Humans as social beings rely on information provided by conspecifics. One important signal in social communication is eye gaze. The current study \((n = 93)\) sought to replicate and extend previous findings of attentional guidance by eye gaze in complex everyday scenes. In line with previous studies, longer, more and earlier fixations for objects cued by gaze compared to objects that were not cued were observed in free viewing conditions. To investigate how robust this prioritization is against top-down modulation, half of the participants received a memory task that required scanning the whole scene instead of exclusively focusing on cued objects. Interestingly, similar gaze cueing effects occurred in this group. Moreover, the human beings depicted in the scene received a large amount of attention, especially during early phases of visual attention, even though they were irrelevant to the current task. These results indicate that the mere presence of other human beings, as well as their gaze orientation, have a strong impact on attentional exploration.

Human beings are social beings who rely on information provided by conspecifics. One important signal in social communication is eye gaze. The current study \((n = 93)\) sought to replicate and extend previous findings of attentional guidance by eye gaze in complex everyday scenes. In line with previous studies, longer, more and earlier fixations for objects cued by gaze compared to objects that were not cued were observed in free viewing conditions. To investigate how robust this prioritization is against top-down modulation, half of the participants received a memory task that required scanning the whole scene instead of exclusively focusing on cued objects. Interestingly, similar gaze cueing effects occurred in this group. Moreover, the human beings depicted in the scene received a large amount of attention, especially during early phases of visual attention, even though they were irrelevant to the current task. These results indicate that the mere presence of other human beings, as well as their gaze orientation, have a strong impact on attentional exploration.

Humans in their social environment rely on the information conspecifics provide. This does not only hold for reading explicit signals, such as verbal communication, but also for implicit signals, such as eye gaze or nonverbal cues. Specifically, if an individual looks into a certain direction, this information is often read spontaneously by an observer who redirects his or her attention towards the referred object or location. Such guidance of someone else’s attention is called gaze following. As a consequence, joint attention is established.

The most frequently used paradigm to investigate such attentional shifts is the so-called gaze cueing paradigm (Driver et al., 1999; Friesen & Kingstone, 1998; Langton & Bruce, 2000; for a review see Frischen et al., 2007). This paradigm has been inspired by classical studies on spatial attention by Posner (1980) and consists of a centrally presented face with varying gaze directions. This face is then followed by a subsequently presented target at either the cued location (i.e., the location that the face is looking at) or an uncued location (i.e., a location that is not being looked at by the face). Studies using this gaze cueing paradigm have demonstrated that gaze cues facilitate target processing as evident in smaller reaction times to targets at cued as compared to uncued locations (Frischen et al., 2007). The paradigm was also used to show that gaze following is shaped by high-level social cognitive processes like group identity (Liuzza et al., 2011), theory-of-mind (Cole et al., 2015; Teufel et al., 2009; Wiese et al., 2012; Wykowska et al., 2014) or physical self-similarity (Hungr & Hunt, 2012).

However, even though gaze cues are crucial for joint attention, this standard gaze cueing paradigm can be criticized for lacking ecological validity. Whereas in the real world, gaze signals occur within a rich context of competing visual information, gaze cueing studies typically used isolated heads (Friesen & Kingstone, 1998; Langton & Bruce, 2000) or even cartoon faces (Driver et al., 1999; Ristic & Kingstone, 2005) as gaze cues (for an overview see: Risko et al., 2012). Although gaze cueing was also found with more naturalistic stimuli (Perez-Osorio et al., 2015), in a recent study in which Hayward et al. (2017)
compared attentional measures of gaze following from laboratory (classical gaze cueing) and real world (real social engagement) settings, they did not find reliable links between those measures.

As a compromise between rich but also less controlled field conditions and standardized but impoverished laboratory studies, complex naturalistic scenes were used to investigate gaze behaviour in laboratory settings (e.g., Fletcher-Watson et al., 2008; Perez-Osorio et al., 2015; Zwickel & Võ, 2010). To specifically explore the influence of gaze cues, Zwickel and Võ (2010) and Perez-Osorio et al. (2015) used pictures of a person (instead of isolated heads or faces) as a directional cue within a naturalistic scene. Zwickel and Võ (2010; in contrast to the gaze cueing task chosen by Perez-Osorio et al., 2015) used a free viewing instruction, meaning that participants had no explicit task to fulfill but should just freely explore the pictures. The authors argued that the lack of a specific task puts gaze following to a stricter test since previous studies frequently used target detection tasks (e.g., Langton et al., 2017) or comprised specific instructions such as asking participants to understand a scene (Castelhano et al., 2007). Consequently, in those latter studies, it remains unclear to what degree gaze following occurred spontaneously or was caused by the specific task at hand. In detail, Zwickel and Võ (2010) presented participants multiple 3D rendered outdoor and indoor scenes for several seconds that always included two clearly visible objects as well as either a person or a loudspeaker that was directed towards one of these objects. The loudspeaker, which also represents an object with a clear spatial orientation, served as a control condition to ensure that gaze cueing effects are due to the social meaning (i.e., the direction of the depicted person’s gaze) as compared to a mere following of any directional cue. The results of the study showed that participants fixated the cued object remarkably earlier, more often and longer than the uncued object. By showing that leaving saccades from the head most often landed onto the cued object, the results gave further evidence for the direct influence of eye gaze on attentional guidance. Crucially, similar effects were not obtained for the loudspeaker. The cued objects were not just focused because they might have been salient by themselves (e.g., due to positioning), or because they were cued by another object, but became more salient by the person’s reference. To sum up, Zwickel and Võ (2010) provide convincing evidence that joint attention is a direct consequence of gaze cues and gaze following, it happens spontaneously and has high relevance even in situations that are more naturalistic (i.e., involve complex scenes and the absence of explicit tasks) than classical gaze cueing studies based on variations of the Posner paradigm.

