Gaze cueing in naturalistic scenes under top-down modulation - Effects on gaze behavior and memory performance

|  |
| --- |
| Jonas D. Großekathöfer, Kristina Suchotzki, & Matthias Gamer |
| Department of Psychology, University of 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 D. Großekathöfer, Marcusstraße 9-11, 97070 Würzburg. E-mail: [jonas.grossekathoefer@uni-wuerzburg.de](mailto: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 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 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:* Gaze Cueing, Joint Attention, Social Attention, Eye-Tracking, Top-Down Control, Memory Performace

Word count: 5325

Gaze cueing in naturalistic scenes under top-down modulation - Effects on gaze behavior and memory performance

# 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, 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, Bayliss, & Tipper, 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 cognitve processes like group identity (Liuzza et al., 2011), theory-of-mind (Cole, Smith, & Atkinson, 2015; Teufel et al., 2009; Wiese, Wykowska, Zwickel, & Müller, 2012; Wykowska, Wiese, Prosser, & Müller, 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, Laidlaw, Freeth, Foulsham, & Kingstone, 2012). **Although gaze cuing was also found with more naturalistic stimuli (Perez-Osorio, Müller, Wiese, & Wykowska, 2015)**, in a recent study in which Hayward, Voorhies, Morris, Capozzi, and Ristic (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 behavior (e.g. Fletcher-Watson, Findlay, Leekam, & Benson, 2008; Perez-Osorio et al., 2015; Zwickel & Võ, 2010). To specifically investigate 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. 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 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. 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 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), nevertheless our motivation was also to improve certain aspects of the study and at the same time extending the line of research.** Due to their low 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 could be rather 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, Cheng, Tracy, Henrich, & Kingstone, 2010)**. Together with Zwickel and Võ (2010), these studies have in common that they manipulate viewing behavior of the participant by manipulating the stimuli. In contrast, in the present study, we tried to modulate viewing behavior via task instructions. 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). 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 by discouraging observers to utilize the information provided by the 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 the uncued object. 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 is 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, 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 = 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- = .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 participated voluntarily. All participant had normal or corrected 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 missing eyetracking data, resulting in a final sample of**  for the analysis with female and male participants between and years ( years, 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 to 23, , ). **From the final sample, one participant actually exceeded the proposed cut-off value of >= 17 for Autistic disorders (, ).**

## 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 (see Figure 1 for an example). The position of the individual in the photograph, was comparable to Zwickel and Võ (2010) stimuli not controlled for, as a consequence 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. Eye movements were tracked with the corneal reflection method and were recorded with an EyeLink1000plus 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.

