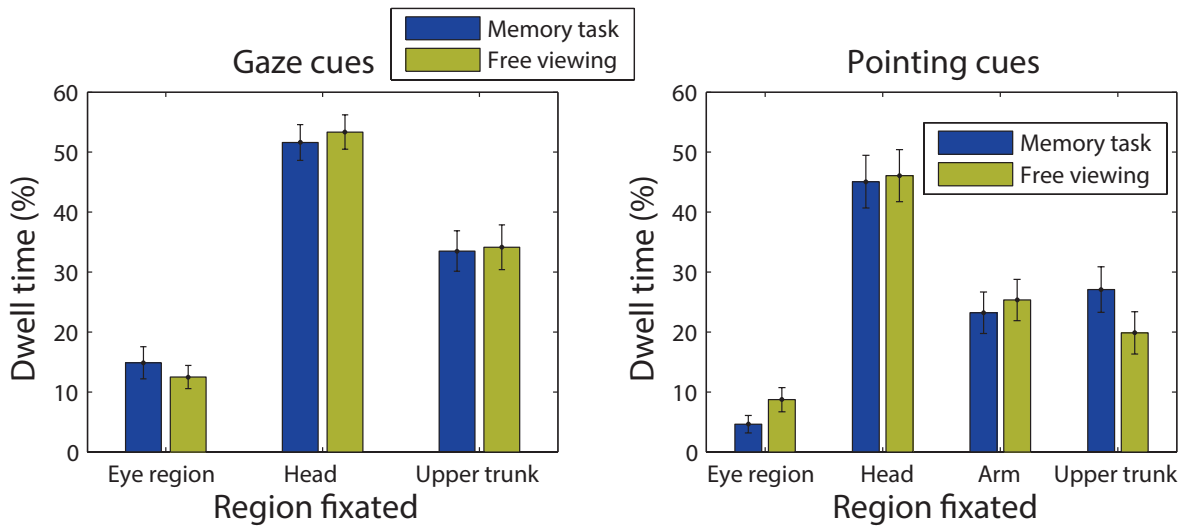
(c) Saccades leaving the cue

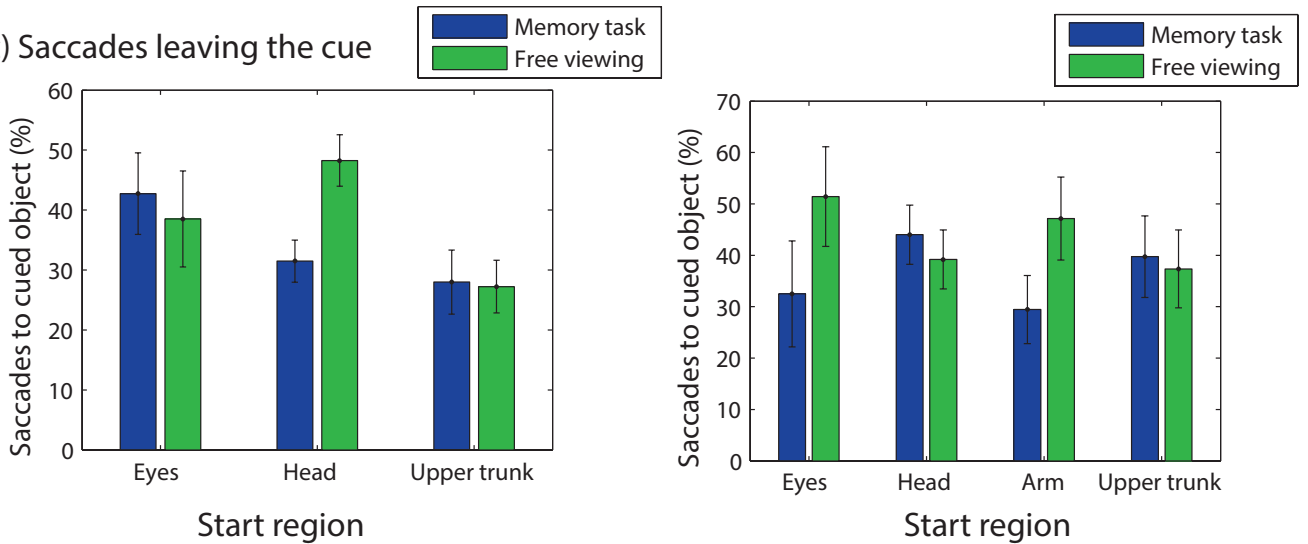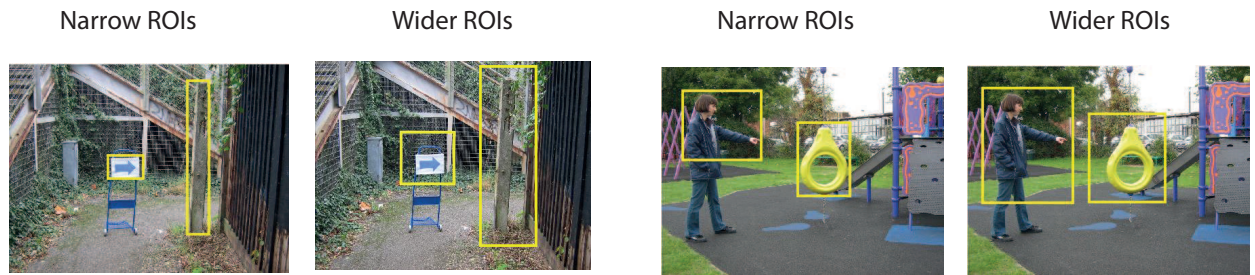*Figure 5*.   (a) Examples of regions of interest within the gaze and pointing cue, defining separate regions for eyes, head (eye region excluded), trunk and arm. (b) Percentage of the overall viewing times on the entire cue on the different regions, showing that the head (excluding the eye region) was relatively often fixated. (c) Percentage of saccades leaving the cue for the cued object, compared to fixations elsewhere in the image.

