Vision Research 216 (2024) 108356

Contents lists available at ScienceDirect

Vision Research

journal homepage: www.elsevier.com/locate/visres

Don't look now! Social elements are harder to avoid during scene viewing

A.P. Martinez-Cedillo [a,b,*], T. Foulsham [b]

[a] Department of Psychology, University of York, York YO10 5DD, England
[b] Department of Psychology, University of Essex, Wivenhoe Park, Colchester, Essex CO4 3SQ, England

A B S T R A C T

Regions of social importance (i.e., other people) attract attention in real world scenes, but it is unclear how automatic this bias is and how it might interact with other guidance factors. To investigate this, we recorded eye movements while participants were explicitly instructed to avoid looking at one of two objects in a scene (either a person or a non-social object). The results showed that, while participants could follow these instructions, they still made errors (especially on the first saccade). Crucially, there were about twice as many erroneous looks towards the person than there were towards the other object. This indicates that it is hard to suppress the prioritization of social information during scene viewing, with implications for how quickly and automatically this information is perceived and attended to.

1. Introduction

Other humans are among the most potent attractors of visual attention. This is the conclusion from attention experiments using arrays of controlled stimuli (e.g., Hershler & Hochstein, 2005; Golan et al., 2014), but also from a variety of studies investigating eye movements in pictures and real-world scenes. For example, early work by Buswell (1935) and Yarbus (1967) reported that observer fixations were often clustered around faces depicted in paintings and photographs, although the scan patterns on display also changed according to the task that was being performed. When participants are simply asked to look at an image, any people in the scene, and particularly their faces and eyes, are one of the first and most frequent regions to be fixated (Bindemann et al., 2010; Birmingham et al., 2008). Faces and eyes also dominate attention during video watching (Foulsham et al., 2010), although in real, face-to-face interactions the drive to look at other people is reduced according to social signalling and social norms (Laidlaw et al., 2011; Risko et al., 2012).

The priority accorded to images of other people is consistent with specialised, rapid processing of "socially relevant" features in our environment. A large body of research suggests that the perception of faces and bodies is associated with specialised regions in the brain (Downing, Peelen, Wiggett & Tew, 2006; Tsao & Livingstone, 2008), and that socially important signals such as gaze direction are also prioritized (Haxby, Hoffman & Gobbini. 2000). Humans and animals can also be detected in pictures extremely rapidly, perhaps with fast "feed-forward" processing or subcortical routes within the visual system. For example, Crouzet et al., (2010) showed that when an observer was asked to saccade to one of two pictures containing a face, they were able to do so in only around 100 ms. This "ultra-rapid" detection was also seen for other categories (animals, vehicles) but was faster and more accurate for human faces.

To what degree is attention towards other people in scenes "automatic"? The first saccade during viewing is often directed towards people in images (Fletcher-Watson et al., 2008), and this happens even if the scene disappears after 200 ms (Rösler, End & Gamer, 2017), which suggests a reflexive behaviour. One different way of addressing automaticity is to examine what happens when participants are given a secondary task which may conflict with any default viewing preferences. Flechsenhar and Gamer (2017) gave participants different tasks involving counting objects or assessing parts of an image. Their results showed that while the eye movements elicited changed with the instructions, there was still a bias to look at the people in the scenes even when not relevant for the task. End and Gamer (2019) asked participants to freely view naturalistic scenes, and in a second condition to specifically try to direct fixations to the socially relevant areas. They found more fixations to heads and bodies in both task instructions, compared to the other areas, with relatively little effect of instructions.

In two recent studies, we have investigated whether the social prioritisation effect is reduced in the presence of varying levels of working memory load. Our findings indicated that the social prioritisation effect persisted in the presence of high memory loads of verbal content (i.e., trying to remember a number), thus highlighting the apparently stubborn preference for looking at people (Martinez-Cedillo, Dent & Foulsham, 2022). We did find a reduction in looks to people when memorising high memory loads of visuospatial content (i.e., trying to

* Corresponding author at: Department of Psychology, University of Essex, Wivenhoe Park, Colchester, CO4 3SQ, England.
E-mail address: a.p.martinezcedillo@essex.ac.uk (A.P. Martinez-Cedillo).

https://doi.org/10.1016/j.visres.2023.108356
Received 27 June 2023; Received in revised form 9 November 2023; Accepted 28 December 2023
Available online 6 January 2024

remember a pattern of dots), thus suggesting some limitations on this prioritisation (Martinez-Cedillo, Dent, & Foulsham, 2023). However, even in that case, socially-relevant regions of interest were more likely to be fixated than non-social objects.

Perhaps the strongest test of volitional vs. automatic attention to social features is to design a task where participants are expressly instructed to avoid attending to these features. This is what Laidlaw and colleagues describe as the 'don't look' task (Laidlaw et al., 2012; 2017). In two experiments, Laidlaw et al. (2012) asked participants to avoid looking at either the eyes or the mouth when viewing photographs of single faces. The results showed that participants found it more difficult to avoid fixating the eyes in the photographs than to avoid fixating the mouth (i.e., they made more frequent errors where they looked to the eyes when they should not have done). Interestingly, this pattern was eliminated when the faces were inverted. This suggests that there is at least some automatic component to fixating the eyes in an image, and that this is associated with the holistic meaning of the face (which is removed when inverted). Thompson et al., (2019) also found that fixations on the eyes were hard to avoid, and that they occurred even when participants were prompted to look at other features in a recognition paradigm. A similar inference can be drawn from studies using face stimuli in an antisaccade task (where participants must make a saccade in the opposite direction of a simple target). For example, Morand et al., (2010) reported that participants made more antisaccade errors when the cue was a face.

Despite this evidence, it is not yet known how such instructions might affect eye guidance during complex scene viewing. In a complex scene, unlike a display with a single face, social information is embedded in a context of other objects and background. People will not always be in the same locations, and while their presence may be determined rapidly (Crouzet et al., 2010), it is not clear whether eye movements towards them are fully under cognitive control or whether they are, at least some of the time, obligatory. In the present study we ask whether participants will be able to avoid looking at people in a complex image.

Research on eye guidance in scene viewing has suggested that contrast in low-level features such as colour, orientation and luminance directly impacts attention and where people look (i.e., saliency; Itti & Koch, 2000). These "bottom-up" targets of attention may be particularly predictive in the absence of a strong task ("free viewing") and, perhaps, specifically in the first eye movements made after the appearance of a scene (Anderson et al., 2015; Anderson & Donk, 2017; Foulsham & Underwood, 2008; Underwood & Foulsham, 2006). However, a range of studies show that, in fact, eye movements are better predicted by the meaning or task-relevance of scene regions than by feature contrast alone (Henderson & Hayes, 2017; 2018; Foulsham & Underwood, 2007; Tatler et al., 2011; Nuthmann & Einhauser, 2015). According to a "cognitive relevance" account (Henderson, 2020), participants must begin to recognise the meaning of possible saccade locations, and then prioritise these according to the current task. In the case where participants must expressly avoid scene regions, this would suggest very few erroneous fixations should be made.

