Idiosyncratic viewing patterns of social scenes reflect individual preferences

Adam M. Berlijn

Department of Experimental Psychology,
Heinrich-Heine-University Düsseldorf,
Düsseldorf, Germany
Institute of Clinical Neuroscience and Medical Psychology,
Medical Faculty, University Hospital Düsseldorf,
Heinrich-Heine University Düsseldorf, Düsseldorf, Germany
Institute of Neuroscience and Medicine (INM-1),
Research Centre Jülich, Jülich, Germany
Department of Psychology, Julius-Maximilians-University
Würzburg, Würzburg, Germany

✉

Lea K. Hildebrandt

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

✉

Matthias Gamer

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

✉

In general, humans preferentially look at conspecifics in naturalistic images. However, such group-based effects might conceal systematic individual differences concerning the preference for social information. Here, we investigated to what degree fixations on social features occur consistently within observers and whether this preference generalizes to other measures of social prioritization in the laboratory as well as the real world. Participants carried out a free viewing task, a relevance taps task that required them to actively select image regions that are crucial for understanding a given scene, and they were asked to freely take photographs outside the laboratory that were later classified regarding their social content. We observed stable individual differences in the fixation and active selection of human heads and faces that were correlated across tasks and partly predicted the social content of self-taken photographs. Such relationship was not observed for human bodies indicating that different social elements need to be dissociated. These findings suggest that idiosyncrasies in the visual exploration and interpretation of social features exist and predict real-world behavior. Future studies should further characterize these preferences and elucidate how they shape perception and interpretation of social contexts in healthy participants and patients with mental disorders that affect social functioning.

Introduction

Social interactions are ubiquitous in everyday life. Even basic interactions, such as avoiding bumping into one another on a crowded street, depend on our ability to recognize others and their intentions. Other people, especially their faces, thus provide a valuable source of information (Birmingham, Bischof, & Kingstone, 2008). Consequently, humans preferentially attend to such social features in their environment – a phenomenon also referred to as *social attention* (Salley & Colombo, 2016).

One factor that drives fixation location – and could potentially explain social attention – is physical saliency (i.e. low-level stimulus features, such as contrasts in color, intensity, or spatial orientation; Borji & Itti, 2013; Onat, Açik, Schuman, & König, 2014). Recent research, however, demonstrated that social features, such as heads, faces, or other body parts, strongly attract attention beyond the mere influence of physical saliency. For example, the relationship between physical saliency and fixation location is significantly reduced when social aspects are featured in the scene (End & Gamer, 2017). The prioritization of social features, especially for the head region, occurs very early during scene viewing (see also Rösler, End, & Gamer, 2017) but can also be influenced by top-down task instructions

(Castelhano, Mack, & Henderson, 2009; DeAngelus & Pelz, 2009; Smith & Mital, 2013). For example, telling participants to focus on social features while freely viewing naturalistic images decreased the latency of fixations on social regions of interest even for the very first fixation on the scene (End & Gamer, 2019). External factors, such as social features and task instructions, thus seem to bias early attentional processing already at the temporal resolution of stimulus-driven attention. Consequently, these findings raise the question whether social attention is purely driven by external factors (e.g. stimulus features and/or task instructions), or whether it also reflects meaningful interindividual differences, such that some individuals prefer social aspects of their environment consistently more than others.

Initial evidence that individuals differ in their gaze preferences for social features primarily came from group-based studies in (sub-)clinical populations. For example, individuals on the autism spectrum show atypical viewing patterns in social scenarios (Fletcher-Watson, Findlay, Leekam, & Benson, 2009) or generally reduced social attention (Speer, Cook, McMahon, & Clark, 2007). Similarly, social anxiety seems to alter gaze behavior toward social features (Schneier, Rodebaugh, Blanco, Lewin, & Liebowitz, 2011; Wieser, Pauli, Alpers, Mühlberger, 2009). Importantly, contrasting findings were also reported for certain experimental scenarios regarding the influence of autism spectrum (Horn, Mergenthaler, & Gamer, 2022) or social anxiety traits (Rösler, Göhring, Strunz, & Gamer, 2021). However, in general, such group-wide effects do not allow to infer that social attention is a meaningful *individual* predisposition. If social attention reflects a trait-like preference, this should be visible in idiosyncratic gaze patterns that are stable across time. In a study by Mehoudar, Arizpe, Baker, and Yovel (2014), participants had to memorize and subsequently recognize faces while eye-tracking data were recorded. The participants repeated this task after 3 days and 18 months. Indeed, the authors found the greatest similarity in gaze patterns across the three time points within participants indicating that the scanning strategy was neither stimulus- nor task-driven but rather represented an idiosyncratic pattern. Kennedy, D'Onofrio, Quinn, Bölte, Lichtenstein, and Falck-Ytter (2017) further demonstrated that fixation patterns of monozygotic twins were more similar than those of dizygotic twins or random pairs of participants, thus arguing for a genetic basis of such viewing preferences. Finally, de Haas, Iakovidis, Schwarzkopf, and Gegenfurtner (2019) showed stable individual viewing preferences for a number of different semantic categories. Interestingly, these preferences were most robust for social aspects, such as faces (Guy, Azulay, Kardosh, Weiss, Hassin, Israel, & Pertzov, 2019; Linka & de Haas, 2020).

It is less clear, however, what these individual differences in gaze patterns mean with respect to scene understanding and behavior, especially in the "real world." First, evidence was reported that a stronger preference to visually explore faces is related to face recognition skills using pictures as stimuli (de Haas et al., 2019). But the reliance of such experiments on the use of static, isolated images to study social attention has also been criticized (e.g. Birmingham, Ristic, & Kingstone, 2012; Risko, Laidlaw, Freeth, Foulsham, & Kingstone, 2012; Risko, Richardson, & Kingstone, 2016) because "real" social attention rarely happens in such impoverished, non-interactive environments. Using pictures of complex social scenes instead of isolated faces as stimuli can be seen as a first improvement (e.g. Birmingham et al., 2008; Cerf, Harel, Einhaeuser, & Koch, 2007; End & Gamer, 2017; End & Gamer, 2019; Flechsenhar, Larson, End, & Gamer, 2018a; Flechsenhar, Rösler, & Gamer., 2018b; Fletcher-Watson, Leekam, Benson, Frank, & Findlay, 2009; Rösler et al., 2017; Xu, Jiang, Wang, Kankanhalli, & Zhao, 2014). However, a natural viewing pattern could already be biased by a lack of responsiveness of the humans depicted (Freeth, Foulsham, & Kingstone, 2013; Großekathöfer, Seis, & Gamer, 2021; Laidlaw, Foulsham, Kuhn, & Kingstone, 2011) and by the laboratory setting itself (Risko et al., 2016). Thus, new approaches are needed to bridge the gap between controlled and natural environments and to consequently increase ecological validity (cf. Schilbach, Timmermans, Reddy, Costall, Bente, Schlicht, & Vogeley, 2013). This is especially true if one aims to assess whether social attention has any real-world relevance. One possibility to assess this relevance is to explicitly ask participants which regions of a stimulus they find interesting, and correlating this measure with their fixations on that stimulus (Masciocchi, Mihalas, Parkhurst, & Niebur, 2009). Another interesting option to achieve this goal was suggested by Wang, Fan, Chen, Hakimi, Paul, Zhao, and Adolphs (2016). They analyzed the content of photographs taken by a sample of patients with autism spectrum disorders compared to healthy control subjects. Professional and naïve raters classified the photographs regarding image properties and content. They discovered that participants on the autism spectrum not only needed longer to complete the task but also took images of worse quality (e.g. more blurry and occluded photographs) and, surprisingly, more photographs containing people. Such self-taken photographs within a natural environment seem to reflect meaningful differences and might thus represent an ecologically valid measure of social preference. However, it is unclear how social preferences as reflected in idiosyncratic gaze patterns relate to such behavioral characteristics.

The aims of the current study were twofold. First, we were interested in whether gaze patterns toward
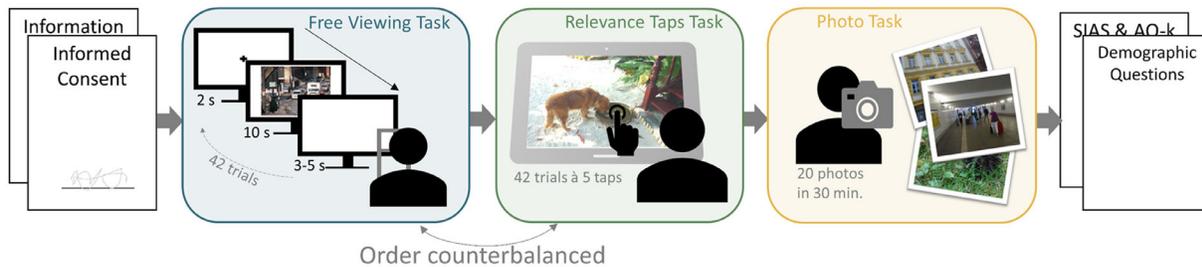
Figure 1. Visual depiction of the procedure.

social information would be intra-individually stable across different stimuli, and, second, whether these gaze patterns would relate to other behavioral measures of social preference. To test these hypotheses, we used three different tasks (see Figure 1): a free viewing task of naturalistic social scenes to measure social attention using eye-tracking methodology, a relevance taps task on a tablet computer using similar stimuli to assess the regions of explicit personal relevance, and a photograph-taking task outside of the laboratory to detect natural preferences for including humans in the pictures. We assumed social attention to constitute an idiosyncratic, stable, and meaningful disposition. Thus, we expected that individual differences in fixation measures on heads and bodies of depicted individuals in the free-viewing task correlate with corresponding metrics in the relevance taps task as well as with the number of human beings portrayed in the photograph-taking task.