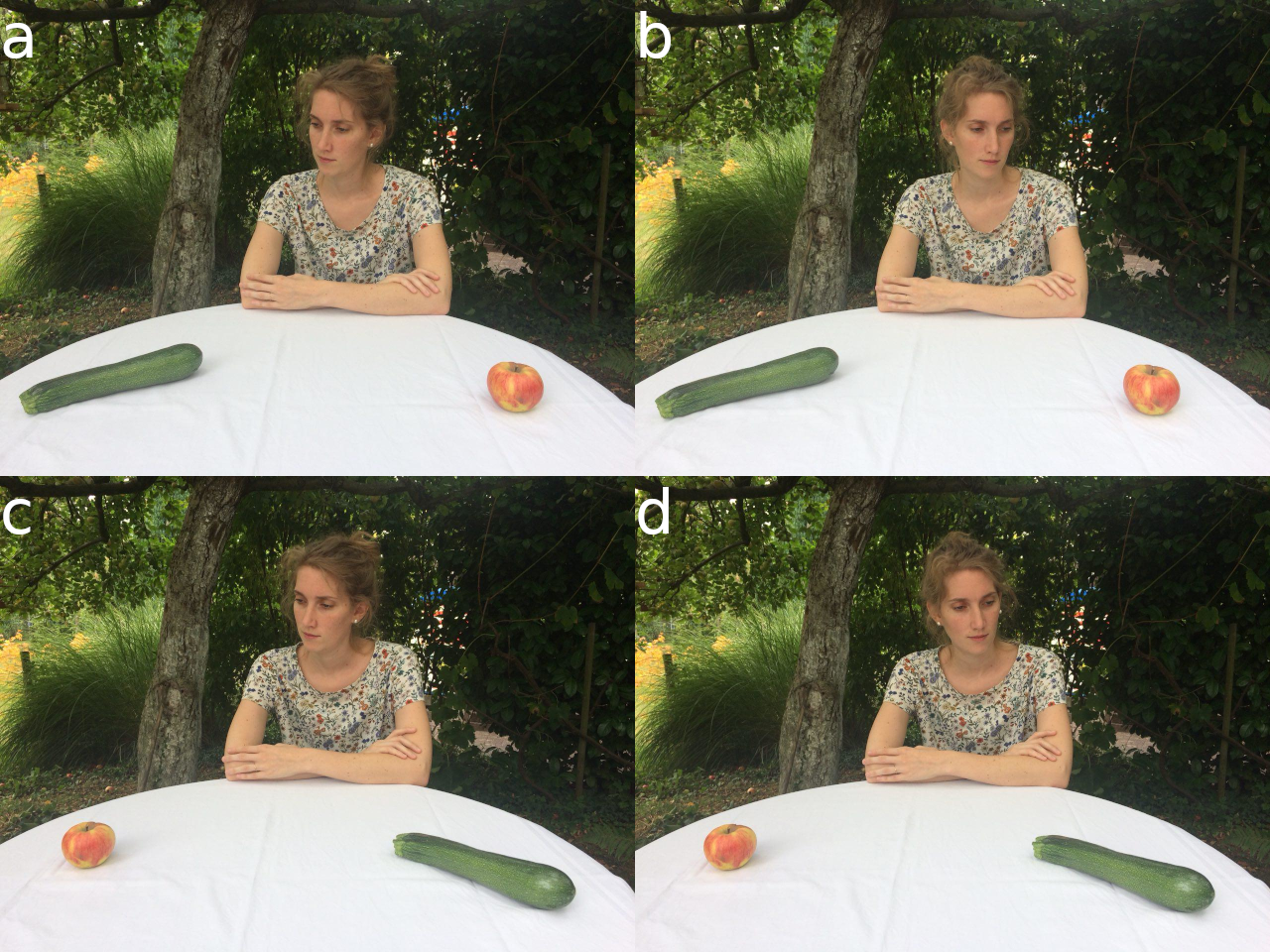


Figure 1: Example photographs of a single scene.Gaze direction and objects were balanced over participants. In total 104 photographs of 26 scenes were used. Since we did not obtain permission for publishing the original stimuli, this image shows an example that was not used in the experiment but taken post-hoc in order to illustrate the generation of the stimulus set.

## Design and Procedure

The experimental design was a 2 x 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 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. This interval was chosen based on our previous studies on social attention (End & Gamer, 2017; Flechsenhar & Gamer, 2017) and 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 seconds. 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. Participants then were asked to recall as many objects from the scene 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 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) and *knitr* (Xie, 2015) and *papaja* (Aust & Barth, 2018) 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 8.000°/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 (average relative size on image: , ), 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 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. 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 package *afex* (Singmann, Bolker, Westfall, & Aust, 2019). The ANOVAs had the between-subject factor instruction group and the within-subject factor object role and were conducted on the dependent variables fixation latency, relative fixation duration, relative numbers 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. **To assess the temporal attentional allocation we included a within-subject factor *time points* with 5 levels into the ANOVA for relative fixation duration as well as relative number of fixations. Time points were bins for the sum of durations and numbers of fixations for 2 seconds, resulting in 5 time points across 10 seconds of stimulus presentation. To follow up the main analysis on gaze following, we calculated contrasts using *emmeans* (Lenth, 2019) for cued vs. uncued object for instruction group and time points to assess the change of gaze following over time and for instruction groups. For social prioritization, the ANOVA was followed by contrasts for free viewing group vs explicit encoding group for head (and body) and time points. Here we were mainly interested in whether prioritization between instruction groups differed for the head over time.**  
**To complement the ANOVA on recall performance we used a generalized linear mixed model using *lme4* (Bates, Mächler, Bolker, & Walker, 2015) to examine the influence of relative fixation duration and relative numbers of fixation on stimulus recall performance. In model 1 instruction group was entered as the main predictor of subsequent recall performance. In the second step we added z-standardized relative fixation duration (i.e., range**  
**-1.12 to 5.98) in model 2a. And analougusly, the z-standardized relative number of fixations (i.e., range -1.23 to 6.14) in model 2b. In the third step we added object role to the previous models. We always tested the new model against the previous model in an ANOVA to test if relative fixation duration and/or relative fixation number had incremental value.**  
For all analysis the apriori significance level of = 0.05 was used. As effect sizes generalized eta-square () for ANOVAs are reported, where guidelines suggest .26 as a large, .13 as a medium and .02 as a small effect (Bakeman, 2005).