When saliency, in terms of bottom-up feature contrast, and social relevance have been pitted against each other, the results have been quite clear. Social information seems to attract attention regardless of the simple conspicuity of the region. Birmingham et al., (2009) reached this conclusion from analysis of the model-predicted saliency of fixated locations in social scenes, with many highly fixated social regions being of low saliency. Nystrom and Holmqvist (2008) digitally altered the contrast of scene regions and found that social regions, in particular, were still fixated even when their saliency had been reduced. Such results complement the findings described above that social information is selected by eye movements even in the presence of other, salient objects (e.g., Flechsenar & Gamer, 2017). As a result, some saliency map models, which aim to predict where people look in scenes, have added a face detection algorithm, showing improved performance (Cerf, Harel, Einhäuser, Koch, 2007). An alternative is for data-driven models to learn to classify fixated locations (Kummerer, Wallis, Gatys, & Bethge, 2017). These models show good predictive power, and may involve "high-level" representations which correspond to the faces and people often selected.

1.1. Present research

The aim of this paper was to investigate whether the fixations made to people in images are elicited automatically, or whether they are under volitional control. We used a similar 'don't look' paradigm to Laidlaw et al., (2012) but which has previously not been applied to complex scenes. Participants were either told to avoid looking at people in the scene, or to avoid looking at the main non-social object in the scene. These conditions were compared to a free viewing condition. Given the results reviewed above, we expected that in the absence of instructions the people in the images would receive many fixations (and more so than non-social objects). We also manipulated the bottom-up saliency of the non-social object in the image, by editing it so that it did or did not stand out from its background. Although many studies have shown that saliency is predictive of fixation during free viewing (Itti & Koch, 2000; Parkhurst et al., 2002), these studies are mostly correlational in nature which means that it may be other aspects of the scene (such as objects or semantic meaning), rather than saliency per se, which causes fixation. Manipulating the saliency of an object should control the content and meaning of that region, and thus provides a different test of whether visual factors alone will lead to increased attention. Previous studies indicate that objects will be fixated more often when they are manipulated to be higher in saliency (Foulsham & Underwood, 2007; Martinez-Cedillo, Dent & Foulsham, 2022). Nonetheless, it is not clear that such an effect will emerge when social information is also present, since several studies have shown that people in the scene attract attention regardless of saliency (Birmingham et al., 2009; Flechsenhar & Gamer, 2017).

The results of the "don't look" tasks will shed light on whether guidance to people and non-social objects is automatic and obligatory or volitional. We hypothesised that if the bias to look at the social information is a consequence of an automatic response, indicative of stronger attentional priority to people than to other objects, then performance in the "don't look social" condition should be worse (i.e., more errors should be made) than performance in "don't look object" condition. If errors are rare, and are not made more often in the "don't look social" condition, then it would indicate that participants have a high level of volitional control over where they look. It may be that such top-down control will take longer to develop, and so we also look specifically at the first eye movement in the scene. If saliency is an important factor in initial guidance to scene regions, we would expect more frequent early fixations to the non-social object when it is higher saliency. Moreover, it might be more difficult to avoid looking at a highly salient object (in the "don't look object" condition).

2. Methods

2.1. Participants

Thirty students from the University of Essex participated (ages 18 – 25, M = 19.53 years, 21 females, 9 males). Although we pre-registered our sample size, we deviated from our plan and recruited fewer participants than planned due to unforeseen difficulties. Our sample size is similar to previous research (Laidlaw et al., 2012; End & Gamer, 2019; Martinez-Cedillo et al., 2022, Experiment 1) and we carry out a sensitivity power analysis (Lakens, 2022; see Results Section 3.1) which demonstrates our ability to detect even modest effects with this sample size and design. The experiment preregistration, data, experimental and analysis code are available via the Open Science Framework at: **https://osf.io/4aewz/**.

All participants reported normal or corrected-to-normal vision. Participants were paid £4 or 1 course credit for their involvement. The ethics board of the University of Essex approved the study. Participants received verbal and written instructions regarding the experimental procedures and gave their informed consent.

2.2. Apparatus and stimuli

The experiment was programmed in MATLAB (version 9.1.0, R2016b; the Mathworks, Natick, MA), using the Psychophysics Toolbox. Eye position was recorded using the SMI RED500, a screen-based eye tracker that samples pupil position at 500 Hz. A 9-point calibration and validation were repeated several times to ensure all recordings had a mean spatial error of better than 0.8 degrees. Head movements were restricted using a chin rest. The experiment took place in a dimly illuminated, sound-attenuated room. Participants sat 60 cm from the monitor, so that the stimuli subtended approximately 43 deg by 28 deg of visual angle at 1680 x 1050 pixels.

The stimuli were 27 colour images depicting indoor and outdoor scenes. These images were taken from the set described previously (Martinez-Cedillo et al., 2022). Each image had two key regions, alongside other objects and background: a social region of interest (a person) and a non-social region of interest (an inanimate object chosen to be similar in size to the social region). The non-social objects varied between trials and included objects such as a pot plant, a lamp and a guitar. The two objects were positioned near to the centre of the left and right half of the image, with similar eccentricity (mean distance from centre: social = 11.3 degrees of visual angle, non-social = 11.1 degrees). These elements were positioned on opposite sides of the image. The photographs were originally sourced from online collections (e.g., Pixabay) and then modified using image editing software (Picmonkey: https://www.picmonkey.com/). Specifically, the non-social object was edited with the aim of either increasing or decreasing the bottom-up saliency of the region. This was done by, for example, changing the colour or brightness of this object to change the contrast between the background and the object. To monitor whether these manipulations had the required effect, we used a widely used model of bottom-up saliency (Itti & Koch, 2000) implemented via the Saliency Toolbox in MATLAB (version 9.1.0, R2016B; *the MathWorks, Natick, MA)* which gives a series of predicted fixation locations according to the feature contrast (brightness, colour and orientation). In the high salience condition, the non-social object stood out from its background and was selected in one of the first three predicted fixations. In the low salience condition, the same object did not stand out and was not selected by the model until much later (and not in the first 5 fixations). In our analysis, we first divide the stimuli into high and low salience conditions, treating saliency as a categorical variable. However, in additional analysis we also examine how these conditions differ according to an alternative saliency model, and use saliency as a continuous predictor for behaviour (see Section 3.5).

We attempted to make the saliency changes realistic, although it was not necessary for our design that the modifications were unnoticed. We also mirror reversed each image, to control for whether the person or the object appeared on the left/right of the scene. This resulted in 4 different versions of each scene (high/low salience x original/reversed), for a total of 108 images which were used equally across the study.