Methods

Participants

Fifty-one volunteers participated in the study. One participant was excluded due to a lack of understanding of the task which led to the final sample size of 50 participants (34 women; age: $M = 25.28$ years; $SD = 7.97$ years). This sample size allowed for the detection of an effect size of $r = 0.34$ on an alpha level of 5% with a power of 80%, which we deemed sufficient. Participants with poor eyesight had to correct their vision with contact lenses or were not able to take part in the study. Recruiting was conducted using the participant pool of the University of Würzburg. All participants gave written informed consent and received either 1.5 hours of course credit or a monetary reward of 15 euros. The study was preregistered on the platform AsPredicted.org (https://aspredicted.org/ry5uz.pdf) and approved by the ethics committee of the University

of Würzburg in accordance with the Declaration of Helsinki.

Questionnaires

Because previous studies reported tentative differences in social preferences in autism spectrum disorder and social anxiety samples (e.g. Fletcher-Watson et al., 2009; Jones & Klin, 2013; Rösler et al., 2021; Speer et al., 2007; Wang, Jiang, Duchesne, Laugeson, Kennedy, Adolphs, & Zhao 2015; Wang et al., 2016; Wieser et al., 2009), we characterized our sample regarding such traits using the German short version of the Autism Spectrum Quotient (AQ-k; English original version: Baron-Cohen, Wheelwright, Skinner, Martin, & Clubley, 2001; German version: Freitag, Retz-Junginger, Retz, Seitz, Palmason, Meyer, Rösler, von Gontard et al., 2007) as well as the Social Interaction Anxiety Scale (SIAS; English original version: Mattick & Clarke, 1998; German version: Stangier, Heidenreich, Berardi, Golbs, & Hoyer, 1999). As expected for such a convenience sample, scores on these two questionnaires were rather low on average (Table 1). Exploratory analyses did not reveal any significant correlations between the AQ-k or SIAS scores and the different measures of social preference across the three tasks (see Supplementary Figure S1). In addition, participants filled in a socio-demographic and a follow-up questionnaire asking four questions regarding the photograph task. The follow-up questionnaire was used to identify whether the individual subjects followed a specific theme (e.g. taking pictures of flowers) and to better characterize their experience and proficiency in using a camera.

	M	SD	Minimum	Maximum	Cronbach's α
AQ-k	7.58	5.10	0	21	0.82
SIAS	18.66	10.14	5	50	0.89

Table 1. Descriptive statistics of the Autism Quotient (AQ-k) questionnaire and the Social Interaction Anxiety Scale (SIAS).

Tasks

Participants carried out three different tasks with the order of the first two tasks being counterbalanced across participants (see Figure 1). In the free viewing task, 42 images of complex social scenes were presented for 10 seconds each while eye-tracking data were recorded. The first two images were used to familiarize participants with the task and were thus not analyzed. A fixation cross was continuously presented during the intertrial interval that lasted 3 to 5 seconds (randomly determined from a uniform distribution). Participants were instructed to look at the fixation cross during the intertrial interval to ensure a constant starting point for their visual exploration but they did not receive specific instructions on how to look at the images.

The so-called relevance taps task was based on previous research using explicit ratings of interestingness in addition to eye-tracking recordings (e.g. Masciocchi et al., 2009). For example, maps of manually selected points of interest were shown to better predict fixation locations than low-level physical features across a variety of different images (Onat et al., 2014). For the current study, we decided to ask participants to mark relevant instead of interesting image regions to specifically assess individual differences in scene understanding. To this aim, participants received a tablet computer and were presented with a total of 42 complex social scenes different from the free viewing task. The first two images were again used for familiarization and removed from further analyses. For each image, participants were instructed to explicitly select five regions they deemed particularly important to understand the respective image content. This was achieved by sequentially tapping with the finger on five areas of highest personal relevance in descending order. Each single tap was visually highlighted on the image as a semitransparent colored circle including the tap number (1 to 5). Participants were free to place these relevance marks within the image region and could also select the same region more than once. Images were displayed until all five regions were selected and participants confirmed their selection. In case of errors, participants could remove previously selected regions and restart the trial. On average, participants needed 13.61 seconds ($SD = 6.88$ seconds) to select the five regions per image, which is roughly similar to the presentation duration in the free viewing task.

Finally, each participant was handed a digital camera and was instructed to leave the laboratory to take 20 pictures within 30 minutes. There was no specific instruction given regarding a motif or theme that could have influenced the content of the photographs taken. The participants were simply asked to take pictures of what they thought was interesting. From the starting point at our laboratory near the city center of Würzburg, Germany, participants could visit very different locations in the available time, including crowded marketplaces, quiet side streets, the main train station, parks, and diverse shops. Moreover, data collection took place in summer between approximately 9 a.m. and 6 p.m. Thus, weather and daylight conditions as well as the possibility to meet other people around the laboratory were largely comparable between participants. In the follow-up questionnaire, 11 participants noted that they talked to other people outside of the laboratory during the photograph task. In addition, most of the participants stated to rarely spend time on taking photographs in their everyday life, $M = 2.51$ ($SD = 1.39$) on a scale of 1 (very little time) to 6 (very much time). However, most participants reported to be somewhat experienced in using a camera, $M = 3.39$ ($SD = 1.19$) on a scale of 1 (unexperienced) to 6 (very experienced). The average time participants needed to complete the photograph task amounted to $M = 27.62$ minutes ($SD = 6.20$ minutes; range = 8–40 minutes).

Stimuli

For the free viewing and the relevance taps task, 80 naturalistic images (1200 × 900 pixels) with social content were used from the stimulus set of End and Gamer (2017) and End and Gamer (2019). Four additional stimuli were added to familiarize participants with the respective tasks. Half of the images were randomly selected for each participant and presented during the free viewing task while recording eye-tracking data. The other half was presented during the relevance taps task on a tablet computer. The images were taken from different sources (Emotional Picture Set; [EmoPicS; Wessa, Kanske, Neumeister, Bode, Heissler, & Schöenfelder 2010]; International Affective Picture System [IAPS; Lang, Bradley, & Cuthbert, 2008]; McGill Calibrated Colour Image Database [Olmos & Kingdom, 2004]; Nencki Affective Picture System [NAPS; Marchewka, Żurawski, Jednoróg, Grabowska, 2014]; Object and Semantic Images and Eye tracking dataset [OSIE; Xu et al., 2014], and the internet [e.g. Google picture search, flickr]) and fulfilled the following criteria. All images were realistic (as opposed to drawn or artificially generated) and had sufficient depth of field and a rather high complexity with distributed visual information across the scene. The images contained naturalistic environments like street scenes, landscapes, and scenarios that were located inside and outside of buildings. A broad variety of social features (i.e. heads and bodies) was depicted within the environments and it was ensured that these social regions of interest did not strongly overlap with physically salient image parts. Finally, in order to reduce the impact of a possible central
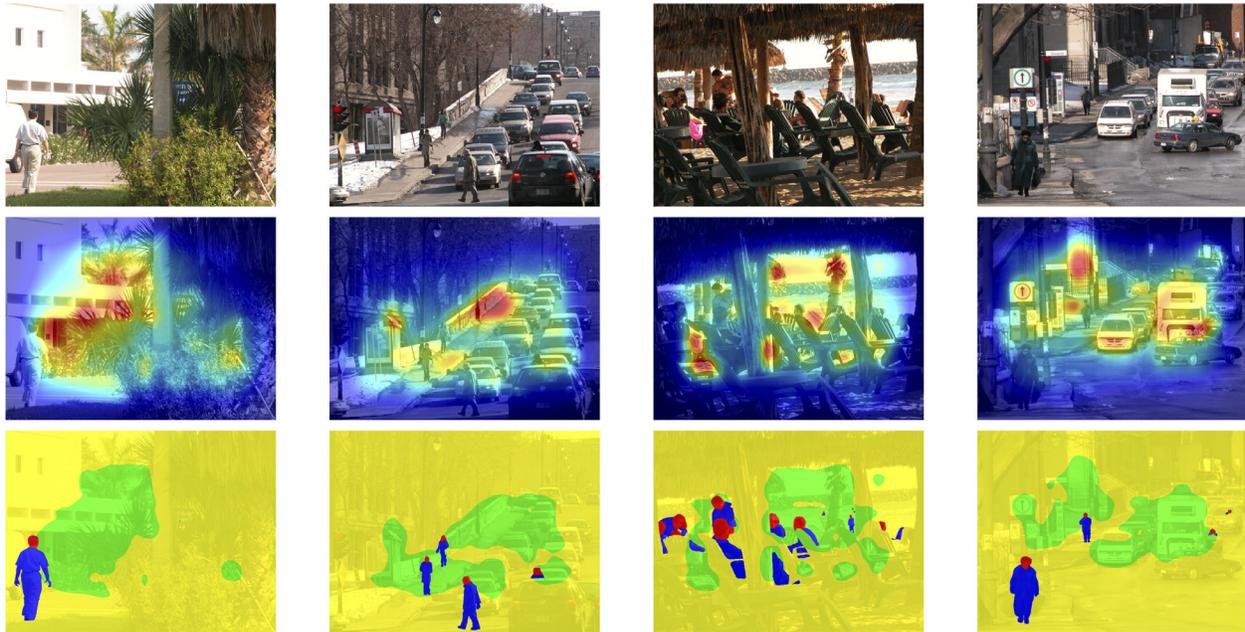
Figure 2. Illustration of the stimulus material and the definition of regions of interest (ROI) for the analyses. First row: Examples of stimuli that were used for the free viewing and the relevance taps task. Second row: The corresponding heat maps for physical saliency (computed according to Harel et al., 2007). Third row: ROI maps (red = heads, blue = bodies, green = high saliency, and yellow = low saliency) for these stimuli.

fixation bias (Foulsham, Walker, & Kingstone, 2011; Tatler, 2007), social elements were distributed across the scene instead of being concentrated at the image center. Example images are depicted in Figure 2.