In the current study, we were first interested in whether the previously reported effects hold when using a different set of stimuli. Replication in itself is a core concept of scientific progress (Schmidt, 2009) and thus relevant for assessing the stability of effects. Nevertheless, our motivation was also to improve certain aspects of the study and at the same time extending this line of research. Due to their low resolution and reduced richness of details, the 3D rendered scenes used by Zwickel and Võ (2010) did not allow for an assessment of the depicted person’s gaze direction. As a consequence, the observed cueing effects could be due to directional information inferred from both the body and head of the person. We, therefore, developed a new set of photographic stimuli that had sufficient resolution to also allow for perceiving gaze direction with clearly visible eyes of the depicted person. These photos always included a human being who directed his/her gaze towards one of two objects that were placed within reaching distance. In order to be consistent with the study of Zwickel and Võ (2010), the depicted person’s head and body were congruently aligned with his/her eye gaze. Second, in order to extend this line of research, we manipulated top-down attentional processes by task instruction to explore the susceptibility of gaze following effects in naturalistic scenes. Earlier research showed that social attention can be influenced by multiple factors like social status of the observed persons (Foulsham et al., 2010), possibility to interact (Hayward et al., 2017; Laidlaw et al., 2011) or social content (Birmingham et al., 2008b). Together with Zwickel and Võ (2010), these studies have in common that they manipulate viewing behaviour of the participant by manipulating the stimuli or environment. In contrast, in the present study, we tried to modulate viewing behaviour via task instructions (for a similar procedure see Flechsenhar & Gamer, 2017). Specifically, half of the participants received an instruction before the viewing task, that they should try to remember as many objects from the scenes as possible (explicit
encoding group). The other half of the participants (free viewing group) merely received the instruction to freely explore the pictures and the memory test that was accomplished after the experiment was unannounced and therefore reflected spontaneous encoding of the respective scene details. The motivation for this manipulation was twofold. First, it was thought to test the robustness of gaze following against top-down processes by discouraging observers to utilize the information provided by eye gaze. Second, it allowed for examining gaze following effects on memory.

We expected to replicate the findings of Zwickel and Võ (2010) in the free viewing group. Specifically, we anticipated to observe an early fixation bias towards cued objects, an enhanced exploration of these details (i.e., more fixations and longer dwell times) and more saccades leaving the head towards the cued as compared to the uncued object. The instruction in the explicit encoding group was thought to induce a more systematic exploration of the presented scenes resulting in higher prioritization of both objects and reduced cueing effects. Furthermore, we anticipated a generally enhanced recall performance in the explicit encoding group. Due to the expected difference in attentional resources spent on the cued and uncued object in the free viewing group, memory performance of the cued object was expected to be better compared to memory performance of the uncued object. Finally, as previous studies showed a strong preference of fixating the head over body and background regions in static images (End & Gamer, 2017; Freeth et al., 2013), we expected to see a similar bias in the current study regarding dwell times, number of fixations and fixation latency. Additionally, we hypothesized that the prioritization for the head decreases when participants follow specific exploration goals such as in the explicit encoding group of the current study (cf., Flechsenhar & Gamer, 2017).

Methods

Participants

The cueing effects in fixations and saccades that were obtained by Zwickel and Võ (2010) can be considered large (Cohen’s $d_z > 0.70$). However, since effects of the top-down modulation implemented in the current study might be smaller, we used a medium effect size for estimating the current sample size. When assuming an effect size of Cohen’s $f = 0.25$ at an $\alpha$ level of .05 and a moderate correlation of .40 between factor levels of the within-subjects manipulation object role (cued vs. uncued), a sample size of 66 participants is needed to reveal main effects of the object role or interaction effects between group and object role at a power of .95. Under such conditions, the power for detecting main effects of group is smaller ($1 - \beta = .67$). As a compromise, we aimed at examining 90 participants (plus eventual dropouts) to achieve a power of .80 for the main effect of group and > .95 for main and interaction effects involving the within-subjects manipulation object role.

Finally, 94 subjects participated voluntarily. All participants had normal or corrected-to-normal vision and were recruited via the University of Würzburg’s online subject pool or by blackboard. Participants received course credit or a financial compensation of 5€. All participants gave written informed consent. One participant was excluded due to problems with the eye-tracking data acquisition, resulting in a final sample of $n = 93$ for the analysis with 64 female and 29 male participants between 18 and 55 years ($M = 24.75$ years, $SD = 5.06$ years). Overall, participants scored very low for autism traits in the Autism-Spectrum Quotient scale (AQ-k, German version, Freitag et al., 2007, Range 0–23, $M = 5.75$, $SD = 3.69$). In the final sample, one participant had an overall score higher than 17 which might reflect the presence of an autistic disorder. However, since we did not specify an exclusion criterion regarding AQ-k values beforehand, we decided to keep this participant in the sample.

Stimuli and apparatus

The experimental stimuli consisted of 26 different indoor and outdoor scenes. In each scene, a single individual was looking at one of two objects that were placed within reaching distance. Thus, there was a total of 52 different objects across all scenes (see online supplement S1 for a complete list of all objects). The direction of the gaze (left/right) and the placement of the objects (object A and B left/right) were balanced by taking four photographs of each scene (see Figure 1 for an outdoor example). Similar to Zwickel and Võ (2010), we did not restrict the position of the individual in the photograph (i.e., the
person could appear in the centre or more peripherally) such that participants could not expect a specific spatial structure of the scene and the gaze cue. This created 104 unique naturalistic pictures in total. For each participant, a set was randomly taken from this pool containing one version of each scene, resulting in 26 trials. The number of stimuli with leftward and rightward gaze of the depicted person, respectively, was balanced within each participant. Eye movements were tracked with the corneal reflection method and were recorded with an EyeLink 1000plus system (tower mount) at a sampling rate of 1000 Hz. The stimulation was controlled via Presentation® (Neurobehavioural Systems). All stimuli had a resolution of 1280 × 960 pixels and were displayed on a 24” LG 24MB65PY-B screen (resolution: 1920 × 1200 pixels, display size: 516.9 × 323.1 mm) with a refresh rate of 60 Hz. The viewing distance amounted to 50 cm thus resulting in a visual angle of 38.03° × 28.99° for the photographs.

**Design and procedure**

The experimental design was a 2 × 2 mixed design. First, as a two-level between-subject factor each participant was either assigned to the free viewing or the explicit encoding group (instruction group). Additionally, as a two-level within subject factor object role was manipulated, with objects being cued or uncued by the depicted individual in the scene.

