Running head: GAZE CUEING IN NATURALISTIC SCENES

Gaze cueing in naturalistic scenes under top-down modulation - A conceptual replication

Jonas Großekathöfer1, Kristina Suchotzki1, & Matthias Gamer1

1 Julius-Maximilians-University, Würzburg

Author Note

Department of Psychology, Julius-Maximilians-University of Würzburg, Würzburg,

Germany.

Enter author note here.

Correspondence concerning this article should be addressed to Jonas Großekathöfer,

Marcusstraße 9-11, 97070 Würzburg. E-mail: jonas.grossekathoefer@uni-wuerzburg.de

# Abstract

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 were observed in free viewing conditions. To investigate how robust this prioritization is against top-down modulation, half of the participants receive 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 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.

*Keywords:* keywords

Word count: ?wordcountaddin

Gaze cueing in naturalistic scenes under top-down modulation - A conceptual replication

# Introduction

Humans in their social environment rely on the information conspecifics provide. This does not only hold for reading explicit signals, such as verbal communication in conversations, 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; for a review see: Frischen, Bayliss, & Tipper, 2007; Langton & Bruce, 2000). This paradigm has been inspired by classical studies on spatial attention by Posner (1980) and consists of a centrally presented face with varying gaze direction 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). However, even though gaze cues are crucial for joint attention, this standard gaze cueing paradigm can be criticized for lacking ecological validity, because these 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, Laidlaw, Freeth, Foulsham, & Kingstone, 2012), whereas in field situations, gaze signals occur within a rich context of competing visual information. For example, in a recent study Hayward, Voorhies, Morris, Capozzi, and Ristic (2017) compared attentional measures of gaze following from laboratory (classical gaze cueing) and real world (real social engagement) and did not find reliable links between those measures.

As a compromise between rich but also less controllable field conditions and standardized by impoverished laboratory gaze cuing studies, Zwickel and Võ (2010) used pictures of a full person (instead of isolated heads or faces) as a directional cue within a naturalistic scene. In this study, the authors used a free viewing instruction, meaning that participants had no explicit task to fulfill. Zwickel and Võ (2010) 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, McIntyre, Hancock, & Leder, 2017) or comprised specific instructions such as asking participants to understand a scene (Castelhano, Wieth, & Henderson, 2007). Consequently, 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, that 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 gaze following triggered by 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. However, this prioritization occurred only when the person looked at the object and was not evident for images that included the loudspeaker. 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 as a non-social directional cue. 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 merely 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.

The given eye-tracking study aims at conceptually replicating and extending Zwickel and Võ’s findings (2010).

First, we were interested in whether the previously reported effects hold when using a different set of stimuli. Due to their resolution and reduced richness of details, the originally used 3D rendered scenes did not allow for an assessment of the depicted person’s gaze direction. As a consequence, the observed cueing effects were rather due to directional information inferred from the body and head of the person that were congruently aligned towards one of the objects. To replicate these findings, we developed a new set of photographic stimuli that had sufficient resolution to also allow for perceiving gaze direction. 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 instructions 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 (Foulsham, Cheng, Tracy, Henrich, & Kingstone, 2010) or action-related expectations (Perez-Osorio, Müller, Wiese, & Wykowska, 2015). These studies have in common (together with Zwickel & Võ, 2010) that they manipulate viewing behavior of the participant by manipulating the stimuli.

In the present study, however, we tried to modulate viewing behavior via top-down instructions. Specifically, half of the participants received an instruction beforehand, that they should try to remember as many objects from the scenes as possible (explicit encoding group). For the other half of the participants (free viewing group), 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. 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 less prioritization of the cued object. With respect to the memory test that was accomplished after viewing the scenes, we expected better memory for cued as compared to uncued objects due to the increased attention on these details. Furthermore, we anticipated a generally enhanced recall performance in the explicit encoding group but a reduced prioritization of cued objects.

Finally, as previous studies showed a strong preference of fixating the head over body and background regions in static images (End & Gamer, 2017; Freeth, Foulsham, & Kingstone, 2013; Zwickel & Võ, 2010), 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 *dz* > 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 ⍺ 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-β = .67). As a compromise, we aimed at examining 90 participants (plus eventual drop-outs) 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 (65 female and 29 male) between 18 and 55 years (M = 24.73 years, SD = 5.04 years) participated voluntarily. All participant had normal or corrected vision and were recruited at the University of Würzburg’s online subject pool or by blackboard. Participants received course credit or a financial compensation of 5€. All participants gave informed consent. One participant was excluded due to missing data (resulting in n = 93 for the eye tracking analysis).