Measurement devices

Eye-tracking was carried out in a dimly lit, sound attenuated room with a stationary Eyelink 1000 Plus system (SR Research, Ottawa, ON, Canada) in the tower mount configuration. The right eye was tracked with a sampling rate of 1000 hertz (Hz). Stimulus presentation and data recording were controlled using Presentation 17.0 (Neurobehavioral Systems, Inc., Berkeley, CA, USA). Images were presented on a 24 inch Asus VG248QE display with an effective screen size of 53.1 × 29.9 cm, a resolution of 1920 × 1080 pixels and a refresh rate of 60 Hz. The fixed viewing distance between the participant and the screen was approximately 53 cm. The size of each stimulus on the screen was 33.2 × 24.9 cm, which corresponds to a visual angle of 34.9 degrees × 26.5 degrees.

Personal relevance taps were recorded on a 10.8 inch tablet computer (Dell Latitude 5175) with a display resolution of 1920 × 1080 pixels and a physical display size of 23.9 × 13.4 cm. Again, the software Presentation was used for the stimulus presentation. The size of each stimulus was 14.9 × 11.2 cm on the computer display. We did not control the viewing distance during the task but for a comfortable position, the tablet was typically held at a distance of approximately 40 cm which corresponds to a visual angle of 21.2 degrees × 15.9 degrees for the depicted images. At such a distance, the diameter of the relevance marks that were placed using finger taps amounted to approximately 2 degrees of visual angle.

A Sony DSC-HX60V cyber shot camera was used by the participants during the photograph task. The automatic mode was activated by default, which ensured the same starting prerequisite for the participants and a similar quality of the photographs, without any post processing. Participants were allowed to use the zoom of the camera to focus on far distance objects. All photographs were taken with a resolution of 5184 × 3888 pixels.

Procedure

At the beginning of each examination, participants gave written informed consent and were given both an oral and detailed written description of the procedure of the study. Subsequently, each participant carried out the three different tasks. The order of the first two tasks was counterbalanced across participants. For the free viewing task, a 9-point calibration, followed by a validation procedure, was conducted. A maximal distance of 1 degree between calibration and validation was tolerated and the procedure was repeated if

necessary. In the relevance taps task, naturalistic scenes were presented on the tablet computer and participants were instructed to mark those five image regions that seemed most relevant to understand the depicted scene. After completing the first two tasks, participants were instructed to leave the laboratory for approximately 30 minutes to take 20 pictures outside for the photograph task. They were briefly introduced to the major functions of the digital camera (e.g. how to zoom, take, and delete a picture). Notably, participants were told not to change the camera settings. After returning to the laboratory, the number of photographs was determined by the experimenter. If the participant took more than 20 photographs, the experimenter asked to decide which photograph should be deleted in order to establish a collection of exactly 20 photographs. While the pictures were counted, the participants filled in the questionnaires (SIAS, AQ-k, the socio-demographic questions, and the follow-up on the photograph task). Finally, participants were thanked and compensated for participating.

Data processing

The eye-tracking data of the free viewing task were preprocessed similarly as in End and Gamer (2017) to obtain the location of the first five fixations as well as the relative fixation density for each stimulus and participant during the free viewing task. EyeLink's default configuration was used to parse gaze data into fixations and saccades. In detail, saccades were identified as eye movements exceeding a velocity threshold of 30 degrees/second or an acceleration threshold of 8000 degrees/second^2. Fixations were identified as time periods between saccades.

To account for and correct drifts in the eye-tracking data over time, the average fixation position during the 300 ms before stimulus onset was subtracted from the trial's fixation data. During this epoch, the central fixation cross was shown on the display screen. To identify trials where participants failed to fixate the cross, we used a previously established recursive outlier detection procedure separately on x- and y-coordinates (see End & Gamer, 2017; End & Gamer, 2019). In detail, the lowest and highest values were temporarily removed from the distribution and it was checked whether one or both deviated more than three SDs from the average of the remaining gaze coordinates. If this was the case, the respective trial was marked as an outlier and the procedure was repeated until no more trial fulfilled this criterion. Finally, outliers ($M = 3.42$ trials, $SD = 2.91$) were replaced by the average of the baseline coordinates of valid trials (average baseline correction of outlier trials: $M = 1.87$ degrees, $SD = 1.34$ degrees of visual angle). Furthermore, single trials were excluded from further analyses if the relative aggregated time of missing data during scene presentation (e.g. due to blinks) exceeded 20% ($M = 0.20$ trials, $SD = 0.57$). All data from a participant would have been excluded if more than 50% of trials were considered invalid. This did not apply to any of the participants in the current study.

Fixation density maps were created as described in End and Gamer (2017). All fixations that started after stimulus onset and occurred during stimulus presentation were considered for further analysis. First, for creating fixation density maps, for each participant and image, an empty two-dimensional map of the size of the stimulus (1200 × 900 pixels) was created. Individual fixations, weighted by their durations in milliseconds, were added to these maps. Afterward, the resulting maps were spatially smoothed using a two-dimensional isotropic Gaussian kernel with a standard deviation of 36 pixels whose width amounts to approximately 2 degrees of visual angle and reflects the size of the functional field of the human fovea centralis. Finally, fixation density maps were normalized to range from 0 to 1. A similar procedure was used to calculate relevance density maps using the coordinates of the relevance taps instead of fixations. No weighting with respect to durations was applied but all other processing steps (i.e. smoothing and normalization) were identical. Finally, in order to facilitate a direct comparison between fixation and relevance density maps, we also constructed fixation density maps for the first five fixations only.

Regions of interest (ROI) were primarily defined with respect to social content. Based on recent findings that fixation densities on head and body regions tend to be anticorrelated (e.g. Broda & de Haas, 2022), we decided to differentiate between these two social elements. Therefore, using the software GIMP 2.8.22 (https://www.gimp.org), human heads and bodies were manually drawn for each picture. Bodies were defined as excluding heads and including torso, arms, hands, legs, and feet. Furthermore, in order to allow for a comparison between social ROIs and areas of high physical saliency (cf. End & Gamer, 2017), we calculated saliency maps using the Graph-Based Visual Saliency (GBVS) model of Harel, Koch, and Perona (2007), that generally performs well in predicting fixation locations (Judd, Durand, & Torralba, 2012). Saliency values of pixels outside social ROIs were classified as reflecting high physical saliency if the respective value exceeded the eighth decile of the saliency distribution. This criterion was, as End and Gamer (2017) already noted, arbitrary, but the cut-off fulfilled the purpose of effectively distinguishing high- from low-saliency regions (see Figure 2 for example images, corresponding physical saliency maps, and resulting ROIs; further characteristics on these ROIs are given in Table S1 of the supplement). Finally, these ROIs were used to determine the trial-wise average fixation or relevance

density, respectively, for head and body ROIs as well as regions of high physical saliency. Therefore, the sum of fixation or relevance density values per ROI was divided by the sum of density values for the whole scene and this proportion was then normalized by the area of the respective ROI to control for the issue that larger areas typically receive more attention than smaller ones. Moreover, in order to compare the current results to the analyses of End and Gamer (2017), we also calculated the relative proportion of fixations/relevance taps on these ROIs for the first five individual fixations/relevance taps and normalized these scores by dividing them by the mean area of the respective ROI across all scenes whose data were represented in the respective relative frequency score.

Photograph classification was realized for 999 photographs because one participant accidentally took 19 instead of 20 photographs. Three additional photographs were excluded because they were classified as blurry. In sum, photograph analysis was conducted for 996 photographs. All photographs were initially examined and classified by one rater. The main variable of interest was the average number of humans and faces in the photographs taken by each individual participant. In addition, we determined the proportion of photographs that included (any number of) humans and faces. People were counted if they were identifiable as individuals and were visible without zooming into the photograph. Faces were counted if they were clear enough to hypothetically recognize a specific person. To ensure robustness of classification criteria, a second rater additionally classified the photographs of a random subset of five participants. Inter-rater agreement was high as reflected in a Kappa value of 0.90 for the binary category "people depicted in the photograph" and 0.74 for the category "faces depicted in the photograph." The same was true for the numerical values concerning the "number of people depicted in the photographs" with an intraclass correlation of 0.97 and for "number of faces depicted in the photographs" with an agreement of 0.80 (examples of photographs from a person with high social preference are shown in Figure S2 and another participant with low social preference in Figure S3 of the supplement).

Statistical analyses

Data were preprocessed and analyzed using R (R Core Team, 2021) and RStudio (2020). The a priori significance level was set to $\alpha = 0.05$. To test whether a preference for social information reliably occurred across stimuli within each task, average fixation and relevance densities were z-standardized for each type of ROI across participants before calculating the split-half reliability as an index of individual stability of viewing patterns.

Regarding this analysis, we deviated from the preregistration protocol and relied on permuted split-half correlations instead of Cronbach's α because individual participants viewed a different, randomly determined set of images in both tasks and thus Cronbach's α could not be directly computed due to different patterns of missing data across participants. For the analysis of split-half correlations, 1000 permutations were carried out. In each permutation, the trials of each participant were randomly subdivided in two halves and the z-transformed fixation densities per ROI were averaged before the Pearson's correlation coefficient of these halves across participants was calculated. The resulting 1000 correlations were Fisher-Z-transformed, averaged, and inversely transformed to yield the average correlation. These mean correlations were additionally adjusted using the Spearman-Brown prophecy formula. This analysis was carried out independently across all fixations as well as the first five fixations of each participant in the free viewing task, and for the relevance taps task data. A similar, non-preregistered permutation analysis was conducted for the photograph task. In this case, the 20 pictures taken per participant were randomly split into two halves, the proportion and average number of humans/faces were calculated per half, and these were correlated across participants. This analysis was also repeated 1000 times and the Pearson's correlation coefficients were averaged using the Fisher-Z-transformation and subsequently Spearman-Brown corrected.