# 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. The main effect of instruction group was also significant, with earlier fixations on both objects in the explicit encoding **(, , ; )** compared to the free viewing group **()**. The interaction effect failed to reach statistical significance (, , ; see Figure  2 A).

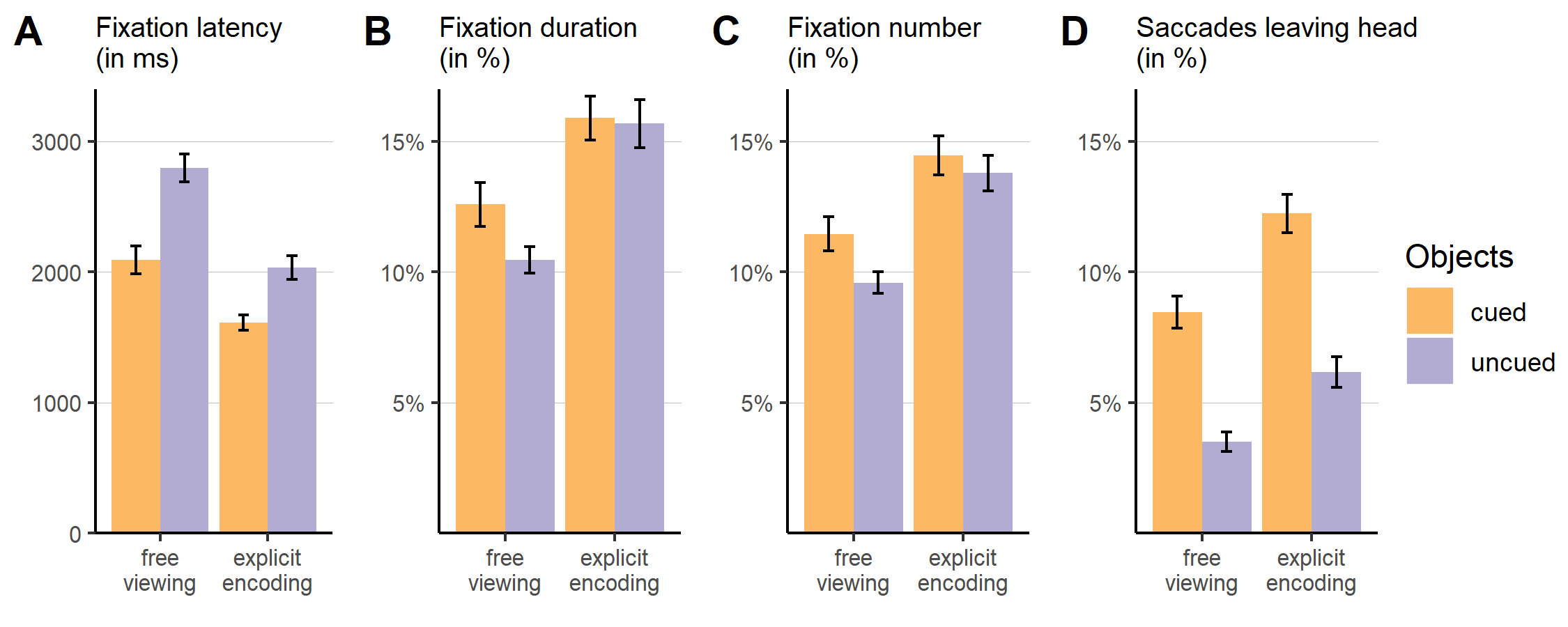


Figure 2: Bar plots of the different prioritization measures for the attentional orienting towards the cued and uncued objects as a function of instruction group. Error bars represent standard errors of the mean.

