scientific reports

Check for updates

OPEN

Looking at faces in the wild

Victor P. L. Varela[1], Alice Towler[1,2], Richard I. Kemp[1] & David White[1]✉

Faces are key to everyday social interactions, but our understanding of social attention is based on experiments that present images of faces on computer screens. Advances in wearable eye-tracking devices now enable studies in unconstrained natural settings but this approach has been limited by manual coding of fixations. Here we introduce an automatic 'dynamic region of interest' approach that registers eye-fixations to bodies and faces seen while a participant moves through the environment. We show that just 14% of fixations are to faces of passersby, contrasting with prior screen-based studies that suggest faces automatically capture visual attention. We also demonstrate the potential for this new tool to help understand differences in individuals' social attention, and the content of their perceptual exposure to other people. Together, this can form the basis of a new paradigm for studying social attention 'in the wild' that opens new avenues for theoretical, applied and clinical research.

A brief glance at a person signals a wealth of critical social information. Information about their emotional state, intentions, demographics and identity all serve to enable us to navigate our social world successfully. Understanding the perceptual processes responsible for these impressive abilities has been an important focus of research in social cognition.

One approach to studying person perception is to examine socially-directed attention by analysing people's eye movements as they view images of people presented on computer screens (e.g.[1–6]). However, photographs of social scenes do not represent the dynamic and multidimensional reality of our social experience. Indeed, participants fixate on faces less in face-to-face interactions than when watching video stimuli[7–10], indicating that contrived laboratory tasks are inadequate analogues of real-world social attention ([11] see also[7,8]).

Surprisingly little is known about how people direct their attention towards others in natural settings. Yet, this information provides valuable constraints to understanding the perceptual processes and mechanisms of attention. For example, researchers have captured the visual experience of babies and toddlers using wearable cameras, enabling researchers to investgate how perceptual expertise with faces develops. This work shows that faces are present in infants' field of view roughly 25% of the time (e.g.[12]), with the vast majority of this exposure being to familiar faces of primary caregivers. In contrast, faces make up a far smaller fraction of children's visual experience beyond their first birthday (~ 10%, e.g.[13,14]). The extent to which babies and children *attend* to these faces is less clear. More generally, quantifying and characterising faces that are attended provides the basis for developing theories of perceptual expertise, by grounding them in the visual information sampled from the environment (e.g. see[15]).

Studies of adults' attention to people in natural settings are extremely rare, and almost all knowledge on this topic comes from tightly controlled laboratory-based research. This laboratory-based research shows, for example, that faces capture attention and are processed preferentially relative to non-face objects and bodies[16–18], and this leads to the view that this process is automatic ([19], for a review see[20]). However, it is not clear whether this holds for ambient environments populated with many competing stimuli each with its unique affordance[21] and where the 'social stimuli' are real people, complete with minds, eyes and legs of their own.

This knowledge gap has increased interest in methods that allow studies of person perception and social attention in immersive environments. One approach has been to use virtual reality, with faces rendered on animated bodies in virtual worlds[22–24]. Another has been to study social attention in "the wild" by studying the eyemovements of participants wearing eye-tracking devices that monitor their fixations as they navigate real-world ambient environments (for recent reviews see[10,25,26].

Wearable eye-tracking offers the advantage of studying social attention and person perception in situ. However, it requires experimenters to manually code what is being fixated on every fixation, amounting to thousands of manual coding decisions even for a single participant in a short 10 min study. Even coding simple aspects of gaze fixations, for example, counting person fixations vs non-person fixations, is extremely time-consuming[25,27,28]. It is therefore impractical to examine social attention in naturalistic settings at the resolution afforded

[1]University of New South Wales, Sydney, Australia. [2]University of Queensland, Brisbane, Australia. ✉email: david.white@unsw.edu.au

by lab-based eye-tracking studies (e.g.[1,4,29]). Experimenters wishing to conduct naturalistic studies of social attention are therefore limited to costly, coarse analysis of fixation patterns.

Here we introduce a novel method that automates fine-grained investigations of naturalistic social attention for the first time. Our 'dynamic regions of interest' (dROI) approach automatically measures social attention in ambient environments frame-by-frame. We achieve this by co-registering eye-movement data from a wearable eye-tracker with body and face landmark positions extracted from video data using a state-of-the-art computer vision algorithm[30]. This encodes eye fixations directed towards people and maps fixations to landmarks on the face and body. A demonstration of the dROI method is available in the Supplementary Video.

Our approach overcomes many significant limitations of prior work on social attention in natural settings, saving substantial research effort by avoiding the need for manual coding of fixations to pre-specified regions (e.g.[27,28,31–33]). In addition to removing the burden of manual coding, our approach also increases temporal resolution and the volume of data, enabling new analytic approaches which open up new avenues to study person perception 'in the wild'.

Given this is the first paper to use this approach, we address some preliminary research questions to demonstrate its potential for answering a diverse range of questions related to social attention. Our primary aim was to quantify the extent to which people attend to bodies and faces of passersby and ask whether faces 'capture' viewers' attention as claimed on the basis of lab-based studies (e.g.[16–18]). We also conducted an exploratory analysis to examine whether patterns of social in natural settings may reflect stable individual differences in observers, both when participants were walking in a public space and when they were engaged in face-to-face social interaction.

Results

Faces of passersby do not capture attention in a live natural setting. Thirty-three participants followed a circular route around a busy university campus wearing a mobile eye-tracking device (see 'Methods—Data Collection' for full technical and procedural details). We show an example video frame illustrating the eye-tracking data provided by the eye-tracker and the detected dynamic regions of interest in Fig. 1A (left panel). Our dynamic region of interest (dROI) analysis of social attention relied on automatic face and body detection algorithms developed by Cao and colleagues (OpenPose:[30]). We verified the accuracy of this algorithm on our video data by comparing its detections to manual coding of body presence by four human observers and found a high level of agreement (see Supplementary Materials Section 1.1; see also[34]).

To calculate the proportion of fixations participants made to faces and bodies, we co-registered fixation locations from the eye-tracker with landmarks on faces and bodies (Fig. 1A, see Methods—Eye gaze processing). The average width of heads detected in the scene measured 2.2° of visual angle from ear-to-ear which roughly corresponds with prior lab-based work showing attention capture by faces (e.g.[16,17]; see Supplementary Material Section. 1.2 for full head size data). For reference, the width of the head in Fig. 1A, as perceived by the participant, measured 4° visual angle from ear to ear, with a body height of 32° from chin to toe.