After arriving at the laboratory individually, participants were asked to give full informed consent. Then the eye-tracker was calibrated for each participant using a 9-point grid. According to the manipulation, half of the participants were told that there was a follow-up memory test for objects that were part of the depicted scenes. All participants were then told to look at the following scenes freely without specifying further exploration goals or mentioning the content of the scenes. The presentation order of the pictures was randomized. Each trial started with the presentation of a fixation cross for one second, followed by the scene for 10 s. This interval was chosen based on our previous studies on social attention (End & Gamer, 2017; Flechsenhar & Gamer, 2017) and was slightly longer than the interval (7 s) that was used by Zwickel and Võ (2010). The inter-trial interval varied randomly between 1 and 3 s. After the last trial, participants filled in demographic
questionnaires and completed the AQ-k. These questionnaires were used for characterizing the current sample of participants, but they were also introduced to reduce recency effects in the memory task that was accomplished afterwards. It took approximately 5–10 min to complete the questionnaires. Participants then were asked to recall as many objects from the scenes as possible and write them down on a blank sheet of paper. No time limit was given but after 10 min, the experimenter asked participants to come to an end. In fact, most participants stopped earlier and indicated that they did not recall further objects. Finally, participants received course credit or payment and were debriefed.

**Data analysis**

For data processing and statistical analysis, the open-source statistical programming language R (R Core Team, 2019) was used with the packages tidyverse (Wickham, 2017), knitr (Xie, 2015) and papaja (Aust & Barth, 2018) for reproducible reporting. All analysis and data are available at https://osf.io/jk9s4/. For the analysis of the eye-tracking data, EyeLink's standard configuration was used to parse eye movements into saccades and fixations. Saccades were defined as eye movements exceeding a velocity threshold of 30 °/s or an acceleration threshold of 8.000 °/s². Fixations were defined as time periods between saccades.

We determined the following regions of interest (ROI) by colour coding respective images regions by hand using GIMP (GNU Image Manipulation Program): the cued object (average relative size on image: $M_{\text{size}} = 1.90\%$, $SD_{\text{size}} = 1.95\%$, average visual degree on image: $M_{\text{deg}} = 5.23\degree$, $SD_{\text{deg}} = 2.90\degree$), the uncued object ($M_{\text{size}} = 1.92\%$, $SD_{\text{size}} = 2.02\%$, $M_{\text{deg}} = 5.19\degree$, $SD_{\text{deg}} = 2.85\degree$), the head ($M_{\text{size}} = 2.01\%$, $SD_{\text{size}} = 1.50\%$, $M_{\text{deg}} = 5.74\degree$, $SD_{\text{deg}} = 1.91\degree$) and the body ($M_{\text{size}} = 7.76\%$, $SD_{\text{size}} = 4.11\%$, $M_{\text{deg}} = 12.55\degree$, $SD_{\text{deg}} = 4.23\degree$) of the depicted person.

Gaze variables of interest were calculated in a largely similar fashion as in Zwickel and Võ (2010). Specifically, we determined the cumulative duration and number of fixations on each ROI per trial. These values were divided by the total time or number of fixations, respectively, to yield proportions. As an additional measure of prioritization, particularly for early attentional allocation, we determined the latency of the first fixation that was directed towards each ROI. These measures allow for effective comparisons of prioritization between the two relevant objects and between the head and the body. To reveal direct relations between the head and the relevant objects, we calculated the proportion of saccades that left the head region of the depicted individual and landed on the cued and uncued objects, respectively. In order to analyze the influence of the experimental manipulations on the eye-tracking data, we carried out separate analyses of variance (ANOVAs) including the between-subject factor instruction group and the within-subject factor object role. ANOVAs were conducted on the dependent variables fixation latency and proportion of saccades from the head towards the object. To examine general effects of social attention, a separate ANOVA with the between-subject factor instruction group and the within-subject factor ROI (head vs. body region) was conducted on fixation latency.

Fixation durations and numbers of fixations were analyzed in more detail by additionally considering the temporal progression of effects. To this aim, we calculated relative fixation durations as well as relative numbers of fixations on each ROI for five time bins of 2 s each spanning the whole viewing duration. These data were analyzed using separate ANOVAs on relative fixation durations and numbers, respectively. The first analyses focused on the temporal progression of cueing effects and included the between-subject factor instruction group and the within-subject factors object role and time point. Subsequent analyses on general effects of social attention included the between-subject factor instruction group and the within-subject factors ROI (head vs. body region) and time point. In case of significant interaction effects, we calculated contrasts using emmeans (Lenth, 2019) as post-hoc tests with $p$ values adjusted according to Tukey’s honest significant difference method.

The memory test was scored manually by comparing the list of recalled objects to the objects that appeared in the scenes. We separately scored whether cued or uncued objects were recalled and ignored any other reported details. Afterwards, we calculated the sum of recalled objects separately for cued and uncued details. These data were analyzed using an ANOVA including the between-subject factor instruction group and the within-subject factor object role. To further assess the influence of visual...
attention on memory, we used a generalized linear mixed model (GLMM) approach implemented via lme4 (Bates et al., 2015). Based on the ANOVA results (see below), we used a sequential model building strategy starting with model 1 including only instruction group as the main predictor of subsequent recall performance. In the second step we added z-standardized relative fixation duration in model 2a and analogously, the z-standardized relative number of fixations in model 2b. In the third step, we added object role and corresponding interaction terms with the other factors to the previous models. We always tested the higher-order model against its lower-order counterpart using an ANOVA approach to examine if relative fixation duration and/or relative number of fixations had incremental value beyond group membership or interacted with object role in predicting recall performance.

For all analyses, the a priori significance level was set to $\alpha = .05$. ANOVAs were computed with the package afex (Singmann et al., 2019). As effect sizes, generalized eta-square ($\hat{\eta}^2_G$) values are reported, where guidelines suggest .26 as a large, .13 as a medium and .02 as a small effect (Bakeman, 2005). Greenhouse–Geisser correction was applied in all repeated-measures ANOVAs containing more than one degree of freedom in the numerator to account for potential violations of the sphericity assumption (Greenhouse & Geisser, 1959).