## 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. The direction of the gaze (left/right) and the placement of the objects (object A and B left/right) were balanced by taking 4 photographs of each scene. 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. Eye movements were tracked with the corneal-reflection method and were recorded with an EyeLink 1000plus tower system at a sampling rate of 1000 Hz. The stimulation was controlled via Presentation®(Neurobehavioral Systems). All stimuli had a resolution of 1280 x 960 pixels and were displayed on a 24” LG 24MB65PY-B screen (resolution: 1920 x 1200 pixels; display size: 516.9 x 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° x 28.99° for the photographs.

## Design and Procedure

The experimental design was a 2 x 2 mixed design. First, as a two-level factor the instruction group was manipulated between participants. Each participant was either assigned to the free viewing or the explicit encoding group. Additionally, as a two-level within subject factor object role was manipulated, with objects either 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, only half of the participants were told that there is a follow-up memory test for objects that are 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 of the pictures was randomized. Each trial started with the presentation of a fixation cross for one second, followed by the scene for 10 seconds. The inter-trial interval varied randomly between 1 and 3 seconds. After the last trial, participants filled in a demographic questionnaire and completed the Autism-Spectrum Quotient scale (AQ-k, German version, Freitag et al., 2007). These questionnaires were mainly introduced to reduce recency effects in the memory task that was accomplished afterwards. Participants 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 minutes, the experimenter asked participants to come to an end. In fact, most participants stopped earlier and indicated that they do not recall further objects. Finally, Participants received course credit or payment and were debriefed.

## Data analysis

For data processing and statistical analyses, the open-source statistical programming language R (R Core Team, 2016)[[1]](#footnote-1) was used with the packages tidyverse (Wickham, 2017c), knitr (Xie, 2015) and papaja (Aust & Barth, 2017) for reproducible reporting. 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°/sec or an acceleration threshold of 8000°/sec². Fixations were defined as time periods between saccades. We determined the following regions of interest (ROI) by color-coding respective images regions by hand using GIMP (GNU Image Manipulation Program): the cued object, the uncued object, the head and the body of the depicted person. Gaze variables of interest were calculated in a largely similar fashion as in Zwickel & Võ (2010). Specifically, we determined tcumulative onperThese values were divided or number of fixations, respectively, to yield proportionsAs ameasurewe determined the of the first fixation that was directed towards each ROI, we calculated the proportion ofthat left the region of the depicted individual and landed on the , respectively

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.

In order to analyze the influence of the experimental manipulations on eye-tracking and memory data, we carried out separate analyses of variance (ANOVAs) with the between-subject factor instruction group and the within-subject factor object role on the dependent variables fixation latency, fixation duration, number of fixations, proportion of saccades from the head towards the object and recalled items. To examine general effects of social attention, separate ANOVAs using the between-subject factor instruction group and the within-subject factor ROI (head vs. body region) were conducted on the dependent variables fixation latency, fixation duration and number of fixations. For all analysis the a priori significance level of α = 0.05 was used. As effect size, η2G is reported.

# Results

## Gaze following

A significant main effect of object role in the analysis of fixation latencies indicates earlier fixations on ,The main effect of iwasearlieron both objects effect failed to reach statistical significance

Largely similar effects were obtained in the analyses of fixation durations and numbers (see Figure 1). Significant main effects of object role indicate that participants fixated the cued object longer, *F*(1*,*91) = 6*.*86, *p* = *.*010, *η²*= *.*017, and more often than the uncued object, *F*(1*,*91) = 10*.*18, *p* = *.*002, *η²*= *.*019. Explicit instructions also led to longer, *F*(1*,*91) = 18*.*02, *p < .*001, *η²*= *.*133, and more fixations on the objects as compared to the free viewing condition, EFFEKT FEHLT. The interaction effects of instruction group and object role were not statistically significant, neither for the durations, *F*(1*,*91) = 2*.*47, *p* = *.*120, *η²*= *.*006, nor for the number of fixations, *F*(1*,*91) = 2*.*09, *p* = *.*152, *η²*= *.*004.

Saccades leaving the head were more likely to land on the cued as compared to the uncued object (see Figure 1) as confirmed by a significant main effect of object role, *F*(1,91) = 39.87, *p* < .001, η² = .151. The main effect for group, *F*(1,91) = 25.14, *p* < .001, η² = .141, shows 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. Again, the interaction effect of instruction group and objet role failes to reach statistical significance, *F*(1,91) = 1.41, *p* = .239, η² = .006.