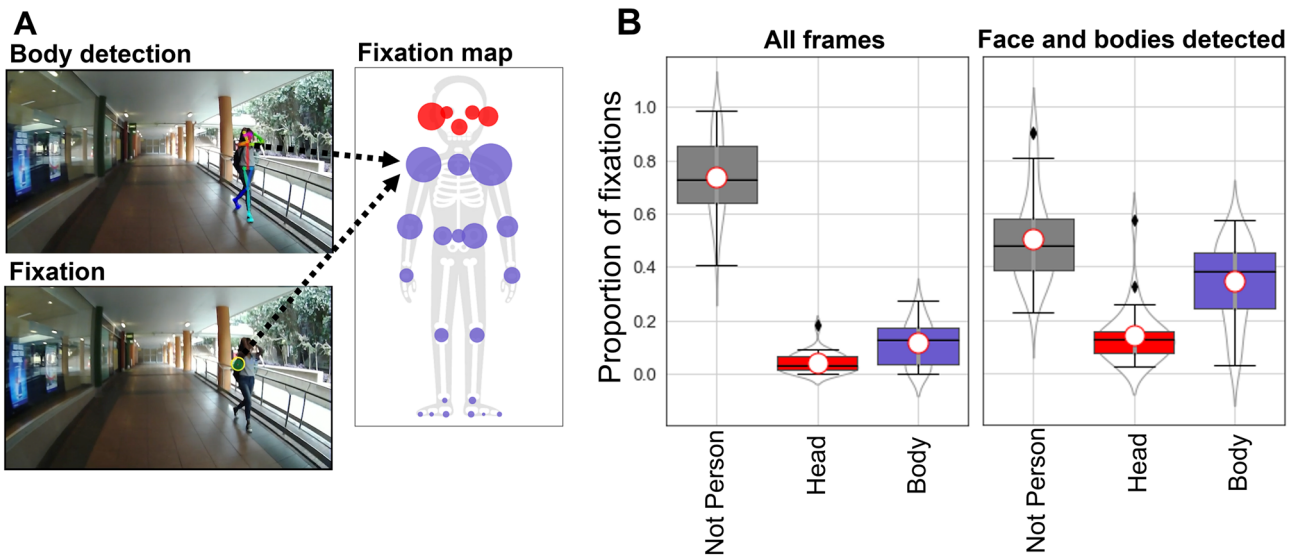

Figure 1. Dynamic region of interest (dROI) analysis of social attention while navigating a university campus. (**A**) Using data from a wearable eye-tracker, we extracted body landmarks from videos using OpenPose (top left) and co-registered viewers' fixations towards these landmarks (bottom left). The skeleton figure shows a participant's relative proportion of fixations to each body landmark, indicated by the size of the marker (all individual participant maps are available in Supplementary Materials, Sect. 1.4 and a video demonstration of the dROI method is shown in the Supplementary Video). (**B**) The left data panel shows boxplots of the proportions of non-person, head and body fixations as a proportion of all fixations in the recordings. The right data panel shows the proportions of non-person, head and body fixations only as a proportion of frames where the algorithm detected heads and bodies. See main text for analysis.

As a function of total fixations (Fig. 1B, left), 16% of fixations were directed to people, with just 4% directed at people's heads (Body: M = 11.6%, SD = 8.3%; Head: M = 4.3%, SD = 3.8%). Restricting the analysis to only frames where faces and bodies were detected by the algorithm, we observed higher proportions of fixation towards people (50%), but fixations to heads remained relatively low at 14% of fixations (Body: M = 34.4%, SD = 14.9%; Head: M = 14.4%, SD = 10.3%). The small proportion of fixations to faces may suggest that the widely reported finding that 'faces capture attention' in lab-based studies does not transfer to encounters with unfamiliar people in public spaces (e.g.[4–6,16,17,35–37]).

Because there were large variations in face width for detected faces in the field of vision (M = 2.24°, SD = 1.28°,min = 0.13°, max = 14.36), we repeated this analysis separately for above- and below-average sized faces and found a modest increase in attention to above-average sized faces (15.2% v 19.1%). However the same pattern of results shown in Fig. 1B was observed (see Supplementary Material Section 1.3).

The effect of frontal vs. averted faces on attention in live natural setting. Previous lab-based studies have shown that frontal faces capture more attention than averted faces (e.g.[38,39]). In a further test of whether unfamiliar faces capture attention in naturalistic settings, we compared the proportions of fixations to people in frames where the algorithm detected full faces (i.e. all facial features) against when the algorithm detected partial faces (i.e. subset of facial features; see Fig. 2 left, and Supplementary Materials Section 2.1 for the manual process used to verify this approach). This provided a test of whether frontal faces are fixated more than averted faces in a natural setting.

Figure 2 shows that participants made more fixations to people with fully visible frontal faces relative to averted face. However, ANOVA simple main effects showed that this increase in attention was distributed evenly between head regions (Partial = 12.3%, Full = 15.1%: $F(1,30) = 7.035$, $p = 0.013$, ç) and body regions (Partial = 32.1%, Full = 36.6%: $F(1,30) = 6.64$, $p < 0.015$, $\eta^2_p = 0.181$; see Supplementary Materials Section 2.2 for full ANOVA). This result suggests that participants were more likely to fixate on *people* when their faces were in full view, but provides no evidence that faces captured this attention any more than other body regions.

Fixation patterns during face-to-face interaction. We also recorded participants' fixation patterns during a face-to-face conversation with the experimenter (see Methods—Data collection). This conversation occurred before the main navigation task, as participants listened to scripted task instructions for about 30 s before asking any follow-up questions. Because participants were closer to the experimenter and faces were larger than in the navigation task (M = 8.17°, SD = 1.17°,min = 3.02°, max = 11.87°; see Supplementary Materials, Fig. S3), this enabled the face detection algorithm to detect 70 facial landmarks (See Methods—Eye gaze process-

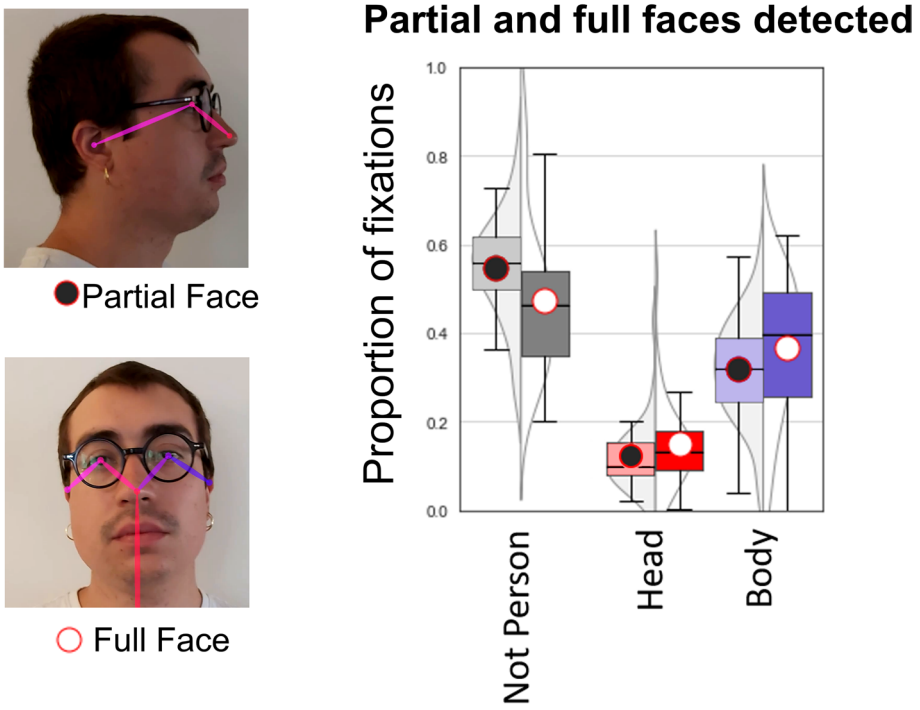

Figure 2. Comparing social attention to faces when they are fully visible versus partially visible in a video frame. Partially visible faces were due to averted head angle, or in some rare cases occlusion (top left). Results show a greater proportion of fixations to people when their faces were fully visible (see figure legend on left). See main text for analysis and Supplementary Materials (Sect. 2) for the full ANOVA and further details of the computational approach used to classify face type.

ing, Data Analysis Strategy). An example of the video recording and the proportion of fixations to each facial landmark for one participant is shown in Fig. 3A.