**Results**

**Gaze following**

A significant main effect of object role in the analysis of fixation latencies indicates earlier fixations on cued compared to uncued objects ($F(1, 91) = 54.84$, $p < .001$, $\hat{\eta}^2_G = .171$; $M_{lat,cued} = 1852ms$, $M_{lat,uncued} = 2413ms$). The main effect of instruction group was also significant, with earlier fixations on both objects in the explicit encoding ($F(1, 91) = 34.91$, $p < .001$, $\hat{\eta}^2_G = .202$; $M_{lat,mem} = 1821ms$) compared to the free viewing group ($M_{lat,free} = 2444ms$). The interaction effect failed to reach statistical significance ($F(1, 91) = 3.48$, $p = .065$, $\hat{\eta}^2_G = .013$; see Figure 2 A).

Comparable effects were obtained for saccades leaving the head which were more likely to land on the cued compared to the uncued object as confirmed by a significant main effect of object role, ($F(1, 91) = 43.17$, $p < .001$, $\hat{\eta}^2_G = .160$; $M_{sac,cued} = 10.35\%$, $M_{sac,uncued} = 4.83\%$). The main effect for group showed that saccades of participants in the explicit encoding group were more often directed towards any of the objects as compared to the free viewing group ($F(1, 91) = 24.71$, $p < .001$, $\hat{\eta}^2_G = .140$; $M_{sac,free} = 5.98\%$, $M_{sac,mem} = 9.20\%$). Again, the interaction effect of instruction group and object role failed to reach statistical significance ($F(1, 91) = 1.14$, $p = .288$, $\hat{\eta}^2_G = .005$; see Figure 2 B).

The time course of fixation durations and numbers on all four ROIs is depicted in Figure 3. In the corresponding ANOVA focusing on the temporal progression of gaze cueing effects, we obtained significant main effects of object role, indicating that participants fixated the cued object longer ($F(1, 91) = 5.60$, $p = .020$, $\hat{\eta}^2_G = .008$; $M_{dur,mem} = 14.40\%$, $M_{dur,uncued} = 13.10\%$), and more often ($F(1, 91) = 8.42$, $p = .005$, $\hat{\eta}^2_G = .009$; $M_{num,cued} = 13.19\%$, $M_{num,uncued} = 12.02\%$), than the uncued object. Explicit instructions also led to longer ($F(1, 91) = 19.31$, $p < .001$, $\hat{\eta}^2_G = .083$; $M_{dur,mem} = 15.91\%$, $M_{dur,free} = 11.59\%$), and more fixations ($F(1, 91) = 18.36$, $p < .001$, $\hat{\eta}^2_G = .082$; $M_{num,cued} = 14.39\%$, $M_{num,free} = 10.82\%$), on the objects as compared to the free viewing condition (see Figure 2 B, C).

The interaction effect of instruction group and object role was only statistically significant for the number of fixations ($F(1, 91) = 4.37$, $p = .039$, $\hat{\eta}^2_G = .005$), but failed statistical significance for the duration of fixations ($F(1, 91) = 2.84$, $p = .096$, $\hat{\eta}^2_G = .004$). However, contrasts of the estimated marginal means for both fixation measures revealed a statistically significant difference for object role only in the free viewing group (duration: $t(91) = 2.85$, $p = .005$, number: $t(91) = 3.51$, $p = .001$), with more and longer fixations on the cued object. In the explicit encoding group, contrasts of object role did not reach statistical significance (both $p > .5$).

The two way interactions between object role and time points were statistically significant for both measures (duration: $F(3.21, 291.82) = 23.93$, $\epsilon = 0.8$, $p < .001$, $\hat{\eta}^2_G = .051$, number: $F(3.11, 283.34) = 26.65$, $\epsilon = 0.78$, $p < .001$, $\hat{\eta}^2_G = .055$). Pairwise contrasts on estimated marginal means revealed that the interaction was mainly driven by more fixations on the cued object than on the uncued object during first and last time point (first: duration: $t(91) = 7.59$, $p < .001$, number: $t(91) = 7.90$,
$p < .001$, last time point: duration: $t(91) = 2.99$, $p = .004$, number: $t(91) = 3.43$, $p = .001$, with all other bins $p > .2$, see Figure 3).

Additionally, time dependent group differences were revealed by significant interactions between instruction group and time points for both measures (duration: $F(2.91, 265.26) = 4.56$, $\epsilon = 0.73$, $p = .004$, $\hat{\eta}^2_p = .012$, number: $F(2.9, 264.03) = 2.92$, $\epsilon = 0.73$, $p = .036$, $\hat{\eta}^2_p = .008$). Pairwise contrasts on estimated marginal means indicate significant differences with longer duration and more fixations on both objects for the explicit encoding group during all except the second time point (first: duration: $t(91) = 6.34$, $p < .001$, numbers: $t(91) = 6.04$, $p < .001$, third: duration: $t(91) = 3.20$, $p = .002$, numbers: $t(91) = 3.08$, $p = .003$, fourth: duration: $t(91) = 4.13$, $p < .001$, numbers: $t(91) = 3.74$, $p < .001$, and last time point: duration: $t(91) = 3.75$, $p < .001$, numbers: $t(91) = 3.87$, $p < .001$).

The three-way interaction of instruction group, object role and time point failed to reach statistical significance for fixations durations ($F(3.21, 291.82) = 1.21$, $\epsilon = 0.8$, $p = .308$, $\hat{\eta}^2_p = .003$) and numbers ($F(3.11, 283.34) = 1.37$, $\epsilon = 0.78$, $p = .253$, $\hat{\eta}^2_p = .003$).

**Memory for objects**

An analysis of the recall data showed, that participants in the explicit encoding group remembered more items than participants from the free viewing group ($F(1, 92) = 33.23$, $p < .001$, $\hat{\eta}^2_p = .234$; $M_{\text{recall,free}} = 11.23$, $M_{\text{recall,mem}} = 18.72$). Neither the main effect of

![Figure 2](image2.png)

**Figure 2.** Bar plots of the different prioritization measures for the attentional orienting towards the cued and uncued objects as a function of instruction group. Note that data were aggregated across time bins for fixation duration and numbers. Error bars represent standard errors of the mean.