To address the second hypothesis, we correlated the z-standardized averaged fixation densities for each ROI (head, body, and high saliency) between both tasks (free viewing and relevance taps) to test whether similar ROIs were preferred across tasks. We included the fixation densities across all fixations in the free viewing task, as well as the fixations densities of only the first five fixations per trial. Similarly, the four measures derived from the photograph task (proportion of photographs containing humans, proportion of photographs containing faces, the average number of depicted humans, and the average number of depicted faces) were correlated with the average fixation density and relevance taps density for the social ROIs (head and body). The Benjamini-Hochberg procedure was used to correct for multiple comparisons.

To replicate the group findings of End and Gamer (2017) and extend them to relevance taps, we analyzed whether the fixation/relevance taps density patterns across ROIs would vary between the two tasks. Therefore, we calculated for each participant the average of relative area-normalized fixation/relevance densities for each ROI across all images. These values were analyzed using a two-way repeated-measures ANOVA with task (free viewing and relevance taps) and ROI (head, body, areas of lower saliency, and areas

of higher saliency) as within-subject factors. Planned contrasts were calculated to further characterize significant interactions between task and ROI. A similar ANOVA was carried out with the fixation densities restricted to the first five fixations per stimulus.

Finally, we were interested in differences in the temporal dynamic of early fixations and relevance taps. We therefore analyzed the relative area-normalized frequency of fixations and relevance taps per ROI, respectively. A three-way repeated-measures ANOVA was conducted relating these values to task (free viewing and relevance taps), ROI (head, body, areas of lower saliency, and areas of higher saliency) and fixation (tap) number (1 to 5). Partial eta-squared values (η_p^2; Bakeman, 2005) are reported as effect size estimates for all ANOVAs.

Results

Reliability of preferences within tasks

Results of the permutation analysis of all fixations of the free viewing data revealed significant split-half correlations for all three ROIs (head, body, and high saliency; see Figure 3). This correlation was especially high for fixations on the head ($r = 0.81$, Spearman-Brown adjusted: $r_{SB} = 0.89$) as well as highly salient regions ($r = 0.66$, $r_{SB} = 0.79$), whereas it was somewhat lower for the body region ($r = 0.51$, $r_{SB} = 0.68$). All 1000 permuted correlations were above zero. The permuted correlations for only the first five fixations were slightly lower (head: mean $r = 0.72$, $r_{SB} = 0.84$, all correlations positive; body: mean $r = 0.52$, $r_{SB} = 0.69$, all correlations positive; high physical saliency: mean $r = 0.25$, $r_{SB} = 0.40$, 99.3% of permuted correlations were positive). Especially the fixations on highly salient regions resulted in rather high split-half correlations when considering all fixations but substantially lower values when analyses were restricted to the first five fixations. Finally, the permutation analysis of the relevance taps task for the same ROIs resulted in the strongest significant correlation for the head region (mean $r = 0.89$, $r_{SB} = 0.94$), followed by body regions (mean $r = 0.76$, $r_{SB} = 0.86$) and physically salient regions (mean $r = 0.50$, $r_{SB} = 0.67$, all correlations above zero). In both tasks, fixations and relevance taps were thus very stable, especially for the head ROI. Participants who looked at or selected a head in one picture likely also looked at or selected heads in other pictures.

For the photograph task, the permuted split-half correlations were high for the proportion of pictures
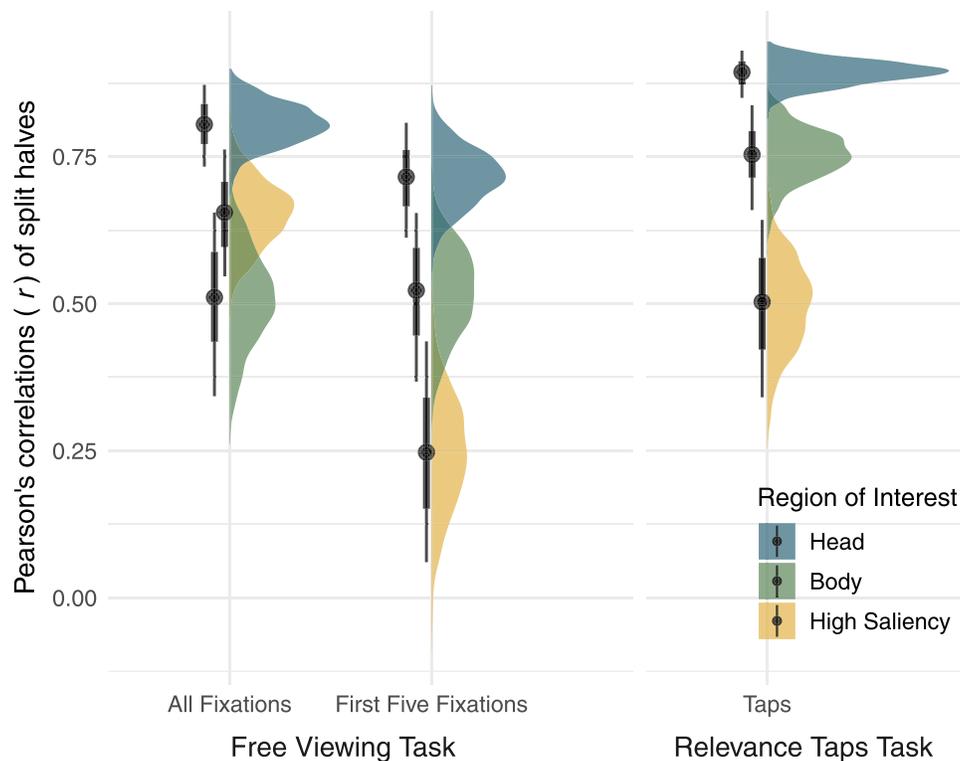

Figure 3. Densities of the 1000 Pearson correlation coefficients between random split halves of the data per ROI, for the Free Viewing and Relevance Taps Tasks, obtained in the permutation analysis. Points on the left-hand side represent median correlation coefficients, the thick interval is the interquartile range (IQR) and the thin line represents the 95% data range.

that contained humans (proportion humans: $r = 0.72$, $r_{SB} = 0.84$, all correlations above zero) and only slightly lower for the proportion of pictures containing faces (proportion faces: $r = 0.59$, $r_{SB} = 0.74$, all correlations above zero). The split-half correlations for the average number of humans or faces depicted per picture were slightly lower but still medium-high to high (number humans: $r = 0.65$, $r_{SB} = 0.79$, all correlations positive; number faces: $r = 0.45$, $r_{SB} = 0.62$, 99% of correlations positive, see Figure 4).

Correlation of preferences across tasks

In order to assess whether social ROIs would be similarly preferred across tasks, we correlated all measures with each other, with the main focus on whether the fixation density on, for example, the head region correlates with the relevance density on the same region. The respective correlations (see highlighted diagonal in Figure 5) were separately analyzed and adjusted using the Benjamini-Hochberg procedure. We found significant moderate correlations between all fixations (free viewing) and relevance taps for both head and highly salient ROIs (see Figure 5A). For the body ROI, the correlation between tasks was non-significant. In addition, the correlations between the tasks for the first five fixations (free viewing) and relevance taps for each ROI showed the same descriptive pattern of results but correlations were lower and did not reach statistical significance (see Figure 5B). Scatterplots for these analyses are shown in Supplementary Figure S4.

To examine whether social preferences generalize to real-world behavior, we correlated fixation and relevance densities on head and body ROIs with
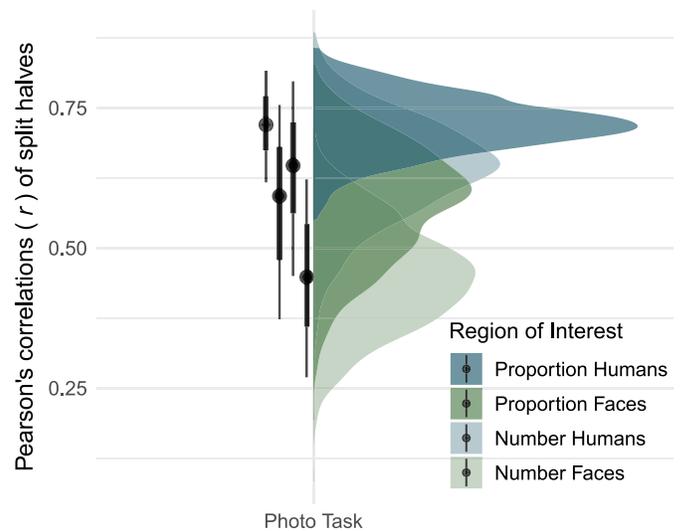

Figure 4. Densities of the 1000 Pearson correlation coefficients between random split halves of the data per measure obtained in the permutation analysis for the Photograph Task. Points on the left-hand side represent median correlation coefficients, the thick interval is the interquartile range (IQR) and the thin line represents the 95% data range.

the content categories of the photograph task (i.e. proportion of photographs containing humans or faces, and the average number of humans or faces depicted in the photographs). All p values were again adjusted for multiple comparisons using Benjamini-Hochberg procedure. First, for all fixations of the free viewing task, head fixations were only slightly and non-significantly correlated with the proportion of photos containing humans, and were moderately and non-significantly correlated with the number of