Unsurprisingly, participants spent far longer looking at faces when engaged in conversation compared to in the navigation task, with an average of 92.9% of fixations to people and 89.5% to faces (Fig. 3B). The proportion of fixations to different regions of the experimenter's face is shown in Fig. 3C, showing a focus on internal facial features, in line with screen-based eye-tracking studies (e.g.[1,40,41]. However, these regions are computed by assigning fixations to the closest landmark, which lacks precision because landmarks are not distributed evenly across the face. To resolve this issue, we used the spatial relations between facial landmarks and fixations to triangulate precise fixation locations (see Methods—Eye gaze processing).

Triangulation enabled us to compute heatmaps of participants gaze patterns distributed continuously across the face, as shown in the average heatmap on Fig. 4 (see Supplementary Materials Section 3.2 for participant's individual heatmaps). This average heatmap shows a tendency for participants to focus on the eyes, nose and mouth in a 'T' shaped distribution, which is a common finding in screen-based eye-tracking studies [e.g.[1,40,41]]). Interestingly, there is also a clear leftward bias observable in Fig. 4. This bias is consistent with previous laboratory-based research investigating people looking at faces on screens to perceive identity[42] and detect emotional expression[43–45].

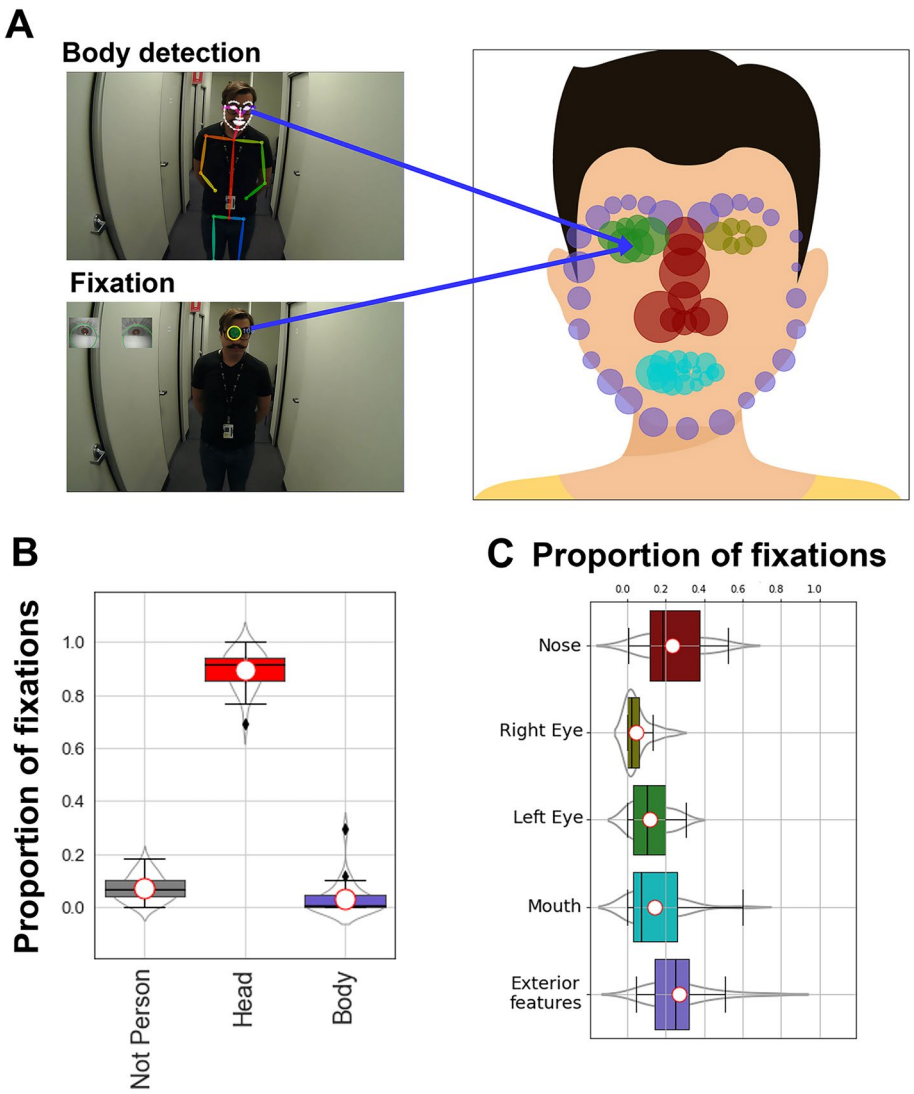

Figure 3. Dynamic region of interest analysis applied to face-to-face interaction. (**A**) We extracted facial landmarks from the video source using OpenPose (left), and these landmarks were used to register the viewers' fixation positions on the face. The width of the head from ear to ear in this image corresponds to 12.6° of visual angle. The size of circles on the schematic face shows the average proportion of fixations participants made to each landmark (individual participant fixation maps are available in Supplementary Material Sect. 3.1); (**B**) Relative frequency of fixations to the experimenter's face and body compared to the surrounding environment; (**C**) Relative frequency of fixations to facial regions indexed by colour coded mapping to landmarks shown in in panel A (right).