![Figure 3](image3.png)

**Figure 3.** Time course of fixation durations and numbers as a function of the region of interest (ROI). Error bars represent standard errors of the mean.
object role \((F(1, 92) = 0.59, p = .444, \hat{\eta}_G^2 = .001)\) nor the interaction effect were statistically significant \((F(1, 92) = 0.02, p = .878, \hat{\eta}_G^2 = .000; \text{see Figure 4})\).

In order to examine the influence of visual exploration on recall performance, we used a GLMM approach starting with a first model where only group assignment was entered. Corresponding to the ANOVA results discussed above, this model revealed a significant effect for group (see Table 1 for model parameters and model selection criteria). Next, we built two extended models incorporating measures of visual attention: Model 2a included the main effect of (z-standardized) relative fixation duration and its interaction with group. To Model 2b we added (z-standardized) relative number of fixations and its interaction with group. Surprisingly, model 2a including the main effect of fixation duration and its interaction with group did not yield a better prediction of recall performance in comparison to model 1 \((p = .168)\). By contrast, model 2b including the main effect of number of fixations and its interaction with group improved the prediction of recalled stimuli \((\chi^2(2) = 9.67, p < .01)\) with a significant weight for the number of fixations. As a last step, we tested whether object role further improves the prediction of recall performance in comparison to model 2b, which was not the case \((p = .924)\).

**Table 1.** Parameters and model selection criteria of general linear mixed models predicting object recall from group, number/duration of fixation and object role.

| Coefficient       | Estimate | SE  | z     | p        | AIC   | BIC   | DIC   | df |
|-------------------|----------|-----|-------|----------|-------|-------|-------|----|
| **Model 1**       |          |     |       |          |       |       |       |    |
| Intercept         | -1.46    | 0.16| -9.16 | <.001    | 4930  | 4956  | 4922  | 4492|
| Group             | 0.79     | 0.14| 5.45  | <.001    |       |       |       |    |
| **Model 2a**      |          |     |       |          |       |       |       |    |
| Intercept         | -1.44    | 0.16| -9.02 | <.001    | 4931  | 4969  | 4919  | 4490|
| Group             | 0.77     | 0.14| 5.3   | <.001    |       |       |       |    |
| Duration          | 0.12     | 0.06| 1.9   | .057     |       |       |       |    |
| Group × duration  | -0.11    | 0.08| -1.39 | .165     |       |       |       |    |
| **Model 2b**      |          |     |       |          |       |       |       |    |
| Intercept         | -1.43    | 0.16| -8.94 | <.001    | 4925  | 4963  | 4913  | 4490|
| Group             | 0.75     | 0.14| 5.19  | <.001    |       |       |       |    |
| Number            | 0.19     | 0.06| 2.96  | <.01     |       |       |       |    |
| Group × number    | -0.12    | 0.08| -1.6  | .11      |       |       |       |    |
| **Model 3b**      |          |     |       |          |       |       |       |    |
| Intercept         | -1.41    | 0.17| -8.37 | <.001    | 4932  | 4996  | 4912  | 4486|
| Group             | 0.74     | 0.16| 4.63  | <.001    |       |       |       |    |
| Number            | 0.19     | 0.08| 2.21  | <.05     |       |       |       |    |
| Object role       | -0.04    | 0.11| -0.41 | .683     |       |       |       |    |
| Group × number    | -0.16    | 0.1  | -1.51 | .13      |       |       |       |    |
| Group × object role| 0.01    | 0.15| 0.07  | 0.943    |       |       |       |    |
| Number × object role | 0    | 0.12| -0.01 | .995     |       |       |       |    |
| Group × number × object role | 0.08 | 0.15| 0.51  | 0.613    |       |       |       |    |

AIC, Akaike information criterion; BIC, Bayesian information criterion; DIC, Deviance information criterion; df, Residual degrees of freedom.

**Social prioritization**

Fixation latencies differed remarkably between the head and the body (see Figure 5 A). Consequently, the ANOVA yielded a significant main effect of ROI, with earlier fixations of the head compared to the body \((F(1, 91) = 216.00, p < .001, \hat{\eta}_G^2 = .585; M_{lat,head} = 722ms, M_{lat,body} = 2609ms)\). There was neither a statistically significant main effect of instruction group \((F(1, 91) = 0.69, p = .407, \hat{\eta}_G^2 = .003)\) nor...
an interaction of both factors \(F(1, 91) = 1.43, p = .235, \hat{\eta}_G^2 = .009\).

In the detailed ANOVA including time bin (see Figure 3), fixation duration and numbers showed a very similar pattern with longer, \(F(1, 91) = 292.51, p < .001, \hat{\eta}_G^2 = .473; M_{dur,head} = 23.22\%, M_{dur,body} = 7.99\%\), as well as more fixations \(F(1, 91) = 238.78, p < .001, \hat{\eta}_G^2 = .413; M_{num,head} = 20.79\%, M_{num,body} = 9.30\%\), on the head than the body. Remarkably, the instruction group did not exhibit a statistically significant main effect, neither for the fixation duration \(F(1, 91) = 3.25, p = .075, \hat{\eta}_G^2 = .011\), nor for the number of fixations, \(F(1, 91) = 1.76, p = .189, \hat{\eta}_G^2 = .006\). Furthermore, the interaction effects of instruction group and ROI failed to reach statistical significance for fixation duration \(F(1, 91) = 3.52, p = .064, \hat{\eta}_G^2 = .011\) and fixation numbers, \(F(1, 91) = 2.58, p = .112, \hat{\eta}_G^2 = .008\); see Figure 5 B, C.

However, the ANOVA yielded a three-way interaction of instruction group, ROI and time point for fixation durations \(F(2.89, 263.12) = 8.26, \epsilon = 0.72, p < .001, \hat{\eta}_G^2 = .016\), as well as for numbers of fixations \(F(2.93, 266.9) = 6.11, \epsilon = 0.73, p = .001, \hat{\eta}_G^2 = .012\). To follow-up on this result, we performed separate ANOVAs for each time point including instruction group and ROI as factors. Interestingly, for both measures, we observed a statistically significant interaction between instruction group and ROI only for the first time point (duration: \(F(1, 91) = 12.36, p = .001, \hat{\eta}_G^2 = .074, \) number: \(F(1, 91) = 8.07, p = .006, \hat{\eta}_G^2 = .053\)). Pairwise contrasts of estimated marginal means for this interval revealed a significant difference between both groups for the head region for fixation duration \(t(91) = 4.57, p < .001\) as well as numbers of fixations \(t(91) = 4.35, p < .001\) with more and longer fixations in the free viewing group (see Figure 3). The fixation duration and fixation number for the body region did not differ between groups during the first time point (both \(p > .1\)). For all other time points follow-up ANOVAs did not yield significant interactions between instruction group and object role, neither for fixation duration nor fixation number (all interactions \(p > .1\), for details see the online supplement, Tables S8–S12).