a) Examples narrow and wider ROIs

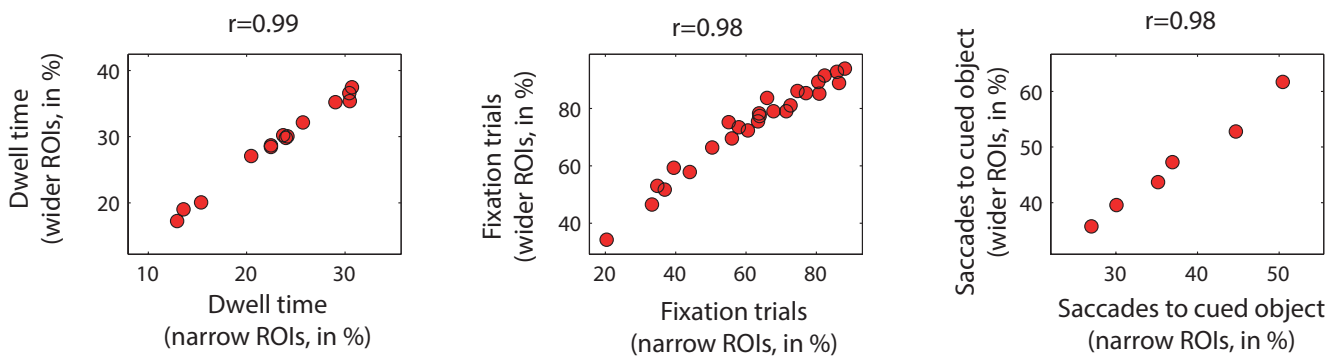## b) Comparison mean data narrow and wider ROIs

Figure 6. (a) Examples of tight regions of interest (ROIs) around the cues and cued objects, and wider ROIs for the same scenes. b) Comparison of the mean data for each of the conditions for three of our data plots, comparing viewing times (Figure 2a), trials with fixations on cues and cued objects (Figure 3a), and saccades leaving the cue for the cued object (Figure 5c). The values in the subplots' titles show the correlation between the mean data for the narrow and wider ROIs.