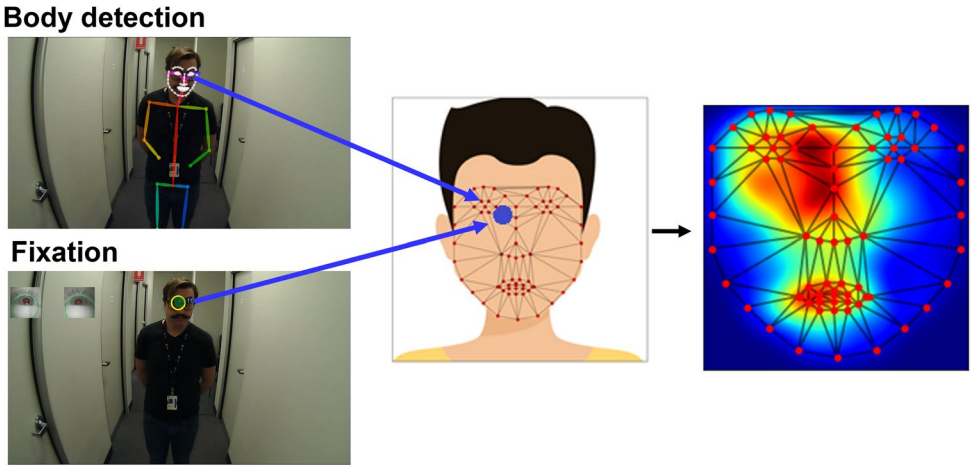

Figure 4. Heatmap analysis of face-to-face interaction. Video and eye movement from the wearable eye-tracker (left) were registered using OpenPose facial landmarks and converted to locations on the standard face template using Delaunay triangulation and affine transformations (middle). This technique enabled face fixation data to be aggregated and recorded as heatmaps (right; see Supplementary Materials Sect. 3.2 for individual participant heatmaps).

Individual differences in naturalistic social attention. Computerised lab-based tests have established that individual differences in social attention are stable across test sessions, and these differences are associated with genetic variation (e.g.[46,47]). Although it is not possible to make strong inferences about individual differences based on our relatively small sample size of 30 participants, we conducted a preliminary correlational analysis of social attention during the navigation task.

We first calculated the correlation between individuals' tendency to fixate on people and faces in two distinct segments of the navigation study route that were separated by a short rest break (see Methods—Data collection). We found a significant correlation (Spearman's rho = 0.58, $p = 0.001$, CI = [0.27,0.78], n = 30; see scatterplot in Supplementary Materials, Fig. S10), indicating that individual differences in people's tendency to fixate on people and faces was relatively stable across our test. Because the university campus was busier for some participants than others, participants' proportion of fixations to people may be affected by number of people available in the scene. We therefore repeated the correlational analysis controlling for the average number of people per frame for each participant, and found the same pattern of stable individual differences (Spearman's rho = 0.532, $p = 0.002$, CI = [0.21,0.75], n = 30; see Supplementary Material Section 4.1 for further details of this analysis).

Participants in our study had completed measures of self-reported and objective face recognition ability (Cambridge Face Memory Test extended version[48], Prosopagnosia Index short version[49]; see Methods—Data Collection). So we also conducted an exploratory analysis to examine if there was evidence of an association between attention to faces and people and face identity processing ability, but we found no significant association (see Supplementary Material Section 4.2 for further details). Exploratory analysis of individual gaze patterns to faces in the face-to-face interaction are also reported in Supplementary Materials Section 4.3 and show a modest but inconclusive association with measures of face identity processing ability.

Capturing exposure to faces in the wild. Analysis in this paper is focused on quantifying viewers' attention to people in their environment. However, the automated approach we have developed can also capture the *content* of person information sampled from these environments. In the field of face perception, the type of face information that is sampled from the environment has special theoretical significance, because exposure to faces is argued to underpin people's specialised expertise in processing faces (e.g. see[50]). The concept of 'face diet' tends to refer to the fact that people tend to be exposed to faces that are from similar demographic groups to themselves, and this has been used to explain the 'other-race effect' whereby people are better at recognising faces of their own ethnicity (e.g.[51,52]). But the composition of face exposure varies on many more dimensions, including transient properties of the face including head angle, lighting conditions and expression. The influence of attention on naturalistic face diets is unknown.

In Fig. 5, we demonstrate how combining automatic face detection with wearable eye-tracking can be used to explore the way that attention shapes 'face diets'. We focus on differences in 'within-face variation' for fixated and non-fixated faces (i.e. differences in transient aspects of facial appearance such as systematic differences in lighting, head angle or expression). We limited this analysis to faces that were fully detected with 70 facial landmarks detected, to allow further image processing. For each face that was both fully detected and fixated, we extracted both the frames in which it was fixated (n = 1601) and the frames in which it was not fixated (n = 4754). We then generated image averages of fixated and non-fixated face images via an image morphing procedure using the 70 face landmark coordinates detected by OpenPose (see Methods—Data analysis).

We computed averages in Fig. 5 so that they contained the same contribution for each face identity in each average. This means that any differences between the images is due to differences in transient aspects of

Figure 5. Image averages showing fixated and non-fixated views of the same faces. Average images of fully detected faces that were not fixated (left) and fixated (right) across all participants. This method is shown here for an illustration of what is achievable using our approach but is not intended as formal analysis. A map showing the locations of face images that contributed to these averages in the field of view is shown in Fig. S4 of Supplementary Material.

appearance, and there do appear to be subtle differences in expression and lighting. Due to the small number of faces that were fully detected in our navigation task, it was not possible to create averages for individual participants in our study, but studies with longer durations and/or higher resolution cameras may not face this same limitation. Therefore, although this preliminary work does not support inferences about systematic biases in face information sampling, it demonstrates future potential to understand the influence of attention on naturalistic face diets, and differences in demographic composition of this diet across different viewer demographics (e.g. see[53]).

Discussion

Our primary research goal was to develop and validate a new research tool to study social attention 'in the wild'. While automated registration of fixations to faces has been applied to static eye-tracking videos viewed on screen[54,55], this is the first time they have been applied to wearable eye-trackers outside of the laboratory. Measures of fixation proportions to people and bodies in a natural setting were broadly consistent with prior research using manual coding of video recordings from wearable eye-trackers[27,28]. Further, there was high agreement between our automated measures and manual experimenter coding (see Fig. S7), and measures of individual participant's fixation patterns were reliable over repeated measurements. Finally, in face-to-face interaction, patterns of fixations across face regions were consistent with general patterns observed in screen-based studies. Together, we interpret this as evidence that dynamic region of interest (dROI) approaches are valid for studying social attention in natural settings.

The dROI approach enabled us to ask some preliminary questions, inspired by screen-based studies of social attention, in natural settings for the first time. We first examined the extent to which faces automatically capture attention as participants navigated a busy public space. Contrary to conclusions based on lab-based experiments (e.g.[4–6,16,17,35–37]), we found no evidence that faces automatically capture attention 'in the wild'. Fixations to faces—when faces were visible in participants' field of view—made up a small proportion of total fixations (14%). Moreover, when comparing attention capture by faces and bodies that were fully visible and those that were only partially viewable, we found that fully visible faces increased fixations to both faces and bodies equivalently. This evidence does not support the idea that people automatically orient their attention to faces, at least for unfamiliar faces in a public space.

As expected, we found that participants spent a far greater proportion of time looking at faces during one-to-one social interaction. This difference is consistent with fixation patterns to faces appearing on-screen, which are also highly context-dependent. For example, attention is dependent on task instructions[56,57], whether the face is moving[58–60], speaking[58,61] and the non-verbal behaviour of the viewed person ([61], for a review see[62]).