**Discussion**

By using naturalistic scenes with rich detail, this study aimed at conceptually replicating previous findings of a general prioritization of social cues (i.e., heads and bodies, Birmingham et al., 2008a; End & Gamer, 2017; Flechsenhar & Gamer, 2017) as well as previously reported gaze cueing effects elicited by a person being directed towards a specific object in the scene (Zwickel & Võ, 2010). Both effects were replicated.

In detail, heads of persons in the scene were fixated earlier and explored more extensively as compared to body regions (and also more than the cued or uncued objects\(^1\)). Additionally, in line with Zwickel and Võ.

---

**Figure 5.** Bar plot of the different prioritization measures for attentional orienting towards and visual exploration of the depicted person’s head and body as a function of instruction group. Note that data were aggregated across time bins for fixation duration and numbers. Error bars represent standard errors of the mean.
J. D. GROBEKATHÖFER ET AL.

(2010), cued objects were preferred over uncued ones. They were fixated remarkably earlier, longer and more often. Thus, gaze following effects did not only occur with respect to a more thorough processing overall but were also evident in an early allocation of attentional resources after stimulus onset. Additional support for such early prioritization was revealed in the temporal analysis of attentional exploration. Fixation durations and numbers differed most between the cued and uncued object during the first 2 s of the 10 s viewing period (see Figure 3).

Moreover, the prioritization of the head and the preference for the cued object indirectly suggest a link between these two regions. To investigate this relationship in more detail, we examined saccades leaving the head towards the cued and uncued object, respectively. Saccades leaving the head were significantly more likely to end on the cued than on the uncued object, directly linking fixations of the head and the cued object. Thereby, current results fully replicate the findings of Zwickel and Võ (2010) with a more naturalistic set of stimuli. As often, by using more naturalistic material, experimental control is reduced. We tried to minimize unsystematic effects by producing the stimuli in the same way as Zwickel and Võ (2010), but using real as compared to 3D rendered scenes. In particular, each scene was photographed four times with gaze direction and object placement being fully counterbalanced. Since four individual photographs of each scene were taken in the current study, we could not fully control all stimulus aspects. However, the full replication of the effects previously obtained with a different set of virtual scenes indicates that these effects generalize to naturalistic conditions and are stable against small variations in scene layout and presentation.

Besides conceptually replicating previous findings, this study also aimed at extending the line of research by testing the robustness of gaze following against top-down modulations. This was achieved by instructing half of the participants to memorize as many details of the presented scenes as possible. Since the depicted human being was not relevant to this task, we expected a generally reduced attention towards head and body regions as well as a more systematic exploration pattern, potentially reducing gaze cueing effects in fixations on and saccades towards cued objects. Unsurprisingly, the memory task that was accomplished after the eye-tracking experiment showed that participants, who knew about the free recall task in advance performed better in recalling items. Interestingly, the hypothesized enhanced attentional preference for the uncued object in the explicit encoding group was only found for fixation numbers. Against our hypothesis, the effect did not reach statistical significance for fixation latencies and durations (while being descriptively in the hypothesized direction, see Figure 2 A–D).

The temporal analysis of attentional allocation furthermore indicates that effects of the instruction were most pronounced during early periods of picture viewing. In the first 2 s, fixations on the head differed clearly between instruction groups, with less social prioritization by participants in the explicit encoding group. In the same interval, however, both groups showed the largest difference in attentional exploration of cued as compared to uncued objects. Overall, the explicit encoding group fixated longer and more often on both objects than the free viewing group but cueing effects were largely unaffected by the explicit task with the only exception of fixation numbers being slightly less biased towards the cued object in the explicit encoding group. Although the time course of attentional exploration (see Figure 3) seems to indicate that the encoding instruction induced a more systematic exploration of the objects particularly at early time points, the three-way interaction failed to reach statistical significance in both ANOVAs.

These findings indicate that the prioritization of social information is largely unaffected by a manipulation of goal-driven attention, although early fixations on the head were slightly inhibited in the current study. The attentional guidance of gaze was effective especially in the early phase of stimulus presentation, even when participants investigated the scenes with an explicit (non-social) task goal. In general, this early attentional preference for cued locations provides support for the automaticity and reflexivity of social attentional processes and is in line with previous studies on gaze cueing within highly controlled setups (e.g., Hayward et al., 2017; Ristic & Kingstone, 2005), more naturalistic laboratory studies (e.g., Castelhano et al., 2007; Zwickel & Võ, 2010) and real-life social situations (e.g., Hayward et al., 2017; Richardson et al., 2007). Moreover, the current results are consistent with recent findings of an early attentional bias towards social information.
in general (End & Gamer, 2017; Rösler et al., 2017) that seems to be relatively resistant against specific task instructions (Flechsenhar & Gamer, 2017).

As expected, participants with specific recall instructions performed better in the subsequent memory task. However, the contribution of the automatic attentional processes to memory encoding remains unclear. In particular, although cued objects were prioritized in the attentional exploration, only the general number of fixations irrespective of object role predicted stimulus recall across both groups (see Table 1). Fixation duration did not add incremental value. This is partially in line with studies on eye movements (e.g., Hollingworth & Henderson, 2002) and (non-social) cueing (Belopolsky et al., 2008; Schmidt et al., 2002), which showed that increased attention results in better memory performance. Originally, we additionally expected the cued object to be better recalled than the uncued one (Belopolsky et al., 2008; Schmidt et al., 2002). However, another study showed that if certain scene details have a special meaning (e.g., by being central to the content of a picture story), attention does no longer predict memory for these details (Kim et al., 2013). With respect to the current study, these findings may indicate that both objects that were placed within reaching distance of the depicted person conveyed such meaning and were therefore remembered with equal probability. Since we only tested for early memory effects, it would be very interesting to delay the memory test by at least 24 h to examine whether memory consolidation differs between cued and uncued objects (Squire, 1993). Another explanation for the currently observed effects might be that exploration time was sufficient to process both objects equally well. It would thus be very interesting for future studies to manipulate viewing durations and examine the effect of such manipulations on memory performance.