## Memory for objects

An analysis of the recall data showed, that participants in the explicit encoding group remembered more items than participants from the other group (see Figure 1), *F*(1*,*92) = 33*.*23, *p < .*001, *η²*= *.*234. Neither the main effect of object role, *F*(1*,*92) = 0*.*43, *p* = *.*516, *η²*= *.*001, nor the interaction effect were statistically significant, *F*(1*,*92) = 0*.*02, *p* = *.*878, *η²*= *.*000.

## Social prioritization

A similar pattern to the objects can be seen when comparing the head with the body region.

Fixationslthe region Consequently, tyielded main of ROI statistically significantmain of group between both factors

Fixation durations and numbers showed a very similar pattern with longer, *F*(1*,*91) = 175*.*08, *p < .*001, *η²*= *.*467, as well as more fixations on the head than the body ROI, *F*(1*,*91) = 144*.*50, *p* < *.*001, *η²*= *.*415 (see Figure 2). Remarkably, the instruction group did not exhibit a statistically significant effect, neither for the fixation duration, *F*(1*,*91) = 2*.*70, *p* = *.*104, *η²*= *.*016, nor for the number of fixations, *F*(1*,*91) = 1*.*25, *p* = *.*266, *η²*= *.*008. Furthermore, the interaction effects of instruction group and ROI failed to reach statistical significance for fixation durations, *F*(1*,*91) = 3*.*25, *p* = *.*075, *η²*= *.*016, and numbers, *F*(1*,*91) = 2*.*03, *p* = *.*158, *η²*= *.*010. These findings indicate that the prioritization of social ROIs was unaffected by the explicit instruction to attend to objects depicted in the scene.

# 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, Bischof, & Kingstone, 2008; 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 could be replicated. In detail, heads of persons in the scene were fixated earlier and explored more extensively as compared to body regions but also cued and uncued objects. Additionally, in line with Zwickel and Võ (2010), cued objects were preferred over uncued objects. They were fixated remarkably earlier, longer and more often. Thus, gaze following effects did not only occur with respect to a more thorough processing during the whole time of scene presentation, but were also evident in an early allocation of attentional resources after stimulus onset. All these findings together indicate a strong and automatic bias of using gaze cues for attentional guidance.

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 cues and uncued object, respectively. As shown in Figure 1, saccades leaving the head were significantly more likely to end on the cued than the uncued object indicating a direct link between head and cued object fixations. So far, the current results fully replicated 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 now 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 for all stimulus aspects. However, the full replication of the effects that were previously obtained with a different set of virtual scenes indicate that these effects generalize to naturalistic conditions and are stable against small variations in scene layout and presentation.

Besides replicating previous findings, this study also aimed at extending this 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 beings are 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. More interestingly, however, social attention as well as cueing effects in viewing behavior were largely unaffected by the explicit instruction to remember as many objects from the scenes as possible. Specifically, whereas participants in the explicit recall group paid more attention towards depicted objects in general, they did not show reduced attention towards head and body regions as compared to the free viewing group. Moreover, gaze cueing effects on fixation latencies and densities as well as the direction of saccades leaving the head region did not differ significantly between both experimental group. These findings indicate that the prioritization of social information in general as well as of cued objects in particular are largely unaffected by top-down instructions. The attentional guidance of gaze was effective even when participants investigated the scenes with an explicit (non-social) task goal. This evidence provides support for the automaticity and reflexivity of social attentional processes and is in line with previous studies on gaze cueing within highly controlled environments (e.g., Ristic & Kingstone, 2005, Hayward et al., 2017), more naturalistic laboratory studies (e.g., Castelhano et al., 2007, Zwickel & Võ, 2010) and even real life social situations (e.g., Foulsham et al., 2010, Freeth et al., 2013, Hayward et al., 2017). Moreover, the current results are consistent with recent findings of an early attentional bias towards social information (End & Gamer, 2017; Roesler et al., 2017) that seems to be relatively resistant against specific top-down instructions (Flechsenhar & Gamer, 2017).

As expected participants with specific recall instructions performed better in the subsequent memory task. However, the contribution of automatic attentional processes to memory encoding remains unclear. In particular, although cued objects were prioritized in the attentional exploration, this did not increase their probability of being recalled. This is in contrast to studies on eye movements and memory performance showing that increased attention results in better memory performance (e.g., Hollingworth & Henderson, 2002). From that perspective, the cued object should be recalled better than the uncued object. 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, Vossel & Gamer, 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 24h to examine whether memory consolidation differs between cued and uncued objects (Squire, Knowlton & Musen, 1993).