Screen-based eye-tracking with free viewing of static scenes—with no task or instruction—show around 60% of fixations to faces where people are prominent in the scene[4,57] although other studies report lower estimates[56]. This variance between studies may be due to simple properties such as the prominence or size of faces in the scene, or to higher-order aspects of the images. The paradigm we have presented here provides a unique opportunity to disentangle how attention is influenced by the stimulus, situational context, as well as the goals and motivations of viewers (see[1,57,63,64] in natural settings.

Our study also points to two additional directions for future research made possible by our dROI approach. First, it opens new possibilities for understanding individual differences in social attention and face-processing ability[65]. Individual differences in attention to people were stable across different sections of the navigation route, which is consistent with screen-based eye-tracking studies that show a strong hereditary influence on patterns of attention to social scenes[46,47]. Although the sample size in our study was not large enough to make strong

inferences about whether these individual differences transfer to naturalistic settings, they do provide some assurance of measurement reliability at the individual level.

This should enable new work aiming to examine questions relating to individual differences. For example, many lab-based studies have found associations between face-processing ability and face information sampling patterns (e.g.[4,66–71] for a review see[72]). Relatedly, developmental disorders have been associated with abonomal social attention in laboratory-based studies (e.g. Autism Spectrum Disorders, see[73,74]; Psychopathy, see[75–77]). Whether these associations generalise from lab-based tests to eye movements in real-world environments is an open question.

Second, using automated face detection combined with eye-movement data enabled us to visualise the faces that people fixated on in our study, offering a window into participants' perceptual experience of faces (see Fig. 3). Limits of the resolution of the video frame meant that we were only able to visualise a small subset of the viewed faces in this study (see Methods), but future methodological and technological development promises to illuminate how social attention shapes a person's face 'diet'. Despite the theoretical importance of this exposure in understanding how our perceptual system develops expertise for faces (see[15]) and the continuing development of this expertise in adulthood[78,79] the amount and quality of perceptual experience people have with faces is currently limited to studies of infants and children (but see[12–14]).

Finally, we believe the present study only scratches the surface of what is possible using this new approach. We expect that the approach will enable new questions to be asked in naturalistic social perception research. Future work could take on ambitious aims, for example, to capture a more complete picture of people's perceptual exposure to faces in their daily lives. Intuitively, this exposure contains rich diversity, such as the familiarity of people we encounter, the contexts and viewing conditions we encounter them in, the nature of our social interactions and the motivations behind them. Characterising the multidimensional nature of this perceptual data and differences in how individuals sample it should offer a critical foundation for the development of theory in this field.

Methods

All methods were carried out in accordance with relevant guidelines and regulations, were approved by UNSW Human Research Ethics Advisory Panel, and informed consent was obtained from all subjects. Where images of people appear in the figures, all subjects have provided informed consent for publication of identifying information in an online open-access publication.

Data collection. *Participants.* Thirty-three university students from UNSW Sydney completed the study in return for course credit (9 male, 24 females; Age M = 21.4, SD = 5.4). We did not record participant's ethnicity. We excluded full data from two participant's because of corrupt eye-tracking data. In addition, procedural issues meant that data from one segment of the navigation task was deleted for one participant, and face-to-face task data were deleted for three participants. This gave a total of 31 participants in the main navigation task analysis ('Faces of passersby do not capture attention in live natural settings'), 30 in the individual differences analysis of the navigation task and 28 in the face-to-face interaction analysis.

Materials. We used a wearable eye-tracking device to record participants' eye gaze data as they completed the study (Pupil Labs Core:[80]). This device recorded videos of participants' field of view and eye gaze coordinates. A set of three cameras achieve this recording, a frontal camera facing the environment and two cameras facing the eyes. The resolution of the frontal camera was 1920 × 1080 pixels at 60 frames per second, and the cameras facing the eyes were both of resolution 192 × 192 pixels at 120 frames per second. The wearable eye-tracker was connected via USB to a laptop (Dell XPS 13 7390 2-in-1 placed inside a backpack worn by the participant. We used Pupil Capture to save video and eye gaze data[80].

After completing the wearable eye-tracking tasks, participants completed a standard measure of unfamiliar face memory ability, the Cambridge Face Memory Test extended version (CMFT+[48]) and a self-report measure of face recognition ability, the Prosopagnosia Index short version (PI-20[49]). This CFMT+ asks for participants to learn and memorise the grayscale faces of 6 caucasian males to be recognised later in 102 three-alternative trials without any time limit. The CFMT+ is a challenging test because the learned faces change in angle of view and image quality in the trials. The PI-20 is composed of 20 questions such as "My face recognition ability is worse than most people", and participants must rank their responses from "Strongly agree" to "Strongly disagree".

Procedure. We conducted the study during term time when the campus was busy. There were no COVID-19 cases in Sydney at the time and so people were not wearing facemasks. We fitted the mobile eye-tracking device to participants (see Materials above). Participants then completed two tasks while the device recorded their eye gaze: a face-to-face interaction task, where they interacted with the experimenter for a brief period; and a navigation task, where they walked around the UNSW Sydney campus following a circular route.

In the face-to-face interaction task, participants stood in an empty corridor, directly facing the experimenter at a distance of 1.5 m (see Fig. 3A). Participants listened to verbal instructions provided by the experimenter about the navigation task, explaining that this was a naturalistic study and that they should walk through campus as they would on a normal day. The experimenter verbally explained the study before asking participants if they understood and had any questions before begining the task. The experimenter delivered these instructions by reciting a pre-defined script, and participants spent an average of 30 s listening and asking questions about the task. This recording was used in the analysis 'Fixation patterns during face-to-face interaction associated with face recognition ability'.

Participants then followed the experimenter to a separate room where a detailed map and pictures of the walk were on the wall showing the study route. When participants indicated they were ready to begin, participants

exited the room with the experimenter and the Navigation task began. Participants navigated a pre-defined circular route via the main campus thoroughfares passing busy places (e.g. coffee shops, library, food court) through indoor and outdoor settings. Participants were always under the experimenter's supervision, who kept a ~ 2.5 m distance behind participants. When participants arrived at the library, we asked them to stop walking and rest for a minute, which divided the study route into two segments. Segment 1 lasted approximately 12 min on average, and segment 2 approximately 4 min.

When the navigation task was complete, participants removed the wearable eye-tracker before completing the PI-20[49] and the CFMT+[48] on a desktop computer. Participants were also asked questions during the debriefing to gauge their awareness of the study's purpose. Only four participants mentioned attention to people or person perception as a potential research topic.

Eye gaze data processing. The eye-tracking device collected raw gaze data of participants. We transformed this data into fixations, saccades and blinks using open-source tools provided by the eye-tracking manufacturer (Pupil Capture and Pupil Player, see https://pupil-labs.com/products/core/). Fixations were defined as per default settings, with saccade dispersion of max 1.5° of visual angle, a minimum duration of 80 ms, and a maximum duration of 220 ms. Fixations were output as coordinates labelled to specific pixels on the frontal camera frames. For analysis, we only considered frames with fixations.