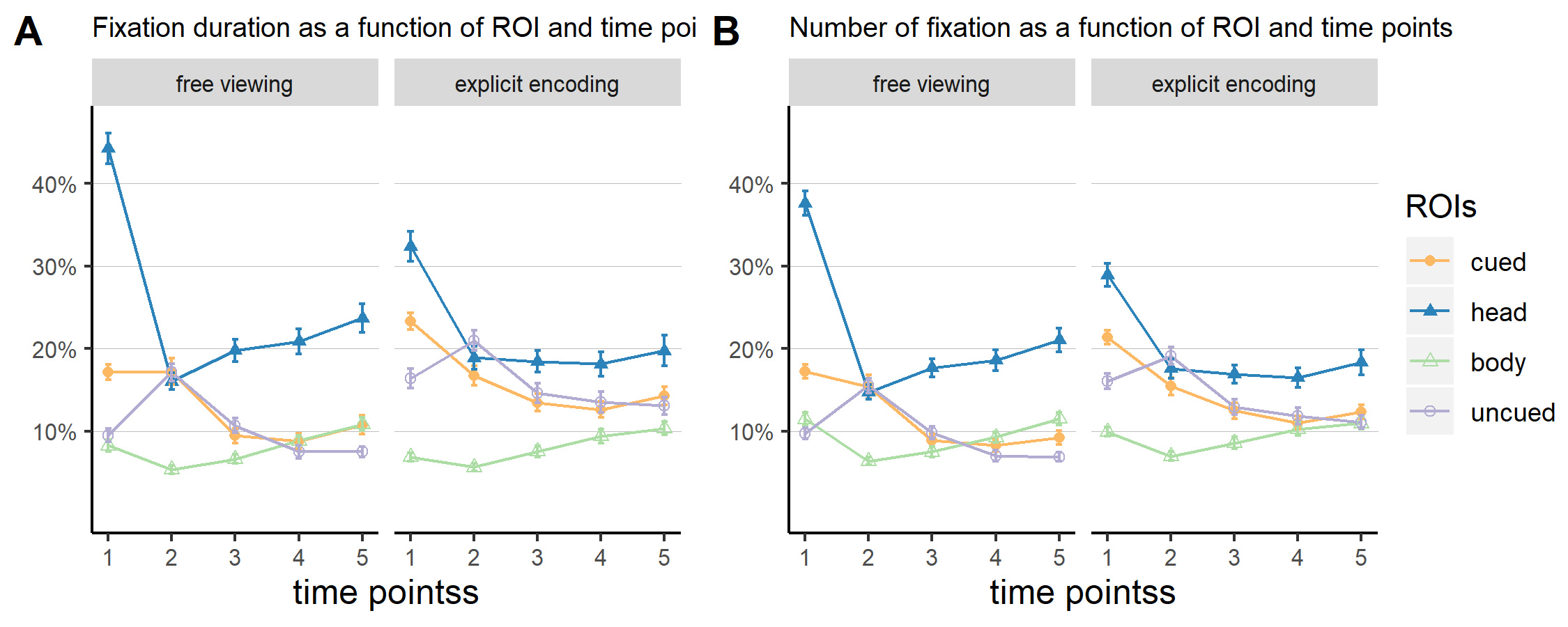


Figure 3: Temporal course of fixation duration and number of fixation. Error bars represent standard errors of the mean.

Largely similar effects were obtained in the analyses of fixation duration and numbers (see Figure 2 B & C). Significant main effects of object role indicated that participants fixated the cued object longer **(, , ; , )**, and more often **(, , ; , )**, than the uncued object. Explicit instructions also led to longer **(, , ; , )**, and more fixations **(, , ; , )**, on the objects as compared to the free viewing condition. The interaction effects of instruction group and object role were not statistically significant, neither for the duration (, , ), nor for the number of fixations (, , ).

**To assess the time course of attentional resources on gaze following we analyzed fixation duration and numbers across different bins on objects (see Figure 3). For both we see an interaction between object role and time points (duration: , , , number: , , ). As well as an interaction between group and time points (duration , , , number: , , ). Pairwise t-test reveal that the interaction between time points and object role was mainly driven by the first and last time point in the free viewing group (first time point: duration: , , number: , , last time point: duration: , , number: , , with all other bins ), whereas for the explicit encoding group the first and the second time point drive the interaction (first time point: duration: , , number: , , second time point: duration: , , number: , , with all other bins ).**

Saccades leaving the head were more likely to land on the cued compared to the uncued object as confirmed by a significant main effect of object role, **(, , ; , )**. 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 **(, , ; , )**. Again, the interaction effect of instruction group and object role failed to reach statistical significance (, , ; see Figure 2 D).