2.3. Procedure

Fig. 1 illustrates the procedure underlying the sequence of the task. At the beginning of each session, the eye tracker was calibrated and validated. The experiment consisted of three blocks, corresponding to three different task instructions. Each trial started with a fixation cross displayed for 500 ms, followed by a reminder of the instructions for 500 ms. At the beginning of the session, participants were told that each image would contain at least two objects: a social element (a person) and an object, which would be presented on the opposite side of each other. The instructions were presented on the screen at the beginning of each block. In the "don't look social" condition, participants were instructed to avoid looking at the person in the image. In the "don't look object" condition, participants were instructed to avoid looking at the main inanimate object in the scene. In the remaining, third condition, participants were told to look at the image however they wished ("free viewing"). Each image was presented for 5 s, which was consistent across all the tasks and chosen to allow enough time to explore the image during free viewing.

The three task conditions were blocked and presented in random order between participants. Within each block, 36 scenes were presented in a random order. We created counterbalanced lists to balance the different image versions across task conditions, and each participant was assigned to one of these lists. Images in each task condition were divided equally between high and low salience and original and flipped versions, and across lists the images in each block were swapped. This meant that, across the whole study, each particular image was seen in

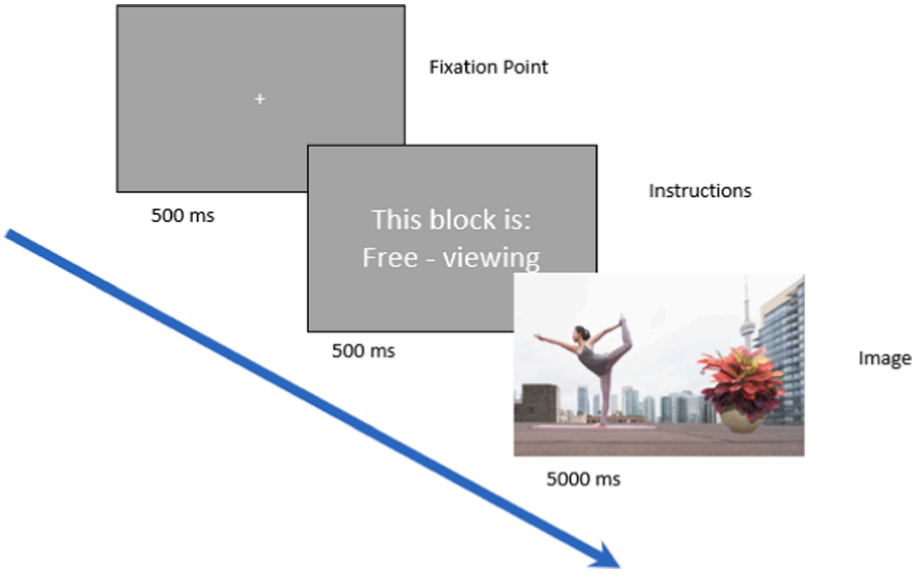

Fig. 1. Experimental procedure showing the sequence of the task, in this case in the free-viewing condition.

each task, each salience condition and in the original and flipped version. The total duration of the experiment was approximately 20 min.

2.4. Data analysis

Fixations were removed if their duration was below 100 ms (0.15 % of all fixations). We also excluded trials where the starting fixation was not recorded on the centre (using a region of interest of 20 pixels around the fixation point; 2.81 % trials were excluded). Following this step, we removed three participants who, due to poor calibration and missing eye tracking data, had fewer than 40 % of their trials remaining, and one additional participant who it transpired had not understood the instructions. We present analysis of the remaining 26 participants. (ages 18–38; $M = 21.76$, $SD = 4.72$; 22 females,4 males).

Fixations (and saliency model predictions, see above) were evaluated according to rectangular interest areas which were drawn around each object. Our analysis focused on the relative frequency of fixations to each of the two regions of interest (social and non-social) according to the task condition. Rather than aggregating the proportion of fixations, we used a generalised linear mixed model (GLMM) approach which allowed us to model outcomes at the level of each fixation (approximately 12,000 data points). Models predicted the binary response and thus in which circumstances participants would fixate on that area or not. We included random effects of the participant and the scene, using the lme4 package in R (Bates et al., 2009) and a binomial function. The contribution of each factor was evaluated with maximum likelihood comparisons. In each case, we first added the fixed effect of task (free viewing, "don't look social" or "don't look object"), with treatment coding comparing each task to the free viewing condition. We then added a fixed effect of the salience factor (high or low) and the interaction between salience and task and compared whether this model produced a better fit.

3. Results

3.1. Sensitivity power analysis

It is not always straightforward to establish what effects can be detected in (G)LMMs. This is particularly the case when random effects of participant and stimulus are included, since in this situation the statistical power depends both on the number of participants and on the number of stimuli (Brysbaert & Stevens, 2018). In the present case, we used simulation in R (package simr: Green & MacLeod, 2016), post hoc, to perform a sensitivity power analysis (Lakens, 2022). This examines the range of effect sizes that can be detected with good statistical power, given a particular design and sample size. Reproducible code is included in our OSF online repository.

The results showed that with our design we had good statistical power (>80 %) for detecting fixed effects of task with an unstandardised β of 0.5 (equivalent to about a 10 % difference in fixation proportions between tasks), and excellent power for larger effects. The contrast between high and low saliency was a more powerful situation and could be reliably detected with smaller β values (>80 % power at 0.15 and higher). The simulations also show good sensitivity to interactions between task and saliency (>80 % power at β values higher than 0.32 or about a 5 % difference in saliency effects in the different tasks).

3.2. Fixations on the two regions of interest

Fig. 2 shows the mean proportion of fixations in each task condition on the social region of interest (the person) and on the other, non-social object. These results confirm that participants were following the task instructions: looking less at the social region when instructed not to do so, and looking less at the non-social object in the "don't look object" condition. The "default" preference in the free viewing condition was to fixate the social region much more often than the non-social region, but this preference was clearly under voluntary control as it changed completely when participants were asked to avoid one of the regions. Nonetheless, fixations on the to-be-avoided region were not completely eliminated. It is also interesting to note that the relative increase in looks to the "permitted" region, as compared to the free viewing baseline, was greater in the "don't look object" condition.

We confirmed this pattern with two separate GLMMs. The first examined the fixed effect of task on fixations to the social region (with random effects of participant and image). Task was a significant predictor (compared to intercept-only: $\chi^2(2) = 119.9$, $p < .001$) and the probability of fixating on the person decreased in "don't look social" ($\beta = -0.744 \pm 0.179$ SE, $p < .001$) but increased with "don't look object" instructions ($\beta = 1.071 \pm 0.117$ SE, $p < .001$). Adding the salience of the non-social object, or its interaction with task, did not improve the model (χ^2 (3) = 2.8, $p = .41$).