Our main methodological advance was to automatically detect the presence of people in the participant's field of view using open-source body and face detection tools (OpenPose:[30]). This tool detects people in video frames and automatically estimates up to 25 landmarks on the body and 70 on the face (if the person is sufficiently close to the viewer). Co-registering fixations with these landmarks enabled us to construct detailed maps of participants' attention to people.

We used two methods to measure participants' attention to faces and people. In the first method (see Fig. 1A), we registered fixations to the closest detected body or face landmark, considering only landmarks that OpenPose detected with a greater than 60% confidence. We chose a 60% confidence rate because our testing suggested this effectively excluded false positive 'phantom' bodies which sometimes briefly appeared in the scene. We calculated the distance between fixation coordinates and landmark coordinates for every frame containing both fixations and landmarks. Where the Euclidean distance between a fixation coordinate and the closest landmark was below a designated threshold we registered a to that landmark. Thresholds varied depending on the spatial resolution of the landmark data being used (navigation task = 70 pixels; face-to-face interaction = 30 pixels). For the purpose of analysis in the navigation task, we clustered 25 landmarks into two categories (face and body; see Fig. 1A) and for the face-to-face interaction task, we clustered 70 facial landmarks into five categories (nose, left/right eye, mouth, and the exterior of the face; see Fig. 3A).

In the second method, we aimed to determine the precise location of fixations in a face to facilitate heatmap analysis in the face-to-face interaction (see Fig. 4). We achieved this by computing the relative position of a given fixation coordinate amongst facial landmarks using Delaunay triangulations[81] followed by Affine transformations. This way, fixation coordinates that landed within a given computed landmark triangle can be projected on the relative triangle in the standard template. This method enabled us to aggregate fixation data to more precise locations on the face to create a heatmap for each participant during the task.

Data analysis. *Navigation task.* In the navigation task we registered fixations as being to the head, body, or 'not-person' fixations. Head and body fixations were registered when OpenPose had greater than 60% confidence in either head or body regions and when a fixation was detected within 70 pixels of a landmark. These criteria were designed to ensure that we did not underestimate the proportion of fixations to faces and bodies because of random error due to accuracy limits of the eye-tracker (0.6° visual angle see: https://pupil-labs.com/products/core/tech-specs/). Probabilities of fixations to each of these three dynamic regions of interest (dROI) were calculated only for frames where a fixation was recorded. These data were filtered based on OpenPose detection as described in the Results section.

Capturing exposure to faces in the wild. First, we collected images of all faces OpenPose detected the full 70 facial landmarks in each participant's video recording, and sorted these according to whether the participant had fixated on them or not. We then used a face recognition algorithm (ResNet50[82] trained on the VGGFace2 Database[83]) to find all instances of fixated faces in participants' recordings. We achieved that by estimating the number of identities in a participant's video file using K-means clustering and the Elbow method to find the most likely number of identities. This produced sets of images of single identities that had been fixated by participants, and we sorted these into fremes where the face had been fixated and frames where it has not been fixated.

This process provided a set of 1601 images that participants had fixated on and an accompanying set of 4754 images that were of these same face identities but which the participant had not fixated on. We then averaged all the images of each persons face to create an average per face identity and then averaged fixated and non-fixated faces separately to create the images shown in Fig. 5. This was achieved by first morphing face images using Delaunay triangulation with affine transform to align the detected face landmarks on each image to a standard face template. Pixel values from all face images contributing to the average were averaged and the resulting pixel information was morphed to the average face landmark locations.

Face-to-Face interaction task. For the face-to-face interaction task, we processed gaze data using landmark and heatmap registration methods. Participant heatmaps were analysed using principal components analysis (PCA) to identify major components (PCs) in the inter-individual variation of heatmaps, returning a set of PCs ranked

according to their explained variance (see also[64,66]).The raw input data for the PCA is shown in Supplementary Materials (Fig. S4).

Data availability

Data supporting analysis are available via https://github.com/UNSWfacelab/Varelaetal_LookingAtFacesInTheWild.

Received: 13 June 2022; Accepted: 28 November 2022
Published online: 16 January 2023