## 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 **(, , ; , )**. Neither the main effect of object role (, , ) nor the interaction effect were statistically significant (, , ; see Figure 4).

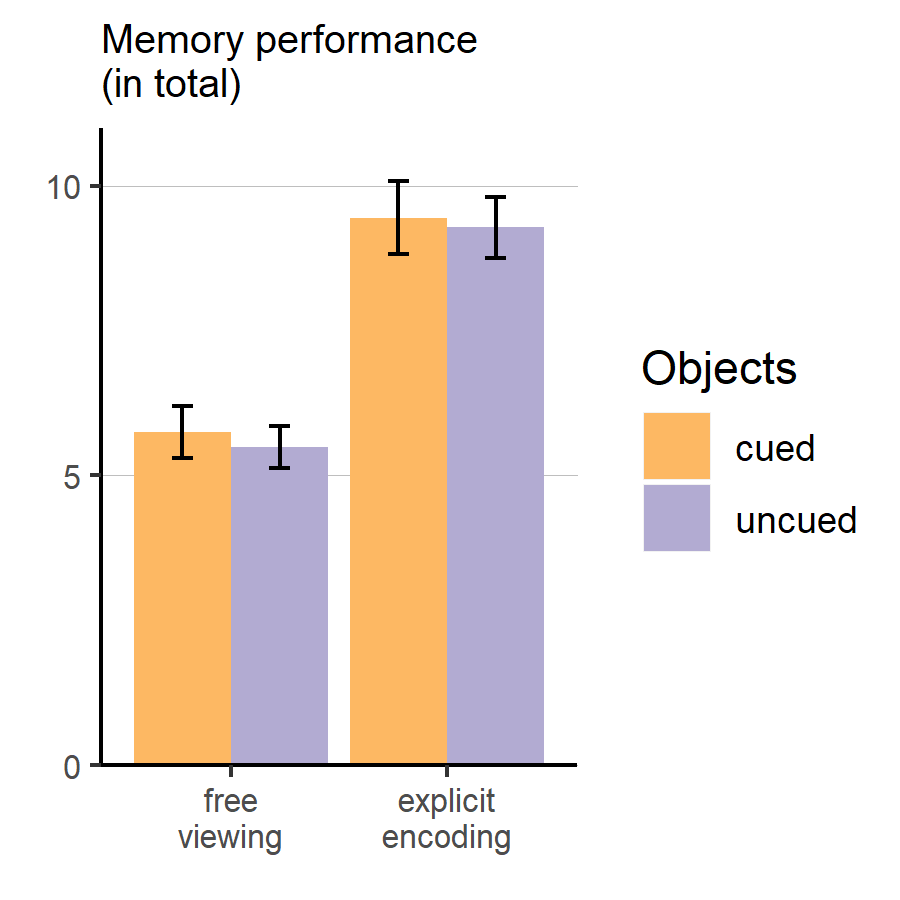


Figure 4: Bar plot of the memory performance for the cued and uncued objects as a function fo instruction group. Error bars represent standard errors of the mean.

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 |  |  |  |  | 4930 | 4956 | 4922 | 4492 |
| INTERCEPT | -1.46 | 0.16 | -9.16 | < .001 |  |  |  |  |
| GROUP | 0.79 | 0.14 | 5.45 | < .001 |  |  |  |  |
| Model 2a |  |  |  |  | 4931 | 4969 | 4919 | 4490 |
| INTERCEPT | -1.44 | 0.16 | -9.02 | < .001 |  |  |  |  |
| 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 |  |  |  |  | 4925 | 4963 | 4913 | 4490 |
| INTERCEPT | -1.43 | 0.16 | -8.94 | < .001 |  |  |  |  |
| 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 |  |  |  |  | 4932 | 4996 | 4912 | 4486 |
| INTERCEPT | -1.41 | 0.17 | -8.37 | < .001 |  |  |  |  |
| 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 | .943 |  |  |  |  |
| NUMBER\*OBJECT ROLE | 0 | 0.12 | -0.01 | .995 |  |  |  |  |
| GROUP\*NUMBER\*OBJECT ROLE | 0.08 | 0.15 | 0.51 | .613 |  |  |  |  |

*Note.* Akaike information criterion; BIC = Bayesian information criterion; DIC = Deviance information criterion; df = Residual degrees of freedom.