Although the current study has several strengths including the systematic generation of novel stimulus material and the large sample size, it also has some limitations. First, although this study shows that humans follow other persons’ gaze implicitly in unconstrained situations, this was shown for situations without real interactions between humans. Research shows that fixation patterns differ remarkably when a real interaction between persons is possible (e.g., Hayward et al., 2017; Laidlaw et al., 2011; for an overview see: Risko et al., 2016). However, our findings add evidence to classic highly controlled laboratory approaches to social attention, yet at the same time better approximates ecological research (Risko et al., 2012). Second, one might criticize that we did not control for directional information from the depicted person’s body in contrast to the head. Earlier studies show that body orientation is relevant for cueing (Hietanen, 1999; Lawson & Calder, 2016) and the influence of body orientation on the cueing effects (e.g., through peripheral vision) cannot be dissociated by our study design. However, our results indicate a direct link between the head and the cued object, as do the results of Zwickel and Võ (2010). In fact, overall the first fixation of the body occurred about 800 ms after first fixation on the cued object. Third, we used a rather long viewing time of 10 s. This time allows for a very detailed exploration of the depicted scene and our analyses of the time courses of attentional measures showed that effects of top-down instructions seemed to be more pronounced during the first few seconds and quickly vanished afterwards. Future studies should, therefore, either use tasks that are cognitively more demanding or focus on a systematic variation of viewing durations to further examine the automaticity of social attention and gaze following.

Overall, the current results provide additional support for previous findings that attention is shifted reflexively to locations where other persons are looking at (e.g., Hayward et al., 2017; Ristic & Kingston, 2005). This evidence, which was previously extended to free viewing of more complex static scenes by Zwickel and Võ (2010), was shown to be valid in more naturalistic scenes and relatively robust against top-down modulation. Even when explicitly directing attention away from depicted individuals by making objects task-relevant, social and joint attention still occurred, and were even largely comparable to the unbiased free viewing condition. These results indicate that the mere presence of other human beings, as well as their gaze orientation, have a strong impact on attentional exploration.

Note

1. A direct comparison of all ROIs, e.g., head with cued object, can be found in the supplementary material, Tables S13–S15 and Figures S1–S3.
Acknowledgements

The authors thank Michael Strunz for his help with data collection.

Disclosure statement

No conflict of interest has been declared by the author(s).

Funding

This work was supported by the European Research Council [grant number ERC-2013-StG #336305].

ORCID

Jonas D. Großekathöfer https://orcid.org/0000-0002-3445-9501
Kristina Suchotzki https://orcid.org/0000-0001-6303-357X
Matthias Gamer https://orcid.org/0000-0002-9676-9038

References

Aust, F., & Barth, M. (2018). *papaja: Create APA manuscripts with R Markdown*. https://github.com/crsh/papaja

Bakeman, R. (2005). Recommended effect size statistics for repeated measures designs. *Behavior Research Methods*, 37(3), 379–384. doi:10.3758/bf03192707

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67(1), 1–48. https://doi.org/10.18637/jss.v067.i01

Belopolsky, A. V., Kramer, A. F., & Theeuwes, J. (2008). The role of awareness in processing of oculomotor capture: Evidence from event-related potentials. *Journal of Cognitive Neuroscience*, 20(12), 2285–2297. https://doi.org/10.1162/jocn.2008.20161

Birmingham, E., Bischof, W. F., & Kingstone, A. (2008a). Gaze selection in complex social scenes. *Visual Cognition*, 16(2–3), 341–355. https://doi.org/10.1080/13506280701434532

Birmingham, E., Bischof, W. F., & Kingstone, A. (2008b). Social attention and real-world scenes: The roles of action, competition and social content. *Quarterly Journal of Experimental Psychology*, 61(7), 986–998. https://doi.org/10.1080/17470210701410375

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. *Journal of Vision*, 7(9), 251–262. https://doi.org/10.1167/7.9.307

Cole, G. G., Smith, D. T., & Atkinson, M. A. (2015). Mental state attribution and the gaze cueing effect. *Attention, Perception, & Psychophysics*, 77(4), 1105–1115. https://doi.org/10.3758/s13414-014-0780-6

Driver, J., Davis, G., Ricciardelli, P., Kidd, P., Maxwell, E., & Baron-Cohen, S. (1999). Gaze perception triggers reflexive visuospatial orienting. *Visual Cognition*, 6(5), 509–540. https://doi.org/10.1080/135602899394920

End, A., & Gamer, M. (2017). Preferential processing of social features and their interplay with physical saliency in complex naturalistic scenes. *Frontiers in Psychology*, 8, https://doi.org/10.3389/fpsyg.2017.00418

Flechsenhar, A. F., & Gamer, M. (2017). Top-down influence on gaze patterns in the presence of social features. *PLoS One*, 12(8), e0183799. https://doi.org/10.1371/journal.pone.0183799

Fletcher-Watson, S., Findlay, J. M., Leekam, S. R., & Benson, V. (2008). Rapid detection of person information in a naturalistic scene. *Perception*, 37(4), 571–583. https://doi.org/10.1068/p5705

Foulsham, T., Cheng, J. T., Tracy, J. L., Henrich, J., & Kingstone, A. (2010). Gaze allocation in a dynamic situation: Effects of social status and speaking. *Cognition*, 117(3), 319–331. https://doi.org/10.1016/j.cognition.2010.09.003

Freeth, M., Foulsham, T., & Kingstone, A. (2013). What affects social attention? Social presence, eye contact and autistic traits. *PLoS One*, 8(1), e53286. https://doi.org/10.1371/journal.pone.0053286