trally presented gaze cues embedded in natural scenes on observers' attention, and found strongest attention effects in the direction of perceived gaze. Castelhano et al. (2007) showed participants sequences of images telling a story and examined people's eye movements, demonstrating that eye movements to the cue were often followed by eye movements to the cued object. Kuhn et al. (2009) asked participants to understand a magic trick played out by an actor and found that observers were influenced by the gaze cues provided. Fletcher-Watson et al. (2008) examined eye movements of observers presented with pairs of images of which one contained an actor, and found that observers tended to fixate the image of the actor more often, and showed evidence of gaze-following once the actor was fixated. Dukewich et al. (2008) examined observers' eye movements while they watched works of arts in which the gaze direction (eyes) of a centrally presented character was manipulated under different instructions (free viewing, spatial memory instruction, social memory instruction). Gaze cueing was found only for the social instruction. Zwickel and Võ (2010) presented social (a person standing and gazing) and symbolic (a standing loudspeaker) cues in a rendered 3D scene and found that observers fixated the social, but not the symbolic cues, with gaze following occurring for the social cues. Taken together,

these studies suggest that people, but not other cues are fixated when observers look at natural scenes with these stimuli embedded in them. The studies also suggest that gaze cues are followed, but other types of cues (standing loudspeakers) are not. While most studies suggest that gaze cueing occurs under free viewing conditions, one study (Dukewich et al., 2008) suggests that gaze following only takes place during a social task.

While these studies provide compelling evidence for gaze cueing in natural scenes which is stronger than for other cues, there are limitations to what evidence the studies provide. In most studies only gaze cues were considered, and it is not clear whether cueing is restricted to these cues, or whether other types of cues (social and/or symbolic) lead to similar effects. Exceptions are the studies by Birmingham et al. (2009a) and Zwickel and Võ (2010), who compared people against arrows and people against standing loudspeakers, respectively. Birmingham et al. (2009a), however, only examined observers' gaze behaviour towards the cues (not the cued objects), and often paired the arrow cues with sections of text, which may have led to people favouring the text over the arrow cues. Zwickel and Võ (2010) relied on computer generated images and may have used a control cue that may not provide a strong sense of direction (a stand-

ing loudspeaker). Several studies only considered centrally presented cues (Dukewich et al., 2008; Kuhn et al., 2009; Langton et al., 2006) and results may not extend to situations where the cues are not presented at fixation (as seems to be the case for isolated cues, e.g., Burton et al., 2009). In some of the studies motion cues may have played a contributing role to the cueing effects (Castelhano et al., 2007; Kuhn et al., 2009). Not all studies have used photographs (Dukewich et al., 2008; Zwickel & Võ, 2010), which may limit the possible conclusions that can be drawn. The role of context and task was studied in only two of the studies (Birmingham et al., 2008; Dukewich et al., 2008), but both suggest that social context is important. However, other studies have found fixations on people and cued objects when no social context or task was provided (e.g., Fletcher-Watson et al., 2008; Zwickel & Võ, 2010). None of the studies have compared gaze cues with other social cues, such as pointing gestures.

The present study was designed to resolve these possible issues, and compared the effects of two social cues (gaze and pointing gestures) with a symbolic cue (an arrow sign) embedded in natural scenes. Same scene images without any cues were also included to examine the saliency of the otherwise cued objects without a cue. We used unedited photographs (with the exception of an occasional left-right flip) with a fair bit of background and actors and objects placed predominantly away from the centre of the image. Two tasks were compared, with identical presentation times of the stimuli (to avoid any differences between tasks that rely solely on a different distribution of fixations across the presentation interval), which varied only in the intend with which participants viewed the images. We analyzed observers' eye movements for fixations on the cues, the cued objects, and elsewhere in the scenes.

Our findings show that social cues are looked at more frequently than symbolic cues (arrows), in agreement with earlier findings (Birmingham et al., 2009a; Fletcher-Watson et al., 2009; Zwickel & Võ, 2010). Within the social cues, observers most often looked at the head of the actor, but not specifically at the eyes region (possibly because it spanned only a small area, or because eye-gaze is less informative in extrafoveal vision, Burton et al., 2009). Comparison of the directions of saccades leaving the cues demonstrate that pointing cues lead to stronger cueing than the other two types of cues (gaze and arrow). This is in agreement with past studies of peripheral cueing, where strongest cueing was found for pointing hands, but only slightly weaker cueing for gaze cues provided by turned heads (Burton et al., 2009; Langton & Bruce, 2000). This suggests that similar cueing effects are obtained for cues presented in isolation and cues embedded in natural scenes, as long as the cues are in the same relative location (at or away from fixation). The fact that cues are fixated after first having been seen in extrafoveal vision does not seem to influence their effectiveness (see also, Hermens & Bindemann, 2014). Cueing was stronger from the eyes and head region than from the trunk of the actors, but only for the gaze cues. Significant differences between viewing times for social and symbolic cues were obtained, but the ob-