Figure 5. Correlation of the mean density between the ROIs of the free viewing and relevance taps tasks for all fixations (**A**) and the first five fixations (**B**). According to our hypotheses, only correlations on the highlighted diagonal were tested for statistical significance with p values adjusted for multiple comparisons according to the Benjamini-Hochberg procedure. *$p < 0.05$ (two-tailed), **$p < 0.01$ (two-tailed), $N = 50$.
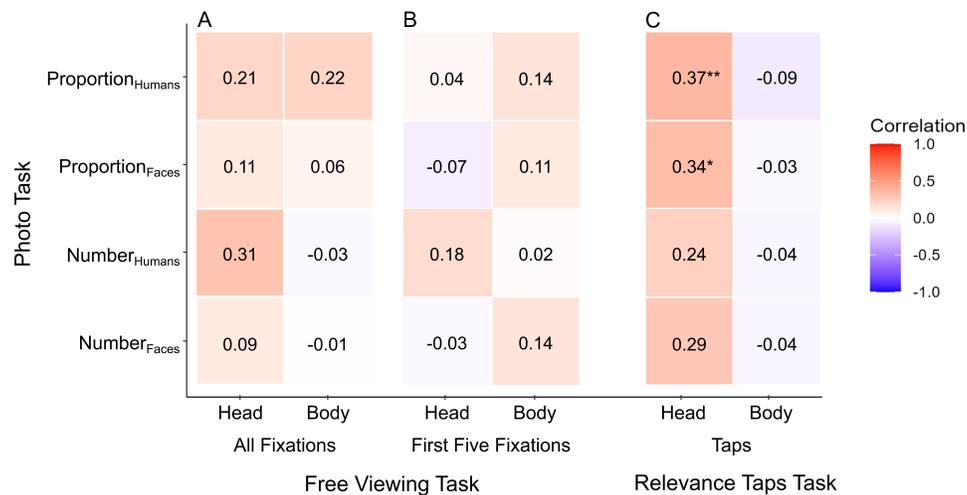
Figure 6. Correlation of the mean densities in head and body regions of interest in the free viewing and the relevance taps task with the social categories of the photograph tasks. (**A**) All fixations, (**B**) first five fixations, (**C**) relevance taps. Photograph content: Proportion$_{Humans}$ = proportion of photographs containing humans, Proportion$_{Faces}$ = proportion of photographs containing faces, Number$_{Humans}$ = average number of humans depicted in the photographs, Number$_{Faces}$ = average number of faces depicted in the photographs. The *p* values were adjusted for multiple comparisons according to the Benjamini-Hochberg procedure. *$p < 0.05$ (two-tailed), **$p < 0.01$ (two-tailed), $N = 50$.

humans depicted in the photographs (Figure 6A). Second, we only observed a small non-significant correlation between body fixations (free viewing) and the proportion of photographs containing humans. A similar pattern but with slightly lower correlations was observed when restricting analyses to the first five fixations (Figure 6B). Third, densities of relevance taps on head regions correlated similarly with all four photograph-based measures (Figure 6C), although the correlation with the number of humans and faces depicted did not reach statistical significance. Last, body relevance densities did not correlate with any of the photograph-based measures. Scatterplots for these analyses are shown in Supplementary Figures S5, S6, and S7.

Group-level differences between the free viewing and relevance taps tasks

In order to replicate and extend previous findings of an attentional prioritization of social ROIs, we carried out a two-way repeated-measures ANOVA on the relative area-normalized sum of density values with the factors task (free viewing and relevance taps) and ROI (head, body, high saliency, and low saliency). The analysis revealed significant main effects of task ($F(1,49)$ = 18.15, $p < 0.001$, $\eta_p^2 = 0.27$), and ROI ($F(3,147)$ = 361.90, $p < 0.001$, $\eta_p^2 = 0.88$), as well as a significant interaction of task and ROI ($F(3,147) = 62.07$, $p < 0.001$, $\eta_p^2 = 0.56$). Visual inspection of Figure 7 and pairwise comparisons (see Table 2) revealed that heads were most frequently fixated/selected, followed by

bodies and areas of higher physical saliency. Areas of lower saliency were looked at and tapped at least. This prioritization of heads was more pronounced in the free viewing than in the relevance taps condition whereas body ROIs were more frequently selected in the latter task. Non-social ROIs (i.e. areas of higher and lower physical saliency) were similarly tapped and fixated in both tasks.

Restricting the analysis to the first five fixations also revealed significant main effects of task, $F(1,49)$ = 274.97, $p < 0.001$, $\eta_p^2 = 0.85$, and ROI, $F(3,147)$ = 522.11, $p < 0.001$, $\eta_p^2 = 0.91$, as well as a significant interaction of both factors, $F(3,147) = 243.98$, $p < 0.001$, $\eta_p^2 = 0.83$ (see Figure 7). Contrasts revealed an even larger viewing preference for heads for the first five fixations as well as a slightly enhanced tendency to fixate bodies rather than selecting them in the relevance taps task (see Table 2). Again, the difference between the two tasks for the non-social ROIs appears to be small but significant. Highly salient regions were fixated more than tapped on, whereas the opposite was true for low salient regions.

Time course analysis of the first five fixations and relevance taps

To shed light on the temporal dynamics of early fixations and relevance taps, we calculated a three-way repeated measures ANOVA on the relative area-normalized frequency of fixations and relevance taps, respectively, with the factors task (free viewing or
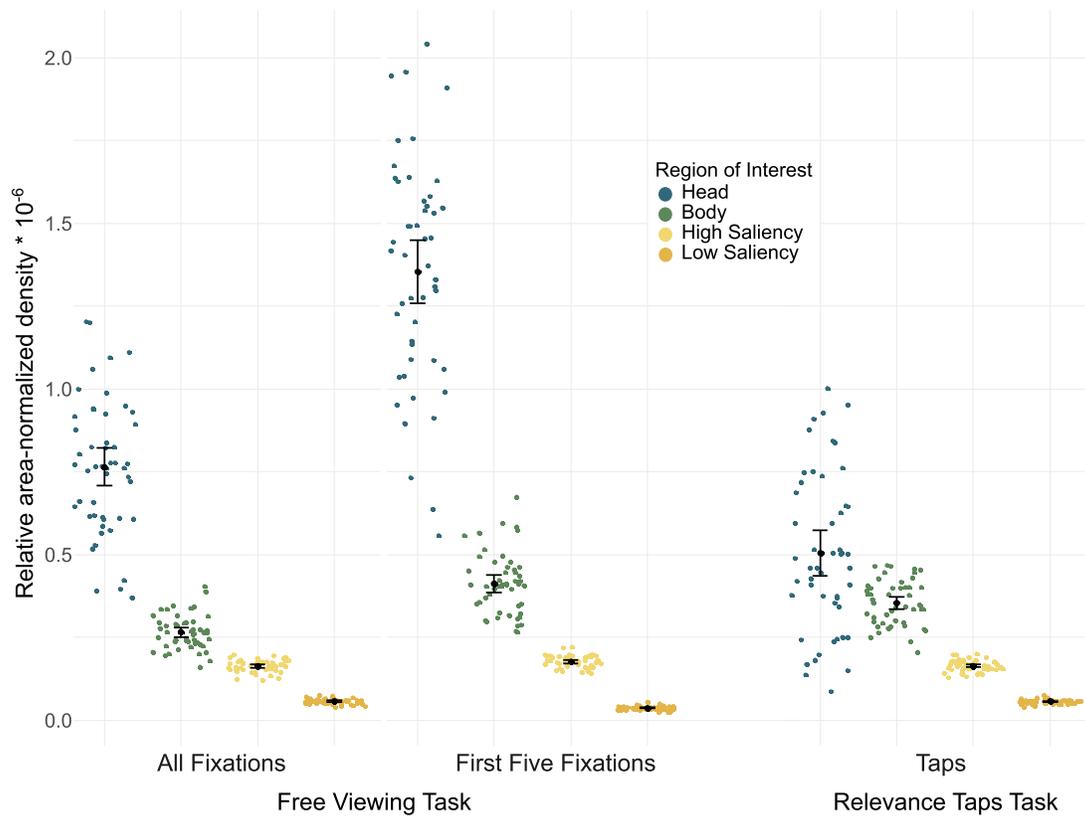
Figure 7. Relative area-normalized fixation and relevance taps density on the four regions of interest as a function of the task. Fixations and relevance taps density were most pronounced on the ROI head followed by body, high and low saliency. Points were jittered in each category across the x-axis. Error bars indicate 95% confidence intervals and black dots the mean in each category.

ROI	Contrast	Estimate	Standard error	Lower CI	Upper CI	$t(49)$	*p* value
Head	FV_{all} - RT	260.31×10^{-8}	3.63×10^{-7}	18.75×10^{-7}	33.32×10^{-7}	7.18	<0.001
Body	FV_{all} - RT	-88.38×10^{-8}	1.35×10^{-7}	-11.56×10^{-7}	-6.12×10^{-7}	-6.52	<0.001
High saliency	FV_{all} - RT	-0.03×10^{-8}	0.29×10^{-7}	-0.58×10^{-7}	0.57×10^{-7}	-0.01	0.993
Low saliency	FV_{all} - RT	0.51×10^{-8}	0.11×10^{-7}	-0.18×10^{-7}	0.28×10^{-7}	0.45	0.654
Head	FV_5 - RT	848.36×10^{-8}	5.01×10^{-7}	74.77×10^{-7}	94.90×10^{-7}	16.94	<0.001
Body	FV_5 - RT	58.43×10^{-8}	1.88×10^{-7}	2.07×10^{-7}	9.62×10^{-7}	3.11	0.003
High saliency	FV_5 - RT	12.86×10^{-8}	0.32×10^{-7}	0.64×10^{-7}	1.94×10^{-7}	3.99	<0.001
Low saliency	FV_5 - RT	-19.49×10^{-8}	0.11×10^{-7}	-2.18×10^{-7}	-1.72×10^{-7}	-17.21	<0.001

Table 2. Contrasts between the free viewing and the relevance taps task within each Region of Interest (ROI). *Note:* FV_{all} = relative area-normalized sum of densities of all fixations during the Free Viewing Task, FV_5 = relative area-normalized sum of densities of the first five fixations per trial in the Free Viewing Task, RT = relative area-normalized sum of densities of the Relevance Taps, CI = confidence interval, *p* values are adjusted for multiple comparisons using the Tukey method.

relevance taps), ROI (head, body, high saliency, and low saliency) and fixation or tap number (1 to 5).