Freitag, C. M., Retz-Junginger, P., Retz, W., Seitz, C., Palmason, H., Meyer, J., Rosler, M., & von Gontard, A. (2007). Evaluation der deutschen version des Autismus-Spektrum-Quotienten (AQ) – die Kurzversion AQ-k. *Zeitschrift Für Klinische Psychologie Und Psychotherapie*, 36(4), 280–289. https://doi.org/10.1026/1616-3443.36.4.280

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

Frischen, A., Bayliss, A. P., & Tipper, S. P. (2007). Gaze cueing of attention: Visual attention, social cognition, and individual differences. *Psychological Bulletin*, 133(4), 694–724. https://doi.org/10.1037/0033-2909.133.4.694

Greenhouse, S. W., & Geisser, S. (1959). On methods in the analysis of profile data. *Psychometrika*, 24(2), 95–112. https://doi.org/10.1007/bf02289823

Hayward, D. A., Voorhies, W., Morris, J. L., Capozzi, F., & Ristic, J. (2017). Staring reality in the face: A comparison of social attention across laboratory and real world measures suggests little common ground. *Canadian Journal of Experimental Psychology/Revue Canadienne de Psychologie Expérimentale*, 71(3), 212–225. https://doi.org/10.1037/cep0000117

Hietanen, J. K. (1999). Does your gaze direction and head orientation shift my visual attention? *NeuroReport*, 10(16), 3443–3447. https://doi.org/10.1097/00001756-199911080-00033

Hollingworth, A., & Henderson, J. M. (2002). Accurate visual memory for previously attended objects in natural scenes. *Journal of Experimental Psychology: Human Perception and Performance*, 28(1), 113. https://doi.org/10.1037/0096-1523.28.1.113-136

Hung, C. J., & Hunt, A. R. (2012). Rapid communication: Physical self-similarity enhances the gaze-cueing effect. *Quarterly Journal of Experimental Psychology*, 65(7), 1250–1259. https://doi.org/10.1080/17470218.2012.690769
Kim, J. S.-C., Vossel, G., & Gamer, M. (2013). Effects of emotional context on memory for details: The role of attention. PLoS One, 8(10), e77405. https://doi.org/10.1371/journal.pone.0077405

Laidlaw, K. E. W., Foulsham, T., Kuhn, G., & Kingstone, A. (2011). Potential social interactions are important to social attention. Proceedings of the National Academy of Sciences, 108(14), 5548–5553. https://doi.org/10.1073/pnas.1017022108

Langton, S. R. H., & Bruce, V. (2000). You must see the point: Sensory coding of observed gaze direction. Current Biology, 10(19), R567–R572. https://doi.org/10.1016/S0960-9822(00)00701-9

Perez-Osorio, J., Müller, H. J., Wiese, E., & Wykowska, A. (2015). Gaze following is modulated by expectations regarding others’ action goals. PLoS One, 10(9), e0143614. https://doi.org/10.1371/journal.pone.0143614

Posner, M. I. (1980). Orienting of attention. Quarterly Journal of Experimental Psychology, 32(1), 3–25. https://doi.org/10.1080/00335558008248231

R Core Team. (2019). R: A language and environment for statistical computing. R Foundation for Statistical Computing. https://www.R-project.org/

Richardson, D. C., Dale, R., & Kirkham, N. Z. (2007). The art of conversation is coordination. Psychological Science, 18(5), 407–413. https://doi.org/10.1111/j.1467-9280.2007.01914.x

Risko, E. F., Laidlaw, K. E. W., Freeth, M., Foulsham, T., & Kingstone, A. (2012). Social attention with real versus reel stimuli: Toward an empirical approach to concerns about ecological validity. Frontiers in Human Neuroscience, 6(May), 1–11. https://doi.org/10.3389/fnhum.2012.00143

Risko, E. F., Richardson, D. C., & Kingstone, A. (2016). Breaking the fourth wall of cognitive science: Real-world social attention and the dual function of gaze. Current Directions in Psychological Science, 25(1), 70–74. https://doi.org/10.1177/0963721415617806

Ristic, J., & Kingstone, A. (2005). Taking control of reflexive social attention. Cognition, 94(3), B55–B65. https://doi.org/10.1016/j.cognition.2004.04.005

Rössler, L., End, A., & Gamer, M. (2017). Orienting towards social features in naturalistic scenes is reflexive. PLoS One, 12(7), e0182037. https://doi.org/10.1371/journal.pone.0182037

Schmidt, S. (2009). Shall we really do it again? The powerful concept of replication is neglected in the social sciences. Review of General Psychology, 13(2), 90–100. https://doi.org/10.1037/a0015108

Schmidt, B. K., Vogel, E. K., Woodman, G. F., & Luck, S. J. (2002). Voluntary and automatic attentional control of visual working memory. Perception & Psychophysics, 64(5), 754–763. https://doi.org/10.3758/bf03194742

Singmann, H., Bolker, B., Westfall, J., & Aust, F. (2019). Afex: Analysis of factorial experiments. https://CRAN.R-project.org/package=afex

Squire, L. (1993). The structure and organization of memory. Annual Review of Psychology, 44(1), 453–495. https://doi.org/10.1146/annurev.psych.44.1.453

Teufel, C., Alexis, D. M., Todd, H., Lawrance-Owen, A. J., Clayton, N. S., & Davis, G. (2009). Social cognition modulates the sensory coding of observed gaze direction. Current Biology, 19(15), 1274–1277. https://doi.org/10.1016/j.cub.2009.05.069

Wickham, H. (2017). Tidyverse: Easily install and load the “tidyverse”. https://CRAN.R-project.org/package=tidyverse

Wiese, E., Wykowska, A., Zwickel, J., & Müller, H. J. (2012). I see what you mean: How attentional selection is shaped by ascribing intentions to others. PLoS One, 7(9), e45391. https://doi.org/10.1371/journal.pone.0045391

Wykowska, A., Wiese, E., Prosser, A., & Müller, H. J. (2014). Beliefs about the minds of others influence how we process sensory information. PLoS One, 9(4), e94339. https://doi.org/10.1371/journal.pone.0094339

Xie, Y. (2015). Dynamic documents with R and knitr (2nd ed.). Chapman, Hall/CRC. https://yihui.name/knitr/

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