**For the follow up analysis of recall performance we started with the first model where only group assignment was entered. Revealing the significant effect for group as shown by the ANOVA previously (see Table 1 for model parameters and model selcetion criteria). Next, we added to separate models our variables of main interest: Model 2a got extended by the main effect of (z-standardized) relative fixation duration and the interaction with group. To Model 2b we added (z-standardized) relative number of fixations and the interaction with group. Surprisingly, the main effect of fixation duration and its interaction with group in model 2a failed to reach statistical significance. As expected, the main effect of number of fixations improved the prediction of recalled stimuli. Again the interaction of group and number of fixations did reach significance. For number of fixations we tested as a last step, whether object role further improves the prediction of recall performance, that was not the case, with the main effect an all other interactions remaining above our significance level.**

## 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 **(, , ; , )**. There was neither a statistically significant main effect of instruction group (, , ) nor an interaction of both factors, (, , ).

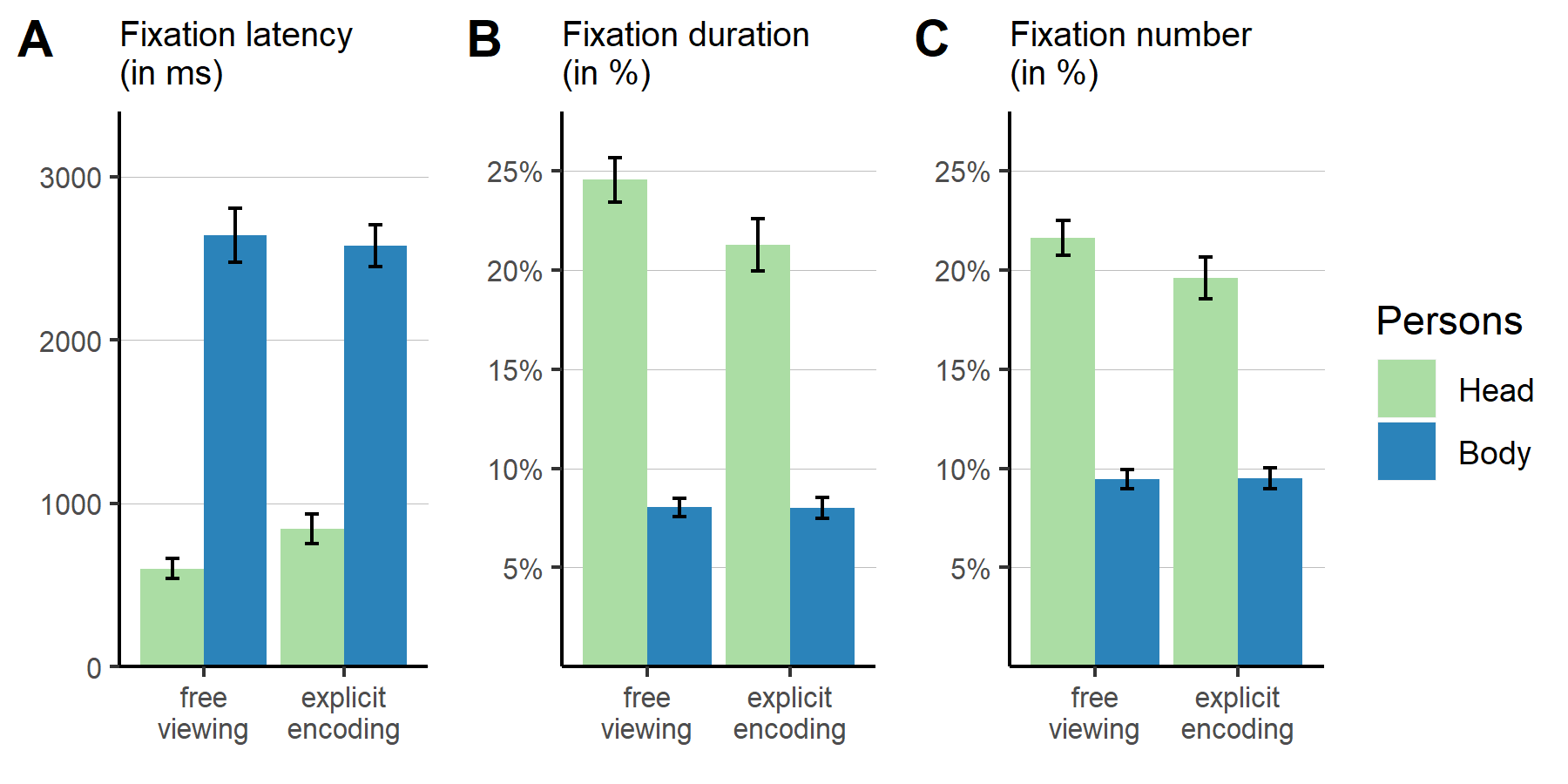


Figure 5: Bar plot of the different prioritization measures for attentional orienting towards and visual exploration of the depcited person’s head and body as a function of instruction group. Error bars represent standard errors of the mean.