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 that need to be mentioned. First, …

Overall, the current results provide additional support to previous findings that attention is shifted reflexively to locations where other persons are looking at (e.g., Ristic & Kingstone, 2005, Hayward et al., 2017). This evidence, which was previously extended to free viewing of more complex static scenes by Zwickel and Võ (2010), was shown to be robust against top-down modulation. Even when explicitly directing attention away from depicted individuals by making objects task-relevant, social and joint attentional shifts were still affected by the mere presence of a person, 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.

# References

Aust, F., & Barth, M. (2017). papaja: Create APA manuscripts with R Markdown. Retrieved from <https://github.com/crsh/papaja>

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

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 International workshop on attention in cognitive systems (Vol. 3, pp. 251–262). Springer; Association for Research in Vision; Ophthalmology (ARVO). doi:[10.1167/3.9.307](https://doi.org/10.1167/3.9.307)

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. doi:[10.1080/135062899394920](https://doi.org/10.1080/135062899394920)

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. doi:[10.1016/j.cognition.2010.09.003](https://doi.org/10.1016/j.cognition.2010.09.003)

Fox, J., & Weisberg, S. (2011). An R companion to applied regression (Second.). Thousand Oaks CA: Sage. Retrieved from <http://socserv.socsci.mcmaster.ca/jfox/Books/Companion>

Freeth, M., Foulsham, T., & Kingstone, A. (2013). What Affects Social Attention? Social Presence, Eye Contact and Autistic Traits. PLoS ONE, 8(1), e53286. doi:[10.1371/journal.pone.0053286](https://doi.org/10.1371/journal.pone.0053286)

Freitag, C. M., Retz-Junginger, P., Retz, W., Seitz, C., Palmason, H., Meyer, J., ... Gontard, A. von. (2007). Evaluation der deutschen Version des Autismus-Spektrum-Quotienten (AQ) - die Kurzversion AQ-k. Zeitschrift Für Klinische Psychologie Und Psychotherapie, 36(4), 280–289. doi:[10.1026/1616-3443.36.4.280](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. doi:[10.3758/BF03208827](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. doi:[10.1037/0033-2909.133.4.694](https://doi.org/10.1037/0033-2909.133.4.694)

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. doi:[10.1037/cep0000117](https://doi.org/10.1037/cep0000117)

Henry, L., & Wickham, H. (2017). Purrr: Functional programming tools. Retrieved from <https://CRAN.R-project.org/package=purrr>

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. doi:[10.1037//0096-1523.28.1.113-136](https://doi.org/10.1037//0096-1523.28.1.113-136)

Langton, S. R. H., & Bruce, V. (2000). You must see the point: Automatic processing of cues to the direction of social attention. Journal of Experimental Psychology: Human Perception and Performance, 26(2), 747–757. doi:[10.1037/0096-1523.26.2.747](https://doi.org/10.1037/0096-1523.26.2.747)

Langton, S. R. H., McIntyre, A. H., Hancock, P. J. B., & Leder, H. (2017). Saccades and smooth pursuit eye movements trigger equivalent gaze-cued orienting effects. The Quarterly Journal of Experimental Psychology, (just-accepted), 1–30. doi:[10.1080/17470218.2017.1362703](https://doi.org/10.1080/17470218.2017.1362703)

Müller, K. (2017). Bindrcpp: An ’rcpp’ interface to active bindings. Retrieved from <https://CRAN.R-project.org/package=bindrcpp>

Müller, K., & Wickham, H. (2017). Tibble: Simple data frames. Retrieved from <https://CRAN.R-project.org/package=tibble>

Perez-Osorio, J., Müller, H. J., Wiese, E., & Wykowska, A. (2015). Gaze following is modulated by expectations regarding others’ action goals. PLoS ONE, 10(11), e0143614. doi:[10.1371/journal.pone.0143614](https://doi.org/10.1371/journal.pone.0143614)

R Core Team. (2016). R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from <https://www.R-project.org/>

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. doi:[10.3389/fnhum.2012.00143](https://doi.org/10.3389/fnhum.2012.00143)

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

Wickham, H. (2009). Ggplot2: Elegant graphics for data analysis. Springer-Verlag New York. Retrieved from [http://ggplot2.org](http://ggplot2.org/)

Wickham, H. (2017a). Forcats: Tools for working with categorical variables (factors). Retrieved from <https://CRAN.R-project.org/package=forcats>

Wickham, H. (2017b). Stringr: Simple, consistent wrappers for common string operations. Retrieved from <https://CRAN.R-project.org/package=stringr>

Wickham, H. (2017c). Tidyverse: Easily install and load the ’tidyverse’. Retrieved from <https://CRAN.R-project.org/package=tidyverse>

Wickham, H., & Henry, L. (2017). Tidyr: Easily tidy data with ’spread()’ and ’gather()’ functions. Retrieved from <https://CRAN.R-project.org/package=tidyr>

Wickham, H., Francois, R., Henry, L., & Müller, K. (2017). Dplyr: A grammar of data manipulation. Retrieved from <https://CRAN.R-project.org/package=dplyr>

Wickham, H., Hester, J., & Francois, R. (2017). Readr: Read rectangular text data.