served differences were not as substantial as in previous studies (Birmingham et al., 2009a; Zwickel & Võ, 2010). There are several possible reasons. For example, Birmingham et al. (2009a) presented the symbolic cues (arrow cues) in their images almost uniquely in conjunction with text. Text is known to strongly attract people's attention (Rayner, Rotello, Stewart, Keir, & Duffy, 2001; Rayner, Miller, & Rotello, 2008), and because of this, people's gaze might have been distracted from the arrows. Zwickel and Võ (2010) used a standing loudspeaker as their symbolic cue. This was done on purpose, to avoid the strong association of direction associated with arrows, but it might also explain why we found stronger attraction to the symbolic cues (arrows) in our study. The intend with which observers viewed the images (for a memory task, or without such a task) did not influence the frequency with which they looked at the cues. After having looked at the cues, observers looked at the cued object more often without the memory task. This suggests that cues attract observers' attention independent of the task, but that cueing is dependent on the task.

Overall, cueing effects were fairly week, with only around 50% of saccades leaving the cue targeted to the cued object, suggesting that observing someone's gaze or pointing gesture does not automatically make the observer look in the same direction. Our percentage of outgoing saccades is difficult to compare to past results, and it is therefore unclear whether cueing in our experiment was exceptionally low. For example, Zwickel and Võ (2010) report the number of saccades rather than a percentage, Fletcher-Watson et al. (2008) provide overall statistics on outgoing saccades but no numbers, and Castelhano et al. (2007) report outgoing saccades in terms of their direction rather than whether they were directed to the cued object, so it is unclear whether our 50% is low in the context of these studies. The image provided by Kuhn and Findlay (2010) suggests that not all observers follow the direction of gaze of their actor (Figure 2 in their paper), but the image in Kuhn, Tatler, Findlay, and Cole (2008) (Figure 2) suggests much stronger cueing. Possibly the task plays a role. Neither of our two tasks had a social context, which has been shown to influence fixations of people in natural scenes (Birmingham et al., 2008) and gaze following (Dukewich et al., 2008).

Past studies have relied on stimuli presented in isolation for good reasons. The use of natural scenes inherently causes difficulties in controlling stimulus parameters. In our experiment we achieved control for several of these parameters. For example, the background for each cueing condition was identical and the presentation duration of the stimuli was the same for the two tasks that we used. However, we did not manage to balance the retinal size of the different cues. The arrow signs in our experiment were smaller than the social gaze and pointing cues. One may therefore argue that the two social cues may have been looked at more often and for longer simply because they occupied a larger region of the image. Indeed, when weighting fixation durations by the area occupied by the cue, the arrow cue came out as the cue looked at for longest. The strength of cueing, however, could be assessed independent of the size of the stimuli by analyz-

ing saccades leaving the cues only (see also Castelhano et al., 2007; Zwickel & Võ, 2010). Another possible issue with using (3D) natural scenes is that maybe for the observers it is not always clear what object the actor or arrow sign is pointing at. Past work, in which movable arrows were placed in a room and in which observers were asked to rotate the arrow so that it pointed at a target object, has shown that observers can make large systematic errors in judging pointing direction (e.g., Doumen, Kappers, & Koenderink, 2006). To avoid such errors as much as possible, we photographed the actors and arrow signs as much as possible from the side, avoiding the 3D aspect in the pointing direction as much as possible. A final note concerns the relatively impoverished scenes that we used. In our experiment, and in agreement with most past studies (Castelhano et al., 2007; Fletcher-Watson et al., 2008; Kuhn & Kingstone, 2009; Zwickel & Võ, 2010) we used scenes with a single individual and without people when an arrow sign was used. This may have made the people and arrow signs in our scenes more prominent, and one may question whether similar cueing effects are obtained when people providing such cues are embedded in a scene with more individuals. Future studies should address this possible issue by varying the number of people in the scene, possibly placing cues in conjunction (e.g., a group of people looking at an object) or in competition (e.g., examining whether observers are more likely to follow arrow signs or an actor pointing).