Fixation duration and numbers showed a very similar pattern with longer, **(, , ; , )**, as well as more fixations **(, , ; , )**, on the head than the body ROI. Remarkably, the instruction group did not exhibit a statistically significant main effect, neither for the fixation duration (, , ), nor for the number of fixation, (, , ). Furthermore, the interaction effects of instruction group and ROI failed to reach statistical significance for fixation duration (, , ) and fixation numbers, (, , ; see Figure 5 B & C). These findings indicate that the prioritization of social ROIs was unaffected by the explicit instruction to attend to objects depicted in the scene.

**To assess instruction effects over time for social prioritization we analysed fixation duration and number across time points (see Figure 3). For both we see an significant interaction between group and time points (duration: , , , number: , , ) as well as between head/body and time points (duration: , , , number: , , ) . Pairwise t-test reveal that the interaction between group and time points was driven by the first and last time point (first time point: , , last time point: , , with all other time points ). For fixation number on the head, surprisingly the first and second time point show significant differences across group assignment (first: , , second: , , except the last time point ( , ), all other time points ). None of the contrasts for body reached significance across instruction group and time points (all ).**

# 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 were 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. **Additional support comes from the investigation of temporal allocation of attention. Differences between our manipulations were most present in the first 2 seconds (see Figure 3). During that time, fixations on the head differed clearly between instruction groups, with more social prioritization during free viewing. Temporal analysis for gaze following showed in the first 2 seconds strongest preference of the cued object in both groups. Interestingly, viewing pattern differd between both groups after the intial 2 seconds. Object prioritization was reversed durint the second time point (seconds 2 - 4) in the explicit encoding group with more and longer fixations on the uncued object. In contrast, participants in the free viewing group payed as much attention on the uncued object as on the cued object during that time period. It seems that the instruction induced a systematic exploration resulting in reducing attention for the cued object.** All these findings together indicate a strong and automatic bias of using gaze cues for attentional guidance **especially in the early phases of stimulus exploration. From the temporal analysis of attentional allocation we can conclude, that the unexpected missing interaction between group and object role comes very likely from the extended stimulus presentation. Our results indicate a rather strong group effects in the first 4 seconds, especially for object role. However, comparing both groups the picture is less clear for the preference of the head. The instruction manipulation seems to affect only very early the attentional allocation which does not persist over time.**

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 an 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 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 were 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, participants in the explicit recall group did not show reduced attention towards head and body regions as compared to the free viewing group. Moreover, whereas they paid more attention towards depicted objects in general, gaze cueing effects on fixation latencies and densities as well as the direction for saccades leaving the head region did not differ significantly between both experimental groups.

These findings indicate that the prioritization of social information in general as well as of the cued objects in particular are largely unaffected by a manipulation of goal-driven attention. The attentional guidance of gaze was effective, even when participants investigated the scenes with an explicit (non-social) task goal. This 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., Ristic & Kingstone, 2005; Hayward et al., 2017), 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, Dale, & Kirkham, 2007). Moreover, the current results are consistent with recent findings of an early attention bias towards social information (End & Gamer, 2017; Rösler, End, & Gamer, 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 number of fixations did improve the prediction of stimulus recall over group assingment (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, Kramer, & Theeuwes, 2008; Schmidt, Vogel, Woodman, & Luck, 2002) which showed that increased attention results in better memory performance. **We found this relationship only for number of fixation**. **Originally we expected the cued object to be recalled better than the uncued object (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, 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, 1993). **Another explanation might be that exploration time was sufficient to process both objects equally well. This interpretation finds support in the fact that the group assignment primary affected the first seconds of stimulus presentation.**

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, 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, Foulsham, Kuhn, & Kingstone, 2011; for an overview see: Risko, Richardson, & Kingstone, 2016). However, our findings add evidence to the classic highly controlled laboratory approaches to social attention, yet at the same time approximates more ecological research (Risko et al., 2012). An additional critique might be 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 does Zwickel and Võ (2010). In fact, overall the first fixation of the body occurs about 1 second after first fixation on the cued object.