All three main effect were significant (task: $F(1,49) = 156.62$, $p < 0.001$, $\eta_p^2 = 0.76$; number: $F(4,196) = 70.82$, $p < 0.001$, $\eta_p^2 = 0.59$; ROI: $F(3,147) = 408.77$, $p < 0.001$, $\eta_p^2 = 0.89$). Similar to the previous analysis, the interaction between task and ROI was also significant, $F(3,147) = 148.37$, $p < 0.001$, $\eta_p^2 = 0.75$, as were the interactions between task and (fixation or tap) number, $F(4,196) = 16.27$, $p < 0.001$, $\eta_p^2 = 0.25$, and between ROI and (fixation or tap) number, $F(12,588) = 39.79$, $p < 0.001$, $\eta_p^2 = 0.45$. The three-way interaction of interest (task × ROI × number) also reached statistical significance, $F(12,588) = 16.44$, $p <$
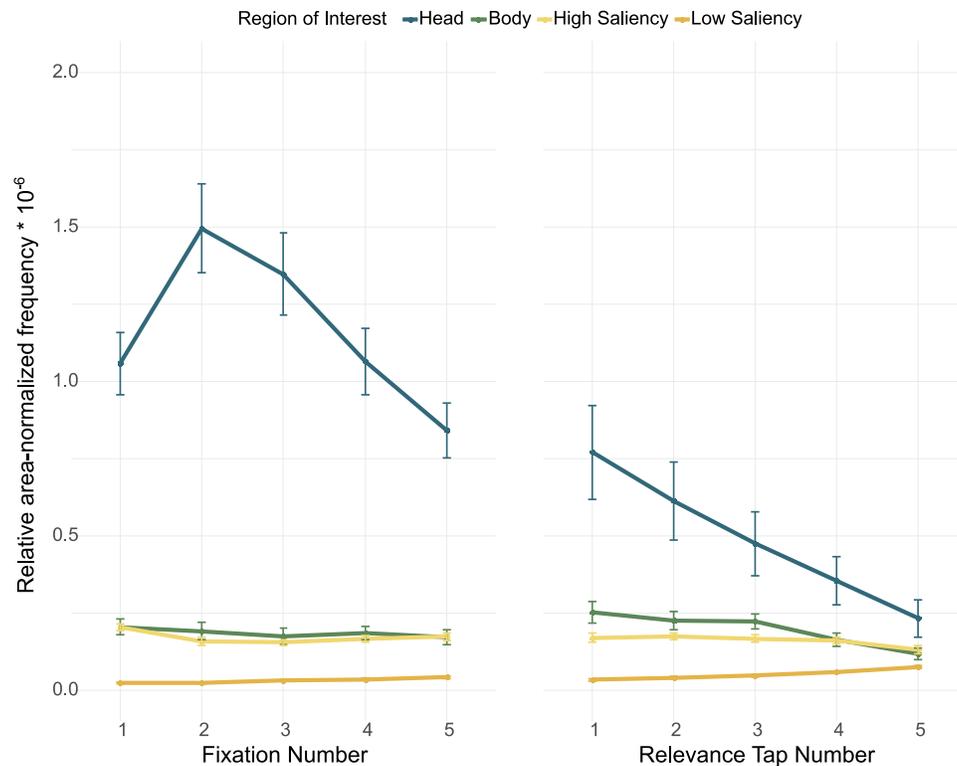
Figure 8. Time course of relative area-normalized fixation and relevance taps frequency on the four regions of interest as a function of the task. Fixations and relevance taps on heads were most pronounced early after stimulus onset. Error bars represent 95% confidence intervals.

0.001, $\eta_p^2 = 0.25$. Visual inspection of Figure 8 revealed that the head region was prioritized in both tasks but received most attention with the second fixation in free viewing conditions. By contrast, relevance taps on the head decreased monotonically as a function of tap number. Body ROIs also tended to receive more attention with earlier fixations/taps but the pattern was less pronounced. Finally, a slight increase in fixation/relevance taps proportions with increasing fixation/tap number was observed for regions of low physical saliency.

Discussion

The goal of the present study was to identify to which extent attentional preferences for social information are stable and relevant for scene perception and real-world behavior. Specifically, we were interested to find out whether social attention, the preferential fixation of social aspects of complex scenes, would reflect an intra-individually stable trait that generalizes to other behavioral measures of social preference. To address this question, participants carried out three tasks: (1) a free viewing task; (2) a relevance taps task, in which they judged which regions of an image are most relevant for understanding a given scene; and (3) a

photograph-taking task, whereby participants could freely take pictures outside the laboratory. We expected that the preferential fixation of social information in the free viewing task would be intra-individually stable and would correlate with personal relevance judgements during the relevance taps task as well as with the social content of self-takes photographs. In addition, we analyzed whether fixation and relevance tap patterns differed between ROIs and time points to further corroborate similarities and differences between the two tasks.

Our results largely supported our hypotheses. Gaze behavior during free viewing revealed stable intra-individual preferences, especially for social ROIs (head and body) and, at least to some extent, also for image regions with high physical saliency. These correlations indicate that participants who preferred to look at social aspects of one stimulus also did so for other scenes which is consistent with previous research (de Haas et al., 2019; Guy et al., 2019; Linka & de Haas, 2020). An even higher intra-individual stability for head and body regions was found in the relevance taps. Thus, participants differed in how relevant they considered the portrayed people to understanding the meaning of the scene, but they did so consistently across the whole stimulus set. The permuted split-half correlations for the photograph task were also surprisingly high. Participants who took pictures of other people did so

across pictures. This was true for both the proportion of pictures displaying humans and faces. The correlations were slightly lower when accounting for the actual number of humans or faces in the pictures.

Moreover, even though the free viewing and the relevance taps task differed in several aspects (e.g. refixations were possible in the former task but relevance taps on the same location – although generally permitted – were rather unlikely), we observed significant correlations of individual preferences between these tasks for head regions and areas of higher physical saliency. Although it has previously been shown that scene locations that are classified as interesting or salient also receive enhanced attention on the group level (Borji, Sihite, & Itti, 2013; Koehler, Guo, Zhang, & Eckstein, 2014; Masciocchi et al., 2009; Onat et al., 2014), we could additionally demonstrate such relationship on the level of individual observers with respect to preferences for social information. Moreover, the current findings extend previous research because we deliberately did not ask participants to select salient or interesting scene locations but to identify those spots that are deemed most relevant for scene comprehension. Our results indicate that observers who preferentially look at heads or faces of conspecifics also assume that these regions are most relevant for understanding the depicted scene. In principle, these aspects might be causally related such that observers first fixate on regions they find important and subsequently select the same regions when asked to do so. Although the current experimental design does not permit to examine such relationship directly because we did not acquire eye-tracking data in the relevance tap condition, it might be an interesting question for future research to elucidate the process by which observers explicitly select regions of personal relevance. Finally, both social fixations and relevance taps on the heads of depicted individuals were also related to the social content of self-taken photographs, although only the relevance taps task was significantly correlated with the proportion of humans/faces when correcting for multiple comparisons. Thus, participants who selected heads more often in the relevance taps task also took more photographs that included other persons or faces when they freely selected motives outside of the laboratory.

On the group level, the current findings replicate previous work showing that social features receive more attention than physically salient image regions (End & Gamer, 2017). This effect was evident from the first fixation onward but persisted for the whole viewing duration indicating that the preference for social features involves early (End & Gamer, 2019; Flechsenhar et al., 2018a; Fletcher-Watson et al., 2008; Rösler et al., 2017) and more sustained aspects of attentional orienting (Flechsenhar et al., 2018a; Rubo & Gamer, 2018). Consequently, it has been discussed that social attention includes reflexive, stimulus-related as well as goal-driven components (Flechsenhar & Gamer, 2017). Importantly, we discovered a similar pattern not only for the first fixations but also for the first taps during the relevance taps task. Fixations and taps were less well predicted by physical saliency than by social stimulus content (cf. Cerf et al., 2007) although this bias was slightly less pronounced in the relevance taps task. Taken together, these results highlight the complex interplay between stimulus-driven and goal-oriented characteristics of social attention, which is reflected by a similar pattern in, as well as a correspondence between, rather reflexive early fixations and more strategic, deliberate behavioral relevance taps and photos. Extending previous research studying social attention on the group level (Birmingham et al., 2008; Birmingham, Bischof, & Kingstone, 2007; End & Gamer, 2017; End & Gamer, 2019; Flechsenhar et al., 2018a; Rösler et al., 2017; Rubo & Gamer, 2018), our results further demonstrate significant interindividual differences in the prioritization of social information in attention, personal relevance judgments and real-world behavior.

When comparing the pattern of correlations between tasks, it becomes evident that results are less consistent for the first five fixations as compared to all fixations that occurred during the 10 second viewing period. In general, the lower correlations for the first five fixations could be simply due to having less data available which is consistent to the lower reliabilities that were observed for this initial exploration period. Remarkably, reliability was especially reduced for fixations on highly salient image regions (see Figure 3). This observation might result from the fact that most of the initial fixations landed on head and body regions (see Figure 8), which implies that less data were available selectively for high-saliency fixations. In addition, correlations between tasks for the head region were also slightly lower when considering the first five compared to all fixations. This pattern possibly reflects that stable social preferences emerge more clearly during elaborate visual exploration. Although previous research has indicated that individual differences in the prioritization of specific semantic categories are already evident in the first fixation (de Haas et al., 2019), reliability estimates increase monotonically with longer viewing durations (Linka & de Haas, 2020) which is in line with our findings.