The second GLMM examined effects on the non-social object. Again,

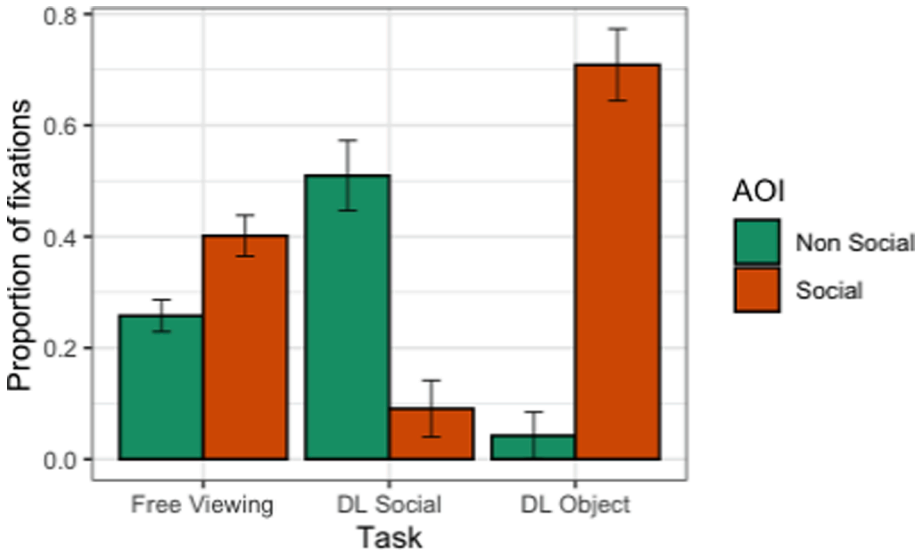

Fig. 2. Mean proportion of all fixations on the two ROIs (social and non-social) as a function of instructions (DL Social ="don't look social", DL Object = "don't look object" and free viewing). Graph shows the mean across participants, with error bars showing 95% confidence intervals, corrected for within-subjects variation using the Cousineau-Morey method (Morey, 2008).

task was a significant predictor (compared to intercept-only: χ^2 (2) = 328.0, p <.001). As expected, the probability of fixating on the non-social element increased in the "don't look social" condition (β = 0.77 ± 0.15 SE, p <.001) and decreased in the "don't look object" instructions (β = -2.18 ± 0.14 SE, p <.001). In this case, adding salience improved model fit and the best fitting model included the fixed effect of salience and also the interaction between condition and salience (χ^2 (3) = 19.3, p <.001). Follow up tests indicated that in the free viewing condition, the non-social object was fixated more often when it was highly salient (M proportion of fixations = 0.285, Cousineau-Morey corrected 95CIs = [0.256, 0.314]) than when it was not (M = 0.250, 95CIs = [0.221, 0.278], β = 0.26 ± 0.07 SE, p <.001). However, this difference was not observed in the "don't look" conditions. In the "don't look social" condition, the high salience object was fixated slightly less often (M = 0.475, 95CIs = [0.412, 0.538]) than the low salience object (M = 0.511, 95CIs = [0.448, 0.574], β = -0.133 ± 0.08 SE, p =.11). In the "don't look object" condition, there was no difference (high salience: M = 0.041, 95CIs = [-0.001, 0.083]; low salience: M = 0.042, 95CIs = [0.000, 0.084]; β = 0.08 ± 0.17 SE, p = 0.63).

Due to the way our images were counterbalanced, participants sometimes saw different versions of the same image later in the experiment. It is possible that their memory for the layout and contents of the image would help in following the instructions when images were repeated. We therefore repeated the analyses with an additional fixed effect of repetition. Repetition or its interaction with condition was not a significant predictor of fixations on the non-social object (χ^2 (3) = 3.3, p =.34). For fixations on the social object, there were numerically more fixations in the "don't look social" condition the first time an image was encountered (M = 0.099, 95CIs = [0.040, 0.157]) than when it was repeated (M = 0.068, 95CIs = [0.012, 0.124]). However, overall the repetition factor was again not a significant predictor (χ^2 (3) = 6.29, p =.098).

3.3. Errors in not looking

We predicted that if looking at people is a consequence of an automatic process, and this process is stronger than looking at other objects, then participants in the "don't look social" condition would perform worse (i.e., make more errors by looking at the social area) than in the "don't look object" condition. We therefore performed a separate analysis of "errors", directly comparing (1) fixations to the social region when participants were told to not look at the social area and (2) fixations to the non-social region when participants were told to not look at the object. Fig. 3 (a) summarises the data on errors in each task. It is clear that there is some variation between participants, with two participants showing a surprisingly large number of errors in the "don't look social" task. We modelled the proportion of errors across all fixations, starting with a fixed effect of task and random effects of participant and image.

This model outperformed a null model with a fixed intercept (χ^2 (1) = 8.67, p =.003). There were about twice as many errors in the "don't look social" condition than in the "don't look object" condition (β = -0.971 ± 0.335 SE, p =.004). Effects of visual salience (and the interaction with task) fell short of statistical significance (χ^2 (2) = 4.97, p =.08). Importantly, there was no difference in the mean proportion of "don't look object" errors in the high and low salience conditions (which were 4 % in each case, see means above).

If errors in the "don't look" tasks are symptomatic of a general problem with controlling attention or impulsivity, then we might expect people with frequent errors in one condition to also make frequent errors in the other condition. We correlated the average proportion of error fixations and found a positive correlation, but one which was weak and non-significant (r(24) = 0.30, p = 0.13). However, since this analysis relies on a single data point per participant per condition, it is likely underpowered and should be treated with caution.