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., 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 valid in more naturalistic scenes and robust against top-down modulation. Even when explicitly directing attention away from depicted individuals by making objects task-relevant, social and joint attention 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.

# Acknowledgements

The authors thank Michael Strunz for his help with data collection. This work was supported by the European Research Council (ERC-2013-StG #336305).

# References

Aust, F., & Barth, M. (2018). *papaja: Create APA manuscripts with R Markdown*. Retrieved from <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](https://doi.org/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. doi:[10.18637/jss.v067.i01](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. doi:[10.1162/jocn.2008.20161](https://doi.org/10.1162/jocn.2008.20161)

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. *Journal of Vision*, *3*(9), 251–262. doi:[10.1167/3.9.307](https://doi.org/10.1167/3.9.307)

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. doi:[10.3758/s13414-014-0780-6](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. doi:[10.1080/135062899394920](https://doi.org/10.1080/135062899394920)

End, A., & Gamer, M. (2017). Preferential processing of social features and their interplay with physical saliency in complex naturalistic scenes. *Frontiers in Psychology*, *8*. doi:[10.3389/fpsyg.2017.00418](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. doi:[10.1371/journal.pone.0183799](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. doi:[10.1068/p5705](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. doi:[10.1016/j.cognition.2010.09.003](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. 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)

Hietanen, J. K. (1999). Does your gaze direction and head orientation shift my visual attention? *NeuroReport*, *10*(16), 3443–3447. doi:[10.1097/00001756-199911080-00033](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. doi:[10.1037//0096-1523.28.1.113-136](https://doi.org/10.1037//0096-1523.28.1.113-136)

Hungr, 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. doi:[10.1080/17470218.2012.690769](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. doi:[10.1371/journal.pone.0077405](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. doi:[10.1073/pnas.1017022108](https://doi.org/10.1073/pnas.1017022108)

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)

Lawson, R. P., & Calder, A. J. (2016). The “where” of social attention: Head and body direction aftereffects arise from representations specific to cue type and not direction alone. *Cognitive Neuroscience*, *7*(1-4), 103–113. doi:[10.1080/17588928.2015.1049993](https://doi.org/10.1080/17588928.2015.1049993)

Lenth, R. (2019). *Emmeans: Estimated marginal means, aka least-squares means*. Retrieved from <https://CRAN.R-project.org/package=emmeans>

Liuzza, M. T., Cazzato, V., Vecchione, M., Crostella, F., Caprara, G. V., & Aglioti, S. M. (2011). Follow my eyes: The gaze of politicians reflexively captures the gaze of ingroup voters. *PLoS ONE*, *6*(9), e25117. doi:[10.1371/journal.pone.0025117](https://doi.org/10.1371/journal.pone.0025117)

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)

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

R Core Team. (2019). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from <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. doi:[10.1111/j.1467-9280.2007.01914.x](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. doi:[10.3389/fnhum.2012.00143](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. doi:[10.1177/0963721415617806](https://doi.org/10.1177/0963721415617806)

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)

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

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. doi:[10.3758/bf03194742](https://doi.org/10.3758/bf03194742)

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. doi:[10.1037/a0015108](https://doi.org/10.1037/a0015108)

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

Squire, L. (1993). The structure and organization of memory. *Annual Review of Psychology*, *44*(1), 453–495. doi:[10.1146/annurev.psych.44.1.453](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. doi:[10.1016/j.cub.2009.05.069](https://doi.org/10.1016/j.cub.2009.05.069)

Wickham, H. (2017). *Tidyverse: Easily install and load the ’tidyverse’*. Retrieved from <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. doi:[10.1371/journal.pone.0045391](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. doi:[10.1371/journal.pone.0094339](https://doi.org/10.1371/journal.pone.0094339)

Xie, Y. (2015). *Dynamic documents with R and knitr* (2nd ed.). Boca Raton, Florida: Chapman; Hall/CRC. Retrieved from <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. doi:[10.3758/PBR.17.2.257](https://doi.org/10.3758/PBR.17.2.257)