Retrieved from <https://CRAN.R-project.org/package=readr>

Xie, Y. (2015). Dynamic documents with R and knitr (2nd ed.). Boca Raton, Florida: Chapman; Hall/CRC. Retrieved from <https://yihui.name/knitr/>

Zhu, H. (2018). KableExtra: Construct complex table with ’kable’ and pipe syntax. Retrieved from <https://CRAN.R-project.org/package=kableExtra>

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

Table 1

*Mean Fixation Latency (in Milliseconds), relative fixation number (FF), and fixation duration as a function of Group (explicit encoding, free viewing) and object (cued, uncued)*

free viewing explicit encoding

cued uncued cued uncued

|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Measurements | mean | sd | se | mean | sd | se | mean | sd | se | mean | sd | se |
| Fixation duration (in %) | 12 | 6 | 1 | 10 | 4 | 0 | 16 | 6 | 1 | 15 | 6 | 1 |
| Fixation latency (in ms) | 2075 | 724 | 107 | 2789 | 707 | 104 | 1606 | 405 | 59 | 2040 | 627 | 91 |
| Fixation number (in %) | 11 | 4 | 1 | 10 | 3 | 0 | 14 | 5 | 1 | 14 | 5 | 1 |

**Note:**

Here is a footnote.

Table 2

*Mean Fixation Latency (in Milliseconds), relative fixation number, and fixation duration as a function of Group (explicit encoding, free viewing) and person (head and body)*

free viewing explicit encoding

head body head body

|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Measurements | mean | sd | se | mean | sd | se | mean | sd | se | mean | sd | se |
| Fixation duration (in %) | 24 | 8 | 1 | 8 | 3 | 0 | 21 | 9 | 1 | 8 | 4 | 0 |
| Fixation latency (in ms) | 601 | 408 | 60 | 2618 | 1087 | 160 | 843 | 617 | 90 | 2572 | 908 | 133 |
| Fixation number (in %) | 22 | 6 | 1 | 9 | 3 | 0 | 20 | 7 | 1 | 9 | 4 | 0 |

**Note:**

Here is a footnote.

Objects

Fixation duration Fixation number

Objects   
cued   
uncued

0

%

5

%

10

%

15

%

%

0

%

5

10

%

15

%

Fixation latency Saccades leaving head Memory performance

0

1000

2000

3000

0

%

5

%

%

10

%

15

0

5

10

free explicit free explicit free explicit

*Figure 1*. Bar plot of the different prioritization measures for the fixation of the cued and uncued objects. Error bars represent standard error.

Head and Body

Fixation number

free

explicit

Fixation latency

0

%

10

%

20

%

free

explicit

Fixation duration

0

%

10

%

20

%

free

explicit

0

1000

2000

Persons

Head

Body

*Figure 2*. Bar plot of the different prioritization measures for the person in the scene for the head and body. Error bars represent standard error.

1. Specifically, R (Version 3.3.1; R Core Team, 2016) and the R-packages *bindrcpp* (Version 0.2; K. Müller,

   2017), *car* (Version 2.1.6; Fox & Weisberg, 2011), *dplyr* (Version 0.7.4; Wickham, Francois, Henry, & Müller,

   2017), *forcats* (Version 0.2.0; Wickham, 2017a), *ggplot2* (Version 2.2.1; Wickham, 2009), *kableExtra* (Version 0.7.0; Zhu, 2018), *knitr* (Version 1.19; Xie, 2015), *papaja* (Version 0.1.0.9655; Aust & Barth, 2017), *purrr*

   (Version 0.2.4; Henry & Wickham, 2017), *readr* (Version 1.1.1; Wickham, Hester, & Francois, 2017), *stringr*

   (Version 1.2.0; Wickham, 2017b), *tibble* (Version 1.3.4; K. Müller & Wickham, 2017), *tidyr* (Version 0.7.2;

   Wickham & Henry, 2017), and *tidyverse* (Version 1.2.1; Wickham, 2017c) is used for all analyses. [↑](#footnote-ref-1)