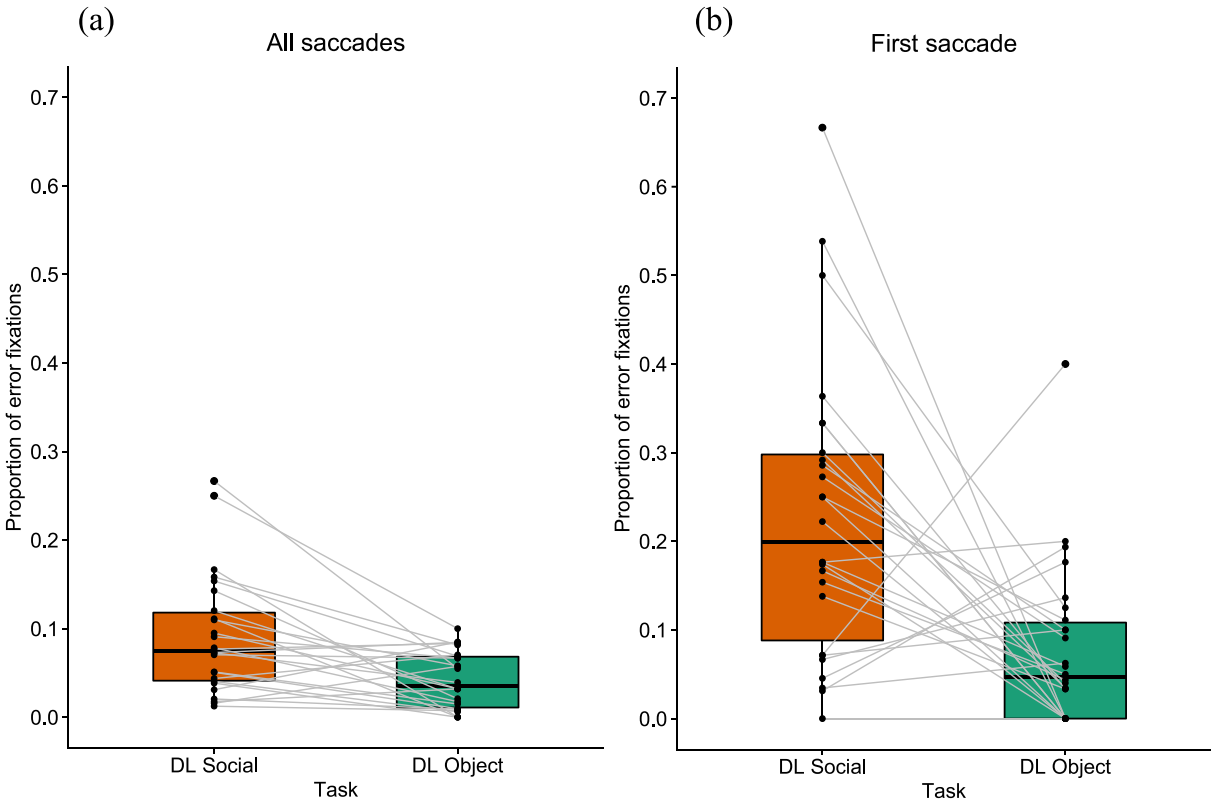

Fig. 3. Proportion of error fixations, where participants looked at the region that they were supposed to be avoiding, across the whole trial (left) and on the first saccade only (right). Boxplots summarise performance across participants, with means for each participant shown as separate datapoints.

3.4. First saccade

We also looked separately at the target of the first free saccade (i.e., the location of the second fixation, since the first fixation was necessarily in the centre of the screen). Performance in this earliest part of the trial might indicate more automatic processing and participants have only peripheral processing on the first fixation to make a decision about which side of the display to fixate (and which to avoid).

The data from first saccades showed a similar pattern to that from the trial as a whole, with a clear effect of instructions – fewer fixations on the person in "don't look social" and fewer fixations on the object in "don't look object". However, the differences in this analysis were not as pronounced and there were relatively more errors compared to the trial overall. A GLMM predicted the proportion of first saccades to the social region by the fixed effect of task (and random effects of participant and image). This model was marginally significant (χ^2 (2) = 5.88, p =.053) and there were fewer fixations on the social object in "don't look social" (M = 0.227, 95CIs = [0.156, 0.299]) than in free viewing (M = 0.406, 95CIs = [0.341, 0.470]; β = -0.906 ± 0.399 SE, p =.023). There were more fixations on the social region in "don't look object" (M = 0.594, 95CIs = [0.541, 0.646]; β = 0.125 ± 0.338 SE, p = 0.712). Salience did not contribute significantly (χ^2 (3) = 2.07, p =.558).

The same analysis on the proportion of first saccades to the non-social region was significant (χ^2 (2) = 16.62, p <.001). First saccades were less common to the non-social object during free viewing (M = 0.118, 95CIs = [0.057, 0.180]) and this increased in the "don't look social" task (M = 0.386, 95CIs = [0.307, 0.466]; β = 1.31 ± 0.329 SE, p <.001) and decreased in the "don't look object" task (M = 0.077, 95CIs = [0.035, 0.118]; β = -0.451 ± 0.327 SE, p = 0.167). Again, the salience of the non-social object was not a significant factor (χ^2 (3) = 5.14, p =.162).

Comparing errors on the first saccade directly, we found a reliable effect of task (χ^2 (1) = 5.64, p =.017), with many more errors in the "don't look social" condition than in the "don't look object" condition (β