Exploring the pattern of correlations further, it seems interesting that our results were much more consistent for head than for body regions. In fact, none of the intercorrelations between tasks for the body ROI reached statistical significance. Within tasks, body regions correlated not at all (all fixations of the free viewing task) or negatively (relevance taps task; see Supplementary Figure S1) with head regions, which indicates that having fixated/selected the head region makes it less likely to revisit the body of the depicted person (or vice versa). This is in line with current

research showing that fixation densities on heads and bodies are negatively correlated (Broda & de Haas, 2022). Social aspects of our surroundings are thus not equally important. Our results indicate that heads seem to be the primary source of information whereas bodies are fixated/selected less often and also less consistently across the stimulus set (for similar results see Birmingham et al., 2008; End & Gamer, 2017). Individual differences in social prioritization thus seem to be particularly driven by heads and faces.

An unresolved question still concerns the origin and consequences of the currently observed individual differences. Previous twin studies have revealed that patterns of visual exploration seem to be heritable (Constantino, Kennon-McGill, Weichselbaum, Marrus, Haider, Glowinski, Gillespie, Klaiman, Klin, & Jones, 2017; Kennedy et al., 2017) and are related to specific clinical constructs, such as autism spectrum disorders (Jones & Klin, 2013; Wang et al., 2015). This might be particularly true for the currently observed social preferences (see review by Chita-Tegmark, 2016). However, consistent with previous research on healthy individuals (Horn et al., 2022; Rösler et al., 2021) we did not observe significant correlations with autism spectrum or social anxiety traits. It seems possible that relationships to clinical constructs increase in clearly pathological conditions (Lazarov, Abend, Bar-Haim, 2016; Wang et al., 2015) or actual instead of depicted social encounters (Hessels, Holleman, Cornelissen, Hooge, & Kemner, 2018; Rubo, Huestegge, & Gamer, 2020). Moreover, future research should also examine to what degree the currently observed differences in social prioritization affect how individuals navigate their social world. It seems interesting whether social attention shapes how individuals interpret social information, how they perceive ambiguous social scenes and how they interact with others in social contexts. Following the cognitive ethology approach (Kingstone, Smilek, & Eastwood, 2008), this research should take into account real social encounters and behavior in social interactions to further explore the consequences of individual differences in social prioritization. The currently used photograph condition can be understood as a first step toward this goal and allowed us to extend laboratory behavior to a natural environment (see also Wang et al., 2016). We assumed that the uninstructed, and individually taken photographs would reflect social preferences on a very personal and unobtrusive level. Relating free viewing and explicitly instructed personal relevance taps to photographs taken outside without constraints was meant to break the often criticized borders of laboratory-based attention research (Birmingham & Kingstone, 2009; Risko, Laidlaw, Freeth, Foulsham, & Kingstone, 2012; Risko et al., 2016) and to identify a preference for social features within a natural setting. Indeed, although our three measures potentially capture different aspects

of social preference, we found correlations among all of them, which provides evidence for a shared social preference trait.

The current study has several strengths, including the broad assessment of social preferences across several laboratory and real-world tasks, but also comes with some limitations. First, although we counterbalanced the order of the first two tasks (free viewing and relevance taps) across participants to avoid a possible order effect, the photograph task was always conducted at the end for practical reasons. Due to the large variety of pictures that were used in the laboratory tasks and the fact that the stimulus set was identical for each observer, we did not expect an influence on how participants would take their photographs, but we cannot completely exclude this possibility. Importantly, however, participants showed a substantial variability regarding photograph content (see Supplementary Figures S2, S3) and the proportion and number of depicted humans and faces (see e.g. Supplementary Figure S5). Executing this task was thus apparently rather unaffected by the stimuli that were viewed in the previous experimental tasks. Second, we used a minimal instruction for the free viewing task in accordance with earlier studies in our laboratory (e.g. End & Gamer, 2017), which Parkhurst, Law, and Niebur (2002) considered close to a natural viewing behavior. However, giving clear task instructions was important for the relevance taps task. In the instructions, we tried to emphasize a general rule (i.e. "select those image regions that are most important for understanding the depicted situation") instead of focusing on specific features. The current findings demonstrate that although different instructions were needed for the free viewing and the relevance taps task, both seemed to capture individual differences in social prioritization. Third, the central fixation bias is a well-known effect in eye-tracking research (Tatler, 2007). We tried to reduce this effect by selecting images that contained social features at different locations of the respective scene and we randomly assigned images to the free viewing and the relevance taps task. Moreover, it seems plausible that a central fixation bias does not only occur in free viewing conditions but also affects how individuals select regions of personal relevance. In the current study, we observed meaningful inter-individual variation among these measures, which seems to be driven by social features. This does not entirely rule out differential central biases across tasks but supports the notion of stable individual social preferences beyond such general attentional biases. The final limitation arises from the currently used stimulus set. Static images in a laboratory setting cannot fully represent the real-world in which other low-level features like motion or high-level features like social interactions influence attentional processes (Rösler et al., 2017). Still, this was the trade-off for a controlled environment, and we

tried to overcome this limitation to a certain extent by including a real-world task. For future studies, it would be interesting to use other, more dynamic media such as self-recorded video sequences or techniques, such as mobile eye tracking in real social encounters. However, although mobile devices proved to be reliable in their accuracy (Ehinger, Groß, Ibs, König, & Negrello, 2019; Großekathöfer et al., 2022), it still seems unclear to what degree the mere fact that someone is taking part in an experiment or wearing eye-tracking glasses influences natural viewing behaviors and social interactions.

Conclusion

In conclusion, we found evidence for a trait-like preference of social information. First, we discovered a stability of gaze patterns during the presentation of complex naturalistic images. Second, we obtained similar effects in a novel task that required participants to actively select image regions that are relevant for understanding social scenes. Finally, social preferences in attention and relevance judgments were correlated and partly predicted social characteristics of self-taken photographs. The current results demonstrate that social attention shares features of stimulus-driven and goal-directed processing and highlights the influence of individual differences on these aspects. Future research should concentrate on the origin and consequences of the currently observed phenomenon and specifically consider new ways to identify such personal preferences and to extend laboratory research to natural environments with the help of new methods and technologies. Last, the implications for clinical conditions associated with deviant social processing, such as autism spectrum and social anxiety disorders should be addressed.

Keywords: social attention, visual system, eye gaze, gaze stability, saliency maps, eye tracking

Acknowledgments

Commercial relationships: none.
Corresponding author: Adam M. Berlijn.
Email: berlijn@uni-duesseldorf.de.
Address: Department of Experimental Psychology, Heinrich-Heine-University Düsseldorf, Universitätsstraße 1, 40225 Düsseldorf, Germany.

References

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

Baron-Cohen, S., Wheelwright, S., Skinner, R., Martin, J., & Clubley, E. (2001). The Autism-Spectrum Quotient (AQ): Evidence from Asperger Syndrome/High-Functioning Autism, Males and Females, Scientists and Mathematicians. *Journal of Autism and Developmental Disorders, 31*(1), 5–17.

Birmingham, E., Bischof, W. F., & Kingstone, A. (2007). Why do we look at people's eyes? *Journal of Eye Movement Research, 1*(1), 1–6.

Birmingham, E., Bischof, W. F., & Kingstone, A. (2008). Gaze selection in complex social scenes. *Visual Cognition, 16*(2-3), 341–355.

Birmingham, E., & Kingstone, A. (2009). Human social attention: A new look at past, present, and future investigations. *Annals of the New York Academy of Sciences, 1156*, 118–140.

Birmingham, E., Ristic, J., & Kingstone, A. (2012). Investigating social attention: A case for increasing stimulus complexity in the laboratory. In J. A. Burack, J. T. Enns, & N. A. Fox (Eds.), *Cognitive neuroscience, development, and psychopathology: Typical and atypical developmental trajectories of attention* (pp. 251–276). Cary, NC: Oxford University Press.

Borji, A., & Itti, L. (2013). State-of-the-art in visual attention modeling. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 35*(1), 185–207.

Borji, A., Sihite, D. N., & Itti, L. (2013). What stands out in a scene? A study of human explicit saliency judgment. *Vision Research, 91*, 62–77.

Broda, M. D., & de Haas, B. (2022). Individual differences in looking at persons in scenes. *Journal of Vision, 22*(12), 9–9.

Castelhano, M. S., Mack, M. L., & Henderson, J. M. (2009). Viewing task influences eye movement control during active scene perception. *Journal of Vision, 9*(3), 6.1–6.15.

Cerf, M., Harel, J., Einhaeuser, W., & Koch, C. (2007). Predicting human gaze using low-level saliency combined with face detection. *Advances in Neural Information Processing Systems, 20*, 1–8.

Chita-Tegmark, M. (2016). Attention Allocation in ASD: a Review and Meta-analysis of Eye-Tracking Studies. *Review Journal of Autism and Developmental Disorders, 3*(3), 209–223.

Constantino, J. N., Kennon-McGill, S., Weichselbaum, C., Marrus, N., Haider, A., & Glowinski, A. L.,

et al. (2017). Infant viewing of social scenes is under genetic control and is atypical in autism. *Nature, 547*(7663), 340–344.

de Haas, B., Iakovidis, A. L., Schwarzkopf, D. S., & Gegenfurtner, K. R. (2019). Individual differences in visual salience vary along semantic dimensions. *Proceedings of the National Academy of Sciences, 24*(116), 11687–11692.

DeAngelus, M., & Pelz, J. B. (2009). Top-down control of eye movements: Yarbus revisited. *Visual Cognition, 17*(6-7), 790–811.

Ehinger, B. V., Groß, K., Ibs, I., König, P., & Negrello, M. (2019). A new comprehensive eye-tracking test battery concurrently evaluating the Pupil Labs glasses and the EyeLink 1000. *PeerJ, 7*, e7086.

End, A., & Gamer, M. (2017). Preferential Processing of Social Features and Their Interplay with Physical Saliency in Complex Naturalistic Scenes. *Frontiers in Psychology, 8*, 418.

End, A., & Gamer, M. (2019). Task instructions can accelerate the early preference for social features in naturalistic scenes. *Royal Society Open Science, 6*(3), 180596.

Flechsenhar, A., & Gamer, M. (2017). Top-down influence on gaze patterns in the presence of social features. *PLoS One, 12*(8), e0183799.