References

1. Yarbus, A. L. *Eye Movements and Vision* (Plenum Press, 1967).
2. Amso, D., Haas, S. & Markant, J. An eye-tracking investigation of developmental change in bottom-up attention orienting to faces in cluttered natural scenes. *PLoS ONE* **9**(1), e85701 (2014).
3. Birmingham, E., Bischof, W. F. & Kingstone, A. Social attention and real-world scenes: The roles of action, competition and social content. *Q. J. Exp. Psychol.* **61**(7), 986–998 (2008).
4. Bobak, A. K., Parris, B. A., Gregory, N. J., Bennetts, R. J. & Bate, S. Eye-movement strategies in developmental prosopagnosia and "super" face recognition. *Q. J. Exp. Psychol.* **70**(2), 201–217 (2017).
5. Rösler, L., End, A. & Gamer, M. Orienting towards social features in naturalistic scenes is reflexive. *PLoS ONE* **12**(7), e0182037 (2017).
6. Gregory, N. J., Bolderston, H. & Antolin, J. V. Attention to faces and gaze-following in social anxiety: Preliminary evidence from a naturalistic eye-tracking investigation. *Cogn. Emot.* **33**(5), 931–942 (2019).
7. Nasiopoulos, E., Risko, E. F. & Kingstone, A. Social attention, social presence, and the dual function of gaze. In *The many Faces of Social Attention* (eds Puce, A. & Bertenthal, B. I.) 129–155 (Springer, 2015).
8. Risko, E. F., Richardson, D. C. & Kingstone, A. Breaking the fourth wall of cognitive science: Real-world social attention and the dual function of gaze. *Curr. Dir. Psychol. Sci.* **25**(1), 70–74 (2016).
9. Laidlaw, K. E., Foulsham, T., Kuhn, G. & Kingstone, A. Potential social interactions are important to social attention. *Proc. Natl. Acad. Sci.* **108**(14), 5548–5553 (2011).
10. Foulsham, T. Beyond the picture frame: The function of fixations in interactive tasks. *Psychol. Learn, Motiv. Adv. Res. Theory* **73**, 33–58 (2020).
11. Kingstone, A. Taking a real look at social attention. *Curr. Opin. Neurobiol.* **19**(1), 52–56 (2009).
12. Sugden, N. A. & Moulson, M. C. These are the people in your neighbourhood: Consistency and persistence in infants' exposure to caregivers', relatives', and strangers' faces across contexts. *Vision. Res.* **157**, 230–241 (2019).
13. Jayaraman, S., Fausey, C. M. & Smith, L. B. The faces in infant-perspective scenes change over the first year of life. *PLoS ONE* **10**(5), e0123780 (2015).
14. Fausey, C. M., Jayaraman, S. & Smith, L. B. From faces to hands: Changing visual input in the first two years. *Cognition* **152**, 101–107 (2016).
15. Young, A. W. & Burton, A. M. Are we face experts? *Trends Cogn. Sci.* **22**(2), 100–110 (2018).
16. Bindemann, M., Burton, A. M., Hooge, I. T., Jenkins, R. & De Haan, E. H. Faces retain attention. *Psychon. Bull. Rev.* **12**(6), 1048–1053 (2005).
17. Theeuwes, J. & Van der Stigchel, S. Faces capture attention: Evidence from inhibition of return. *Vis. Cogn.* **13**(6), 657–665 (2006).
18. Morrisey, M. N., Hofrichter, R. & Rutherford, M. D. Human faces capture attention and attract first saccades without longer fixation. *Vis. Cogn.* **27**(2), 158–170 (2019).
19. Yan, X., Young, A. W. & Andrews, T. J. The automaticity of face perception is influenced by familiarity. *Atten. Percept. Psychophys.* **79**(7), 2202–2211 (2017).
20. Palermo, R. & Rhodes, G. Are you always on my mind? A review of how face perception and attention interact. *Neuropsychologia* **45**(1), 75–92 (2007).
21. Gibson, J. J. *The Ecological Approach to Visual Perception* (Houghton, 1979).
22. Bindemann, M. *et al.* Face identification in the laboratory and in virtual worlds. *J. Appl. Res. Memory Cognit.* **11**, 120 (2021).
23. Fysh, M. C. *et al.* Avatars with faces of real people: A construction method for scientific experiments in virtual reality. *Behav. Res. Methods* **54**, 1461 (2021).
24. Bülthoff, I., Mohler, B. J. & Thornton, I. M. Face recognition of full-bodied avatars by active observers in a virtual environment. *Vision. Res.* **157**, 242–251 (2019).
25. Foulsham, T., Walker, E. & Kingstone, A. The where, what and when of gaze allocation in the lab and the natural environment. *Vision. Res.* **51**(17), 1920–1931 (2011).
26. Tatler, B. W., Hansen, D. W. & Pelz, J. B. Eye movement recordings in natural settings. In *Eye Movement Research: An Introduction to its Scientific Foundations and Applications* (eds Klein, C. & Ettinger, U.) 549–592 (Springer, 2019).
27. Hessels, R. S., van Doorn, A. J., Benjamins, J. S., Holleman, G. A. & Hooge, I. T. Task-related gaze control in human crowd navigation. *Atten. Percept. Psychophys.* **82**(5), 2482–2501 (2020).
28. De Lillo, M. *et al.* Tracking developmental differences in real-world social attention across adolescence, young adulthood and older adulthood. *Nat. Hum. Behav.* https://doi.org/10.1038/s41562-021-01113-9 (2021).
29. Rice, A., Phillips, P. J., Natu, V., An, X. & O'Toole, A. J. Unaware person recognition from the body when face identification fails. *Psychol. Sci.* **24**(11), 2235–2243 (2013).
30. Cao, Z., Hidalgo, G., Simon, T., Wei, S. E. & Sheikh, Y. OpenPose: Realtime multi-person 2D pose estimation using part affinity fields. *IEEE Trans. Pattern Anal. Mach. Intell.* **43**(1), 172–186 (2019).
31. Mele, M. L. & Federici, S. Gaze and eye-tracking solutions for psychological research. *Cogn. Process.* **13**(1), 261–265 (2012).
32. Benjamins, J. S., Hessels, R. S., & Hooge, I. T. (2018). GazeCode: Open-source software for manual mapping of mobile eye-tracking data. In *Proceedings of the 2018 ACM Symposium on Eye-Tracking Research & Applications* (pp. 1–4).
33. Haensel, J. X. *et al.* Culture modulates face scanning during dyadic social interactions. *Sci. Rep.* **10**(1), 1–11 (2020).
34. Jongerius, C. *et al.* Eye-tracking glasses in face-to-face interactions: Manual versus automated assessment of areas-of-interest. *Behav. Res. Methods* **53**(5), 2037–2048 (2021).
35. Vuilleumier, P. Faces call for attention: Evidence from patients with visual extinction. *Neuropsychologia* **38**(5), 693–700 (2000).
36. Gamer, M. & Büchel, C. Amygdala activation predicts gaze toward fearful eyes. *J. Neurosci.* **29**(28), 9123–9126 (2009).
37. Ro, T., Russell, C. & Lavie, N. Changing faces: A detection advantage in the flicker paradigm. *Psychol. Sci.* **12**(1), 94–99 (2001).
38. Shirama, A. Stare in the crowd: Frontal face guides overt attention independently of its gaze direction. *Perception* **41**(4), 447–459 (2012).
39. Palanica, A. & Itier, R. J. Eye gaze and head orientation modulate the inhibition of return for faces. *Atten. Percept. Psychophys.* **77**(8), 2589–2600 (2015).