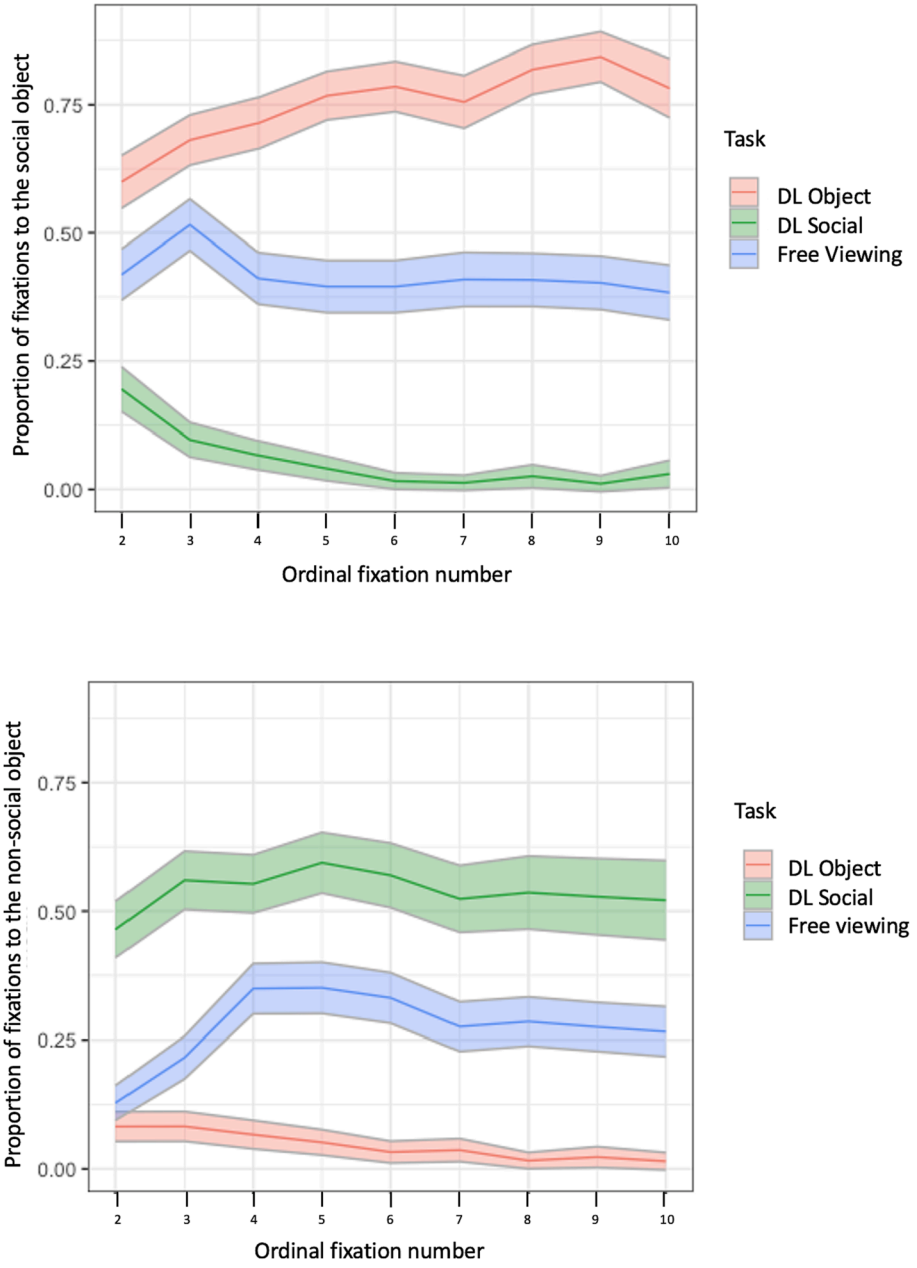

Fig. 4. Mean proportion of all fixations on the two ROIs binned by the fixation number. The shaded region indicates the 95% confidence interval. Note that the fixation count begins from the second fixation, as the initial one is directed towards the centre of the screen.

= -1.04 ± 0.436 SE, p =.017; see Fig. 3b). In both tasks, there were about twice as many errors on the first saccade compared to the trial as a whole.

Fig. 4 shows the time course of fixations to the two regions of interest across the first 10 fixations. This plot confirms that the effects of condition are seen on the first free saccade and maintained over the course of the trial. Errors in the "don't look social" condition are highest at the start of the trial and decrease over time.

3.5. Post hoc saliency analyses

The images in this experiment were manipulated so that the non-social object could be either high or low in saliency. Such a manipulation is important because it ensures that the region has the same semantic meaning (it is the same object) while being relatively higher or lower in saliency. The relative saliency was confirmed by looking at the predictions of a classic saliency map model (Itti & Koch, 2000), which predicts the order of fixations and therefore provides a good way to evaluate the priority of different regions (see Foulsham & Underwood, 2007). We also carried out several post hoc analyses using the model-predicted saliency of our regions of interest (for a review of this approach, see Foulsham, 2019; Borji and Itti, 2012).

First, we calculated the mean saliency values for each region, using both the original Itti and Koch model (implemented in the Saliency Toolbox) and the Graph-Based Visual Saliency model (GBVS; Harel et al., 2007). GBVS is also a bottom-up model which combines measures of feature contrast and does not require training, and it continues to perform well at fixation prediction and be used by perception researchers (Flechsenhar & Gamer, 2017; Kiat et al., 2022). In both cases, saliency maps were normalised to a fixed range of [0,1]. Confirming our manipulation, the object had a higher mean saliency in the high saliency condition than in the low saliency condition, a result which replicated using both models (Itti & Koch: low saliency M = 0.08, SEM = 0.01 vs. high saliency M = 0.12, SEM = 0.01; GBVS: M = 0.31, SEM = 0.02 vs M = 0.37, SEM = 0.2). This difference was statistically significant across the different scenes ($t(26) = 2.96$, p =.006 and $t(26) = 2.26$, p =.03 for Itti & Koch and GBVS, respectively). For comparison, the social region of interest tended to be of low saliency (Itti & Koch: M = 0.08, SEM = 0.01; GBVS: M = 0.30, SEM = 0.02).

Next, we repeated our main GLMM analyses, but this time including the mean GBVS of the region as a continuous effect (we chose to use the GBVS as it is more common in the recent literature, although predictions from the two models are correlated and tend to make the same predictions). For the analysis of fixations to the person, the saliency value of that region (which was not manipulated in our design) had no main effect (χ^2 (1) = 0.28, p =.59) but was implicated in a reliable interaction with the task (χ^2 (3) = 10.8, p =.013). Follow-up analysis showed that while there was no effect of mean GBVS in free viewing or in the "don't look object" condition, there was an effect in the "don't look social condition" (β = -7.05 ± 1.9 SE, p <.001). Intriguingly, this was in the opposite direction to what one might expect, with people having higher saliency values less likely to be fixated (when the instruction was to avoid them). For fixations to the non-social object, the GBVS value of that region was not a significant predictor and did not interact with task (χ^2 (3) = 4.4, p =.22).

Finally, we might expect the "don't look" task to be most difficult when the to-be-avoided object is highly salient and the other object is not (i.e., when there is a large difference). To capture this, we calculated a difference score for each image by subtracting the mean GBVS of the non-social object with the mean of the social region. A more negative score indicates that the object is more salient than the person, while a positive score indicates that the person is more salient than the object. However, this score had no significant effect on fixations to the person in free viewing (χ^2 (1) = 2.4, p =.12) and it was also not a predictor of the errors in the "don't look" conditions (χ^2 (2) = 2.8, p =.25).

4. Discussion

The current study used an image-viewing task to investigate visual attention towards social and non-social elements. The participants were instructed to keep their gaze away from certain areas (social or non-social) or to freely view the image. We expected that social information would be potent at capturing attention and fixations, but that it might also be treated differently when participants were asked to avoid looking. The results showed a clear tendency for participants to look at the person in the scene, which is consistent with our previous results using similar stimuli (Martinez-Cedillo, Dent, & Foulsham, 2022, 2023) as well as many other studies (Crouzet et al., 2010; End & Gamer, 2019; Birmingham, Flechsenhar, & Gamer, 2017). People were looked at more often than the non-social objects, even on the first saccade (about 40 % of first saccades to the social object compared to 12 % to the non-social object). Although we can think of the two regions of interest as competing for attention in this task, the saliency of the non-social object had no effect on looks to the social object during free viewing. These findings confirm a strong and robust early preference for looking at people in images.

Interestingly, the effect of the "don't look" instructions showed that it was harder to avoid looking at the social region of interest than the non-social region of interest. Although the frequency of errors (looking at the object that should have been avoided) was low across the whole trial, there were significantly more errors made to the person. Participants were able to avoid the specified region in most cases, indicating that fixations are under volitional control, but this was more difficult in the "don't look social" condition. This is similar to the effect found by Laidlaw et al., (2012) who used single faces as stimuli and observed that it was harder to avoid looking at the eyes than at the mouth. Our results provide strong evidence that attentional orienting to people in scenes has an automatic component which is not entirely under cognitive control. Since this automatic orienting was seen, to a greater degree, on the first saccade, it must depend on information extracted from the first, central fixation on the scene. In this sense, participants may be using specialised processing for faces and animate objects which interacts with attentional selection at an early stage of processing (as discussed by Crouzet et al., 2010; Morand et al., 2010).