Flechsenhar, A., Larson, O., End, A., & Gamer, M. (2018). Investigating overt and covert shifts of attention within social naturalistic scenes. *Journal of Vision, 18*(12), 11.

Flechsenhar, A., Rösler, L., & Gamer, M. (2018). Attentional Selection of Social Features Persists Despite Restricted Bottom-Up Information and Affects Temporal Viewing Dynamics. *Scientific Reports, 8*(1), 12555.

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.

Fletcher-Watson, S., Leekam, S. R., Benson, V., Frank, M. C., & Findlay, J. M. (2009). Eye-movements reveal attention to social information in autism spectrum disorder. *Neuropsychologia, 47*(1), 248–257.

Foulsham, T., Walker, E., & Kingstone, A. (2011). The where, what and when of gaze allocation in the lab and the natural environment. *Vision Research, 51*(17), 1920–1931.

Freeth, M., Foulsham, T., & Kingstone, A. (2013). What affects social attention? Social presence, eye contact and autistic traits. *PLoS One, 8*(1), e53286.

Freitag, C. M., Retz-Junginger, P., Retz, W., Seitz, C., Palmason, H., & Meyer, J., et al. (2007).

Evaluation der deutschen Version des Autismus-Spektrum-Quotienten (AQ) - die Kurzversion AQ-k. *Zeitschrift Für Klinische Psychologie Und Psychotherapie, 36*(4), 280–289.

Großekathöfer, J. D., Seis, C., & Gamer, M. (2022). Reality in a sphere: A direct comparison of social attention in the laboratory and the real world. *Behavior Research Methods, 54*(5), 2286–2301.

Guy, N., Azulay, H., Kardosh, R., Weiss, Y., Hassin, R. R., & Israel, S., et al. (2019). A novel perceptual trait: Gaze predilection for faces during visual exploration. *Scientific Reports, 9*(1), 10714.

Harel, J., Koch, C., & Perona, P. (2007). Graph-Based Visual Saliency. In B. Schölkopf, J. C. Platt, & T. Hofmann (Eds.), *Advances in Neural Information Processing Systems 19* (pp. 545–552). Cambridge, MA: MIT Press.

Hessels, R. S., Holleman, G. A., Cornelissen, T. H. W., Hooge, I. T. C., & Kemner, C. (2018). Eye contact takes two – autistic and social anxiety traits predict gaze behavior in dyadic interaction. *Journal of Experimental Psychopathology, 9*(2), jep.062917.

Horn, A., Mergenthaler, L., & Gamer, M. (2022). Situational and personality determinants of social attention in a waiting room scenario. *Visual Cognition, 30*(1-2), 86–99.

Jones, W., & Klin, A. (2013). Attention to eyes is present but in decline in 2–6-month-old infants later diagnosed with autism. *Nature, 504*(7480), 427–431.

Judd, T., Durand, F., & Torralba, A. (2012). *A Benchmark of Computational Models of Saliency to Predict Human Fixations*. Technical Report. Cambridge, MA: Massachusetts Institute of Technology.

Kennedy, D. P., D'Onofrio, B. M., Quinn, P. D., Bölte, S., Lichtenstein, P., & Falck-Ytter, T. (2017). Genetic Influence on Eye Movements to Complex Scenes at Short Timescales. *Current Biology, 27*(22), 3554–3560.e3.

Kingstone, A., Smilek, D., & Eastwood, J. D. (2008). Cognitive Ethology: A new approach for studying human cognition. *British Journal of Psychology, 99*(3), 317–340.

Koehler, K., Guo, F., Zhang, S., & Eckstein, M. P. (2014). What do saliency models predict? *Journal of Vision, 14*(3), 14.

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 of the United States of America, 108*(14), 5548–5553.

Lang, P. J., Bradley, M. M., & Cuthbert, B. N. (2008). *Technical report A-8*. Gainesville, FL: University of Florida.

Lazarov, A., Abend, R., & Bar-Haim, Y. (2016). Social anxiety is related to increased dwell time on socially threatening faces. *Journal of Affective Disorders, 193*, 282–288.

Linka, M., & de Haas, B. (2020). Osieshort: A small stimulus set can reliably estimate individual differences in semantic salience. *Journal of Vision, 20*(9), 13.

Marchewka, A., Żurawski, Ł., Jednoróg, K., & Grabowska, A. (2014). The Nencki Affective Picture System (NAPS): Introduction to a novel, standardized, wide-range, high-quality, realistic picture database. *Behavior Research Methods, 46*(2), 596–610.

Masciocchi, C. M., Mihalas, S., Parkhurst, D., & Niebur, E. (2009). Everyone knows what is interesting: Salient locations which should be fixated. *Journal of Vision, 9*(11), 25.1–25.22.

Mattick, R. P., & Clarke, J. (1998). Development and validation of measures of social phobia scrutiny fear and social interaction anxiety. *Behaviour Research and Therapy, 36*(4), 455–470.

Mehoudar, E., Arizpe, J., Baker, C. I., & Yovel, G. (2014). Faces in the eye of the beholder: Unique and stable eye scanning patterns of individual observers. *Journal of Vision, 14*(7), 6.

Olmos, A., & Kingdom, F. A. A. (2004). A Biologically Inspired Algorithm for the Recovery of Shading and Reflectance Images. *Perception, 33*(12), 1463–1473.

Onat, S., Açık, A., Schumann, F., & König, P. (2014). The contributions of image content and behavioral relevancy to overt attention. *PLoS One, 9*(4), e93254.

Parkhurst, D., Law, K., & Niebur, E. (2002). Modeling the role of salience in the allocation of overt visual attention. *Vision Research, 42*(1), 107–123.

R Core Team. (2021). *A language and environment for statistical computing.* 2015. Vienna, Austria: R foundation for statistical computation, https://www.R-project.org/.

Risko, E., Laidlaw, K., 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*, 143.

Risko, E., Richardson, D. C., & Kingstone, A. (2016). Breaking the Fourth Wall of Cognitive Science. *Current Directions in Psychological Science, 25*(1), 70–74.

Rösler, L., End, A., & Gamer, M. (2017). Orienting towards social features in naturalistic scenes is reflexive. *PLoS One, 12*(7), e0182037.

Rösler, L., Göhring, S., Strunz, M., & Gamer, M. (2021). Social anxiety is associated with heart rate but not gaze behavior in a real social interaction. *Journal of Behavior Therapy and Experimental Psychiatry, 70*(101600), 1–9.

RStudio Team. (2020). *RStudio: Integrated Development for R.* Boston, MA: RStudio, PBC, http://www.rstudio.com/.

Rubo, M., & Gamer, M. (2018). Social content and emotional valence modulate gaze fixations in dynamic scenes. *Scientific Reports, 8*(1), 3804.

Rubo, M., Huestegge, L., & Gamer, M. (2020). Social anxiety modulates visual exploration in real life – but not in the laboratory. *British Journal of Psychology, 111*(2), 233–245.

Salley, B., & Colombo, J. (2016). Conceptualizing Social Attention in Developmental Research. *Social Development (Oxford, England), 25*(4), 687–703.

Schilbach, L., Timmermans, B., Reddy, V., Costall, A., Bente, G., & Schlicht, T., et al. (2013). Toward a second-person neuroscience. *Behavioral and Brain Sciences, 36*(4), 393–414.

Schneier, F. R., Rodebaugh, T. L., Blanco, C., Lewin, H., & Liebowitz, M. R. (2011). Fear and avoidance of eye contact in social anxiety disorder. *Comprehensive Psychiatry, 52*(1), 81–87.

Smith, T. J., & Mital, P. K. (2013). Attentional synchrony and the influence of viewing task on gaze behavior in static and dynamic scenes. *Journal of Vision, 13*(8), 16.

Speer, L. L., Cook, A. E., McMahon, W. M., & Clark, E. (2007). Face processing in children with autism: Effects of stimulus contents and type. *Autism, 11*(3), 265–277.

Stangier, U., Heidenreich, T., Berardi, A., Golbs, U., & Hoyer, J. (1999). Die Erfassung sozialer Phobie durch Social Interaction Anxiety Scale (SIAS) und die Social Phobia Scale (SPS). [Assessment of social phobia by the Social Interaction Anxiety Scale (SIAS) and the Social Phobia Scale (SPS).]. *Zeitschrift Für Klinische Psychologie, 28*(1), 28–36.

Tatler, B. W. (2007). The central fixation bias in scene viewing: Selecting an optimal viewing position independently of motor biases and image feature distributions. *Journal of Vision, 7*(14), 4.

Wang, S., Fan, S., Chen, B., Hakimi, S., Paul, L. K., & Zhao, Q., et al. (2016). Revealing the world of autism through the lens of a camera. *Current Biology, 26*(20), R909–R910.

Wang, S., Jiang, M., Duchesne, X. M., Laugeson, E. A., Kennedy, D. P., & Adolphs, R., et al. (2015). Atypical Visual Saliency in Autism Spectrum

Disorder Quantified through Model-Based Eye Tracking. *Neuron, 88*(3), 604–616.

Wessa, M., Kanske, P., Neumeister, P., Bode, K., Heissler, J., & Schönfelder, S. (2010). EmoPics: subjektive und psychophysiologische Evaluation emotionalen Bildmaterials zur klinischen und biopsychologischen Forschung. *Z. Klin. Psychol. Psychother. Suppl., 39*, 77.

Wieser, M. J., Pauli, P., Alpers, G. W., & Mühlberger, A. (2009). Is eye to eye contact really threatening and avoided in social anxiety? An eye-tracking and psychophysiology study. *Journal of Anxiety Disorders, 23*(1), 93–103.

Xu, J., Jiang, M., Wang, S., Kankanhalli, M. S., & Zhao, Q. (2014). Predicting human gaze beyond pixels. *Journal of Vision, 14*(1), 28.