In conclusion, while social cues embedded in natural scenes were looked at more often than the symbolic (arrow) cue, cueing was stronger only for the pointing cue. The task mainly influenced cueing, with stronger cueing when freely viewing the images. Future work could examine the joint influence of multiple cues in a scene, or the role of cueing in dynamic scenes (video clips).

References

Birmingham, E., Bischof, W. F., & Kingstone, A. (2008). Social attention and real-world scenes: The roles of action, competition and social content. *Quarterly Journal of Experimental Psychology*, *61*, 986–998. doi:10.1080/17470210701410375.

Birmingham, E., Bischof, W. F., & Kingstone, A. (2009a). Get real! Resolving the debate about equivalent social stimuli. *Visual Cognition*, *17*(6), 904–924. doi: 10.1080/13506280902758044.

Birmingham, E., Bischof, W. F., & Kingstone, A. (2009b). Saliency does not account for fixations to eyes within social scenes. *Vision Research*, *49*, 2992–3000. doi:10.1016/j.visres.2009.09.014.

Burton, A. M., Bindemann, M., Langton, S. R., Schweinberger, S. R., & Jenkins, R. (2009). Gaze perception requires focused attention: Evidence from an interference task. *Journal of Experimental Psychology: Human Perception & Performance*, *35*(1), 108-118. doi: 10.1037/0096-1523.35.1.108.

Castelhano, M. S., Wieth, M., & Henderson, J. M. (2007). I see what you see: Eye movements in real-world scenes are affected by perceived direction of gaze. In *Attention in cognitive systems. theories and systems from an interdisciplinary viewpoint* (pp. 251–262). Springer.

DeAngelus, M., & Pelz, J. (2009). Top-down control of eye movements: Yarbus revisited. *Visual Cognition*, *17*, 790–811. doi: 10.1080/13506280902793843.

Doumen, M. J., Kappers, A. M., & Koenderink, J. J. (2006). Effects of context on a 3d pointing task. *Journal of Vision*, *6*(6), 728–728. doi:10.1068/p5550.

Driver, J., Davis, G., Kidd, P., Maxwell, E., Ricciardelli, P., & Baron-Cohen, S. (1999). Shared attention and the social brain: Gaze perception triggers automatic visuo-spatial orienting in adults. *Visual Cognition*, *6*, 509–554. doi: doi=10.1.1.212.3316.

Dukewich, K. R., Klein, R. M., & Christie, J. (2008). The effect of gaze on gaze direction while looking at art. *Psychonomic Bulletin & Review*, *15*(6), 1141–1147. doi: 10.3758/PBR.15.6.1141.

Fletcher-Watson, S., Findlay, J. M., Leekam, S. R., & Benson, V. (2008). Rapid detection of person information in a naturalistic scene. *Perception*, *37*, 571–583. doi: doi:10.1068/p5705.

Fletcher-Watson, S., Leekam, S. R., Benson, V., Frank, M. C., & Findlay, J. M. (2009). Eye-movements reveal attention to social information in autism spectrum disorder. *Neuropsychologia*, *47*, 248–257. doi: doi:10.1016/j.neuropsychologia.2008.07.016.

Friesen, C. K., & Kingston, A. (1998). The eyes have it! Reflexive orienting is triggered by nonpredictive gaze. *Psychonomic Bulletin and Review*, *5*(3), 490–495. doi: 10.3758/BF03208827.

Friesen, C. K., Ristic, J., & Kingstone, A. (2004). Attentional effects of counterpredictive gaze and arrow cues. *Journal of Experimental Psychology: Human Perception & Performance*, *30*, 319–329. doi: 10.1037/0096-1523.30.2.319.

Frischen, A., Bayliss, A. P., & Tipper, S. P. (2007). Gaze cueing of attention: visual attention, social cognition, and individual differences. *Psychological bulletin*, *133*(4), 694.

Hermens, F., & Bindemann, M. (2014). Social cueing in peripheral vision. *Paper presented at the EPS conference in Newcastle, UK*.

Itier, R. J., Villate, C., & Ryan, J. D. (2007). Eyes always attract attention but gaze-orienting is task dependent: evidence from eye movement monitoring. *Neuropsychologia*, *5*, 1019–1028. doi: 10.1016/j.neuropsychologia.2006.09.004.