There were more errors in general on the first saccade in the image. It may be that this is because top-down control takes some time to exert, something consistent with reports of "reflexive" capture by salient items on the very first saccade (Anderson & Donk, 2017). On the other hand, there is evidence that task priorities (i.e., looking for a target) override salient items in a scene even at the start of scene viewing (Einhauser, Rutishauser & Koch, 2008). Importantly, in the present study, the location of the to-be-avoided region was not known in advance. This means that, at least some of the time, participants may have had to explore the image in order to decide what it is that they should be avoiding, and making errors in the process. In future research it would be interesting to provide participants with prior knowledge about where the person was, and then see whether errors could be completely eliminated. It might also be possible to increase the motivation of participants to comply with the instructions (e.g., by rewarding them contingent on not making erroneous saccades), although it should be noted that participants in the present study clearly were following the instructions (and mostly performing quite well).

One possible limitation of the current design was that the non-social object changed on every trial and varied in identity, unlike the social region of interest (which varied in appearance, but was always a person). However, the variability of the non-social object is unlikely to have caused our key effect – that looking away from a person leads to more errors. Having a non-social object that changed on every trial presumably made it harder for participants to adopt a template for what to avoid (in the "don't look object" condition) and the uncertainty might have made it more likely that participants would have to search around and identify the object. These possibilities would act against our

prediction that social information would be harder to avoid. But, in fact, there is no evidence in the current data that participants in a particular trial were unsure about the identity of the object, and they were largely very good at avoiding it. In future it could be better to keep the non-social object constant, and this would provide a test of whether having a consistent object to avoid will make exercising top-down control more effective.

In the present study the manipulated saliency of the non-social object had relatively minor effects. Participants were more likely to look at the non-social object when it was more salient, but this effect was only seen during free viewing (when the social object received much more attention) and not during the "don't look" conditions. The saliency conditions were confirmed by additional analysis with a more recent saliency map model, but the mean model-predicted saliency of a region had few effects on the fixation probability (and the effect of the non-social object saliency was not replicated). It is possible that our sample size of participants and images was not large enough to capture saliency effects robustly. It is also possible that a larger difference in feature contrast would have produced different effects, and some features such as a sudden movement would be more likely to grab attention even in a complex scene (Mital et al., 2011). Objects which were higher in salience, and which attracted more attention during free viewing, were not more distracting in the "don't look object" condition, and neither did they garner more fixations in the "don't look social" condition (when they were a permitted place to look). This suggests that saliency does not act as a guidance signal when strong top-down instructions are in place. It would also be possible to modify the saliency of the social region of interest (which was not done here). Previous research indicates that saliency has rather little effect on the bias towards looking at people (Birmingham et al., 2009; Flechsenhar & Gamer, 2017). Interestingly, in our analysis of mean saliency values we found that social regions which were more salient were actually *less* likely to be looked at in error during the "don't look social" condition. Although this pattern was not seen in free-viewing of for the non-social object, it is possible that saliency here was confounded by peripheral visibility (and so more salient people were also easier to identify and avoid without fixating).

When examining the proportion of errors made by different participants (see Fig. 3) it is clear that there is quite a wide variation. Our statistical approach helped to control for this (using random effects) but it is interesting to observe the individual differences. For example, some participants made zero "errors", even on the first saccade, while others fixated the to-be-avoided person on more than 50 % of trials. We reported a weak correlation between errors in the two "don't look" conditions, which should be confirmed in future research. Such individual differences could, in future, be linked to traits in impulsivity or ADHD symptomology. Such traits seem to correlate with errors in antisaccade tasks (e.g., Maron et al., 2021), and they may also predict scan patterns in scenes (Hayes & Henderson, 2017; 2018). In a previous study we found that those with high levels of ADHD-like traits were less likely to look at social information in some cases, but there were no consistent effects on distractibility by a dual memory task (Martinez-Cedillo, et al., 2022).

In conclusion, this study demonstrates that social regions are prioritized by fixations in scenes. This prioritization can be overcome by the task, but there remains a proportion of automatic eye movements which select people in scenes despite an instruction not to look there. This is consistent with early, preferential processing of social information even in a complex image.

CRediT authorship contribution statement

A.P. Martinez-Cedillo: Writing – original draft, Visualization, Investigation, Methodology, Formal analysis. **T. Foulsham:** Conceptualization, Methodology, Formal analysis, Investigation, Supervision, Writing – original draft.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

The experiment was preregistered. Data, experimental and analysis code are available via the Open Science Framework at: https://osf.io/4aewz/

References

Anderson, N. C., & Donk, M. (2017). Salient object changes influence overt attentional prioritization and object-based targeting in natural scenes. *PLoS One, 12*(2), e0172132.

Anderson, N. C., Ort, E., Kruijne, W., Meeter, M., & Donk, M. (2015). It depends on when you look at it: Salience influences eye movements in natural scene viewing and search early in time. *Journal of Vision, 15*(5), 9.

Bates, D., Maechler, M., Bolker, B., Walker, S., Christensen, R. H. B., Singmann, H., & Grothendieck, G. (2009). Package 'lme4'. *URL http://lme4. r-forge. r-project. org*.

Bindemann, M., Scheepers, C., Ferguson, H. J., & Burton, A. M. (2010). Face, body, and center of gravity mediate person detection in natural scenes. *Journal of Experimental Psychology: Human Perception and Performance, 36*(6), 1477–1485. https://doi.org/10.1037/a0019057

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

Birmingham, E., Bischof, W. F., & Kingstone, A. (2009). Saliency does not account for fixations to eyes within social scenes. *Vision research, 49*(24), 2992–3000.

Borji, A., & Itti, L. (2012). State-of-the-art in visual attention modeling. *IEEE transactions on pattern analysis and machine intelligence, 35*(1), 185–207.

Brysbaert, M., & Stevens, M. (2018). Power Analysis and Effect Size in Mixed Effects Models: A Tutorial. *Journal of. Cognition, 1*(1). https://doi.org/10.5334/joc.10

Buswell, G. T. (1935). *How people look at pictures: A study of the psychology and perception in art*. Chicago Press: Univ.

Cerf, M., Harel, J., Einhäuser, W., & Koch, C. (2007). Predicting human gaze using low-level saliency combined with face detection. In *Advances in neural information processing systems* (p. 20).

Crouzet, S. M., Kirchner, H., & Thorpe, S. J. (2010). Fast saccades toward faces: face detection in just 100 ms. *Journal of vision, 10*(4), 16.

Downing, P. E., Peelen, M. V., Wiggett, A. J., & Tew, B. D. (2006). The role of the extrastriate body area in action perception. *Social Neuroscience, 1*(1), 52–62.

Einhauser, W., Rutishauser, U., & Koch, C. (2008). Task-demands can immediately reverse the effects of sensory-driven saliency in complex visual stimuli. *Journal of vision, 8*(2), 2.

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

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.

Flechsenhar, A. F., & Gamer, M. (2017). Top-down influence on gaze patterns in the presence of social features. *PLoS One1, 12*(8), e0183799.