40. Arizpe, J., Walsh, V., Yovel, G. & Baker, C. I. The categories, frequencies, and stability of idiosyncratic eye-movement patterns to faces. *Vision. Res.* **141**, 191–203 (2017).
41. Blais, C., Jack, R. E., Scheepers, C., Fiset, D. & Caldara, R. Culture shapes how we look at faces. *PLoS ONE* **3**(8), e3022 (2008).
42. Wolff, W. (1933). The experimental study of forms of expression. *Character & Personality; A Quarterly for Psychodiagnostic & Allied Studies*.
43. Heller, W. & Levy, J. Perception and expression of emotion in right-handers and left-handers. *Neuropsychologia* **19**(2), 263–272 (1981).
44. David, A. S. Spatial and selective attention in the cerebral hemispheres in depression, mania, and schizophrenia. *Brain Cogn.* **23**(2), 166–180 (1993).
45. Ferber, S. & Murray, L. J. Are perceptual judgments dissociated from motor processes? A prism adaptation study. *Cogn. Brain Res.* **23**(2–3), 453–456 (2005).
46. Constantino, J. N. *et al.* Infant viewing of social scenes is under genetic control and is atypical in autism. *Nature* **547**(7663), 340–344 (2017).
47. Kennedy, D. P. *et al.* Genetic influence on eye movements to complex scenes at short timescales. *Curr. Biol.* **27**(22), 3554–3560 (2017).
48. Russell, R., Duchaine, B. & Nakayama, K. Super-recognizers: People with extraordinary face recognition ability. *Psychon. Bull. Rev.* **16**(2), 252–257 (2009).
49. Shah, P., Gaule, A., Sowden, S., Bird, G. & Cook, R. The 20-item prosopagnosia index (PI20): A self-report instrument for identifying developmental prosopagnosia. *R. Soc. Open Sci.* **2**(6), 140343 (2015).
50. Rhodes, G. *et al.* Adaptation and face perception: How aftereffects implicate norm-based coding of faces. In *Adaptation and After-Effects in High-Level Vision* (Oxford University Press, 2005).
51. Crookes, K. & McKone, E. Early maturity of face recognition: No childhood development of holistic processing, novel face encoding, or face-space. *Cognition* **111**(2), 219–247 (2009).
52. McKone, E. *et al.* A critical period for faces: Other-race face recognition is improved by childhood but not adult social contact. *Sci. Rep.* **9**(1), 1–13 (2019).
53. Stelter, M., Simon, D., Calanchini, J., Christ, O. & Degner, J. Real-life outgroup exposure, self-reported outgroup contact and the other-race effect. *Br. J. Psychol.* https://doi.org/10.1111/bjop.12600 (2022).
54. Hessels, R. S., Benjamins, J. S., Cornelissen, T. H. & Hooge, I. T. A validation of automatically-generated areas-of-interest in videos of a face for eye-tracking research. *Front. Psychol.* **9**, 1367 (2018).
55. Zhou, X., Vyas, S., Ning, J. & Moulson, M. C. Naturalistic face learning in infants and adults. *Psychol. Sci.* **33**(1), 135–151 (2022).
56. Short, L. A., Semplonius, T., Proietti, V. & Mondloch, C. J. Differential attentional allocation and subsequent recognition for young and older adult faces. *Vis. Cogn.* **22**(9–10), 1272–1295 (2014).
57. DeAngelus, M. & Pelz, J. B. Top-down control of eye movements: Yarbus revisited. *Vis. Cogn.* **17**(6–7), 790–811 (2009).
58. Buchan, J. N., Paré, M. & Munhall, K. G. Spatial statistics of gaze fixations during dynamic face processing. *Soc. Neurosci.* **2**(1), 1–13 (2007).
59. Foulsham, T., Cheng, J. T., Tracy, J. L., Henrich, J. & Kingstone, A. Gaze allocation in a dynamic situation: Effects of social status and speaking. *Cognition* **117**(3), 319–331 (2010).
60. Scott, H., Batten, J. P. & Kuhn, G. Why are you looking at me? It's because I'm talking, but mostly because I'm staring or not doing much. *Attent. Percep. Psychophys.* **1**(1), 109–118 (2019).
61. Võ, M. L. H., Smith, T. J., Mital, P. K. & Henderson, J. M. ˜ Do the eyes really have it? Dynamic allocation of attention when viewing moving faces. *J. Vis.* **12**(13), 1–14 (2012).
62. Hessels, R. S. How does gaze to faces support face-to-face interaction? A review and perspective. *Psychon. Bull. Rev.* **27**(5), 856–881 (2020).
63. Borji, A. & Itti, L. Defending Yarbus: Eye movements reveal observers' task. *J. Vis.* **14**(3), 29–29 (2014).
64. Han, N. X. & Eckstein, M. P. Gaze-cued shifts of attention and microsaccades are sustained for whole bodies but are transient for body parts. *Psychon Bull. Rev.* **29**, 1854–1878. https://doi.org/10.3758/s13423-022-02087-z (2022).
65. Broda, M. D. & de Haas, B. Individual differences in looking at persons in scenes. *J. Vis.* **22**(12), 9 (2022).
66. Dunn, J. D. *et al.* Face information sampling in super-recognizers. *Psychol. Sci.* **33**(9), 1615–1630 (2022).
67. Wilcockson, T. D., Burns, E. J., Xia, B., Tree, J. & Crawford, T. J. Atypically heterogeneous vertical first fixations to faces in a case series of people with developmental prosopagnosia. *Vis. Cogn.* **28**(4), 311–323 (2020).
68. Varela, V. P., Ribeiro, E., Orona, P. A., & Thomaz, C. E. (2018, October). Eye movements and human face perception: An holistic analysis and proficiency classification based on frontal 2D face images. In *Anais do XV Encontro Nacional de Inteligência Artificial e Computacional* (pp. 48–57). SBC.
69. Bird, G., Press, C. & Richardson, D. C. The role of alexithymia in reduced eye-fixation in autism spectrum conditions. *J. Autism Dev. Disord.* **41**(11), 1556–1564 (2011).
70. Bal, E. *et al.* Emotion recognition in children with autism spectrum disorders: Relations to eye gaze and autonomic state. *J. Autism Dev. Disord.* **40**(3), 358–370 (2010).
71. Riby, D. M. & Hancock, P. J. Do faces capture the attention of individuals with Williams syndrome or autism? Evidence from tracking eye movements. *J. Autism Dev. Disord.* **39**(3), 421–431 (2009).
72. Avidan, G. & Behrmann, M. Spatial integration in normal face processing and its breakdown in congenital prosopagnosia. *Ann. Rev. Vis. Sci.* **7**, 301–321 (2021).
73. Guillon, Q., Hadjikhani, N., Baduel, S. & Rogé, B. Visual social attention in autism spectrum disorder: Insights from eye tracking studies. *Neurosci. Biobehav. Rev.* **42**, 279–297 (2014).
74. Chita-Tegmark, M. Social attention in ASD: A review and meta-analysis of eye-tracking studies. *Res. Dev. Disabil.* **48**, 79–93 (2016).
75. Dadds, M. R. *et al.* Attention to the eyes and fear-recognition deficits in child psychopathy. *Br. J. Psychiatry* **189**(3), 280–281 (2006).
76. Gehrer, N. A., Duchowski, A. T., Jusyte, A. & Schönenberg, M. Eye contact during live social interaction in incarcerated psychopathic offenders. *Personal. Disord. Theory Res. Treat.* **11**(6), 431–439 (2020).
77. Gehrer, N. A., Scheeff, J., Jusyte, A. & Schönenberg, M. Impaired attention toward the eyes in psychopathic offenders: Evidence from an eye tracking study. *Behav. Res. Ther.* **118**, 121–129 (2019).
78. Germine, L. T., Duchaine, B. & Nakayama, K. Where cognitive development and aging meet: Face learning ability peaks after age 30. *Cognition* **118**(2), 201–210 (2011).
79. Dunn, J. D., Summersby, S., Towler, A., Davis, J. P. & White, D. UNSW face test: A screening tool for super-recognisers. *PLoS ONE* **15**(11), e0241747 (2020).
80. Kassner, M., Patera, W., & Bulling, A. (2014). Pupil: An open source platform for pervasive eye tracking and mobile gaze-based interaction. In *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct Publication* (pp. 1151–1160).
81. Delaunay, B. (1934). Sur la sphere vide. Izv. Akad. Nauk SSSR, Otdelenie Matematicheskii i Estestvennyka Nauk, 7(793–800), 1–2.
82. He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 770–778).
83. Cao, Q., Shen, L., Xie, W., Parkhi, O. M., & Zisserman, A. (2018). Vggface2: A dataset for recognising faces across pose and age. In *2018 13th IEEE international conference on automatic face & gesture recognition (FG 2018)* (pp. 67–74). IEEE.

Acknowledgements
This project was supported by funding from the Australian Research Council (DP190100957, FT200100353).

Author contributions
V.P.L.V.: Project design, data collection, data analysis, manuscript preparation and revision. A.T.: project design, manuscript preparation and revision; R.K.: project design, manuscript preparation and revision. D.W.: acquiring funding, project design, data analysis, manuscript preparation and revision.

Competing interests
The authors declare no competing interests.

Additional information
Supplementary Information The online version contains supplementary material available at https://doi.org/10.1038/s41598-022-25268-1.

Correspondence and requests for materials should be addressed to D.W.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.