Ivanoff, J., & Saoud, W. (2009). Nonattentional effects of non-predictive central cues. *Attention, Perception & Psychophysics*, *71*(4), 872–880. doi: 10.3758/APP.71.4.872.

Kingstone, A. (2009). Taking a real look at social attention. *Current Opinion in Neurobiology*, *19*, 52–56. doi: 10.1016/j.conb.2009.05.004.

Kingstone, A., Smilek, D., & Eastwood, J. D. (2008). Cognitive Ethology: A new approach for studying human cognition. *British Journal of Psychology*, *99*, 317–340. doi: 10.1348/000712607X251243.

Kuhn, G., & Findlay, J. M. (2010). Misdirection, attention and awareness: inattentional blindness reveals temporal relationship between eye movements and visual awareness. *Quarterly Journal of Experimental Psychology*, *63*(1), 136–146. doi: 10.1080/17470210902846757.

Kuhn, G., & Kingstone, A. (2009). Look away! Eyes and arrows engage oculomotor responses automatically. *Attention, Perception & Psychophysics*, *71*, 314–327. doi: 10.3758/APP.71.2.314.

Kuhn, G., Tatler, B. W., & Cole, G. (2009). You look where I look! Effect of gaze cues on overt and covert attention in misdirection. *Visual Cognition*, *17*(6-7), 925–944. doi: 10.1080/13506280902826775.

Kuhn, G., Tatler, B. W., Findlay, J. M., & Cole, G. G. (2008). Misdirection in magic: Implications for the relationship between eye gaze and attention. *Visual Cognition*, *16*(2-3), 391–405. doi: 10.1080/13506280701479750.

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 & Performance*, *26*(2), 747–757. doi: 10.1037/0096-1523.26.2.747.

Langton, S. R., 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*(12), 2056–2064. doi: 10.1080/17470210600917884.

Nummenmaa, L., & Hietanen, J. K. (2009a). How attentional systems process conflicting cues? The superiority of social over symbolic orienting revisited. *Journal of Experimental Psychology: Human Perception & Performance*, 1738–1754. doi: 10.1037/a0016472.

Orquin, J. L., Ashby, N. J., & Clarke, A. D. (2015). Areas of interest as a signal detection problem in behavioral eye-tracking research. *Journal of Behavioral Decision Making*.

Posner, M. I. (1980). Orienting of attention. *Quarterly Journal of Experimental Psychology*, *32*, 3–25. doi: 10.1080/00335558008248231.

Posner, M. I., & Cohen, Y. (1984). Components of visual orienting. In H. Bouma & D. Bouwhuis (Eds.), *Attention and Performance Vol. X, Erlbaum* (p. 531-556). Eindhoven, The Netherlands: Institute for Perception Research IPO.

Rayner, K., Miller, B., & Rotello, C. M. (2008). Eye Movements When Looking at Print Advertisements: The Goal of the Viewer Matters. *Applied Cognitive Psychology*, *22*, 697–707. doi: 0.1002/acp.1389.

Rayner, K., Rotello, C. M., Stewart, A. J., Keir, J., & Duffy, S. A. (2001). Integrating text and pictorial information: Eye movements when looking at print advertisements. *Journal of Experimental Psychology: Applied*, *7*, 219–226. doi: 10.1037//1076-898X.7.3.219.

Sato, W., Kochiyama, T., Uono, S., & Yoshikawa, S. (2009). Commonalities in the neural mechanisms underlying automatic attentional shifts by gaze, gestures, and symbols. *Neuroimage*, *45*(3), 984–992. doi: 10.1016/j.neuroimage.2008.12.052.

Tatler, B. W., Wade, N. J., Kwan, H., Findlay, J. M., & Velichkovsky, B. M. (2010). Yarbus, eye movements, and vision. *i-Perception*, *1*(1), 7–27. doi: 10.1068/i0382.

Tipples, J. (2008). Orienting to counterpredictive gaze and arrow cues. *Perception & Psychophysics*, *70*, 77–87. doi: 10.3758/PP.70.1.77.

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

Zwickel, J., & Võ, M. L.-H. (2010). How the presence of humans biases gaze movements. *Psychonomic Bulletin & Review*, *17*(2), 257–262. doi: 10.3758/PBR.17.2.257.