Foulsham, T. (2019). Scenes, saliency maps and scanpaths. In C. Klein & U. Ettinger (Eds.), Eye movement research. Studies in neuroscience, psychology and behavioral economics (pp. 197–238). Springer. 10.1007/978-3-030-20085-5_6.

Foulsham, T., & Underwood, G. (2007). How does the purpose of inspection influence the potency of visual salience in scene perception? *Perception, 36*(8), 1123–1138.

Foulsham, T., & Underwood, G. (2008). What can saliency models predict about eye movements? Spatial and sequential aspects of fixations during encoding and recognition. *Journal of Vision, 8*(2), 6.

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.

Golan, T., Bentin, S., DeGutis, J. M., Robertson, L. C., & Harel, A. (2014). Association and dissociation between detection and discrimination of objects of expertise: Evidence from visual search. *Attention, Perception, & Psychophysics, 76*(2), 391–406. https://doi.org/10.3758/s13414-013-0562-6

Green, P., & MacLeod, C. J. (2016). SIMR: An R package for power analysis of generalized linear mixed models by simulation. *Methods in Ecology and Evolution, 7*(4), 493–498. https://doi.org/10.1111/2041-210X.12504

Harel, J., Koch, C., & Perona, P. (2007). Graph-based visual saliency. In *Advances in neural information processing systems* (pp. 545–552). Cambridge, MA: Massachusetts Institute of Technology.

Haxby, J. V., Hoffman, E. A., & Gobbini, M. I. (2000). The distributed human neural system for face perception. *Trends in Cognitive Sciences, 4*, 223–233.

Hayes, T. R., & Henderson, J. M. (2017). Scan patterns during real-world scene viewing predict individual differences in cognitive capacity. *Journal of Vision, 17*(5), Article 23. https://doi.org/10.1167/17.5.23

Hayes, T. R., & Henderson, J. M. (2018). Scan patterns during scene viewing predict individual differences in clinical traits in a normative sample. *PLOS ONE, 13*(5), Article e0196654. 10.1371/journal.pone.0196654.

Henderson, J. M. (2020). Meaning and attention in scenes. *Psychology of learning and motivation, 73*, 95–117.
Henderson, J. M., & Hayes, T. R. (2017). Meaning-based guidance of attention in scenes as revealed by meaning maps. *Nature human behaviour, 1*(10), 743–747.
Henderson, J. M., & Hayes, T. R. (2018). Meaning guides attention in real-world scene images: Evidence from eye movements and meaning maps. *Journal of Vision, 18*(6), 10.
Hershler, O., & Hochstein, S. (2005). At first sight: A high-level pop out effect for faces. *Vision Research, 45*(13), 1707–1724. https://doi.org/10.1016/j.visres.2004.12.021
Itti, L., & Koch, C. (2000). A saliency-based search mechanism for overt and covert shifts of visual attention. *Vision research, 40*(10–12), 1489–1506.
Kiat, J. E., Hayes, T. R., Henderson, J. M., & Luck, S. J. (2022). Rapid extraction of the spatial distribution of physical saliency and semantic informativeness from natural scenes in the human brain. *Journal of Neuroscience, 42*(1), 97–108.
Kummerer, M., Wallis, T. S., Gatys, L. A., & Bethge, M. (2017). Understanding low-and high-level contributions to fixation prediction. In *In Proceedings of the IEEE international conference on computer vision* (pp. 4789–4798).
Laidlaw, K. E., Risko, E. F., & Kingstone, A. (2012). A new look at social attention: Orienting to the eyes is not (entirely) under volitional control. *Journal of Experimental Psychology: Human Perception and Performance, 38*(5), 1132.
Laidlaw, K. E., 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.
Laidlaw, K. E., & Kingstone, A. (2017). Fixations to the eyes aids in facial encoding; covertly attending to the eyes does not. *Acta Psychologica, 173*, 55–65.
Lakens, D. (2022). Improving Your Statistical Inferences. Retrieved from https://lakens.github.io/statistical_inferences/. 10.5281/zenodo.6409077.
Martinez-Cedillo, A. P., Dent, K., & Foulsham, T. (2022). Do cognitive load and ADHD traits affect the tendency to prioritise social information in scenes? *Quarterly Journal of Experimental Psychology, 75*(10), 1904–1918.
Martinez-Cedillo, A. P., Dent, K., & Foulsham, T. (2023). Social prioritisation in scene viewing and the effects of a spatial memory load. *Attention, Perception, & Psychophysics*, 1–11.
Maron, D. N., Bowe, S. J., Spencer-Smith, M., Mellahn, O. J., Perrykkad, K., Bellgrove, M. A., & Johnson, B. P. (2021). Oculomotor deficits in attention deficit hyperactivity disorder (ADHD): A systematic review and comprehensive meta-analysis. *Neuroscience & Biobehavioral Reviews, 131*, 1198–1213.
Mital, P. K., Smith, T. J., Hill, R. L., & Henderson, J. M. (2011). Clustering of gaze during dynamic scene viewing is predicted by motion. *Cognitive computation, 3*, 5–24.
Morand, S. M., Grosbras, M. H., Caldara, R., & Harvey, M. (2010). Looking away from faces: Influence of high-level visual processes on saccade programming. *Journal of Vision, 10*(3), 16.
Morey, R. D. (2008). *Tutorials in Quantitative Methods for Psychology, 4*, 61–64.
Nuthmann, A., & Einhäuser, W. (2015). A new approach to modeling the influence of image features on fixation selection in scenes. *Annals of the New York Academy of Sciences, 1339*(1), 82–96.
Nyström, M., & Holmqvist, K. (2008). Semantic Override of Low-level Features in Image Viewing – Both Initially and Overall. *Journal of Eye Movement Research, 2*(2). https://doi.org/10.16910/jemr.2.2.2
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.
Risko, E. F., Laidlaw, K. E., 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.
Rösler, L., End, A., & Gamer, M. (2017). Orienting towards social features in naturalistic scenes is reflexive. *PLoS One1, 12*(7), e0182037.
Tatler, B. W., Hayhoe, M. M., Land, M. F., & Ballard, D. H. (2011). Eye guidance in natural vision: Reinterpreting salience. *Journal of vision, 11*(5), 5.
Thompson, S. J., Foulsham, T., Leekam, S. R., & Jones, C. R. (2019). Attention to the face is characterised by a difficult to inhibit first fixation to the eyes. *Acta psychologica, 193*, 229–238.
Tsao, D. Y., & Livingstone, M. S. (2008). Mechanisms of face perception. *Annual Review of Neuroscience, 31*, 411–437.
Underwood, G., & Foulsham, T. (2006). Visual saliency and semantic incongruency influence eye movements when inspecting pictures. *Quarterly Journal of Experimental Psychology, 59*(11), 1931–1949.
Yarbus, A. L. (1967). Eye movements during perception of complex objects. *Eye movements and vision*, 171–211.