SUPPLEMENTARY MATERIALS

Looking at faces in the wild

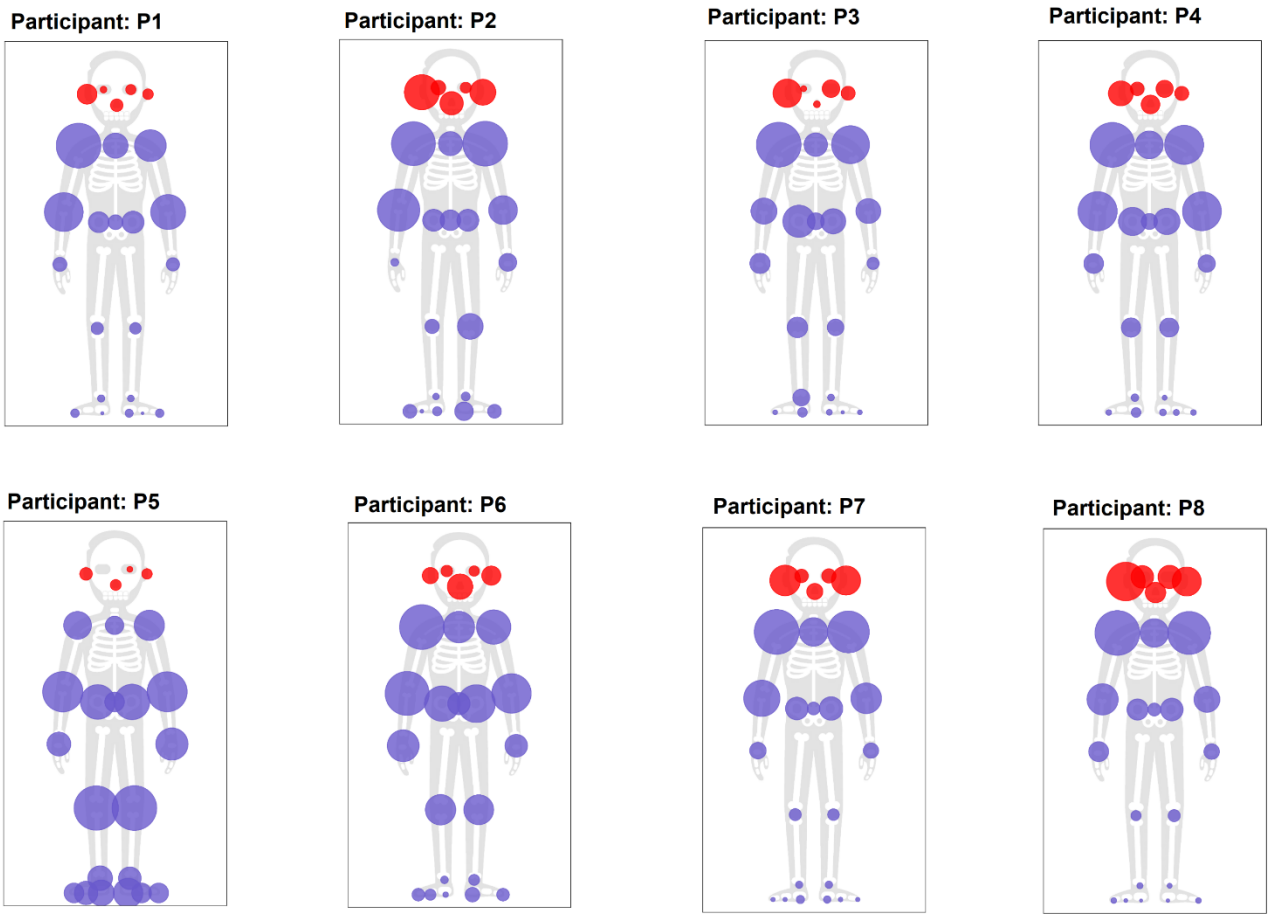
Victor PL Varela[a], Alice Towler[a], Richard Kemp[a], David White[a*]

[a]*University of New South Wales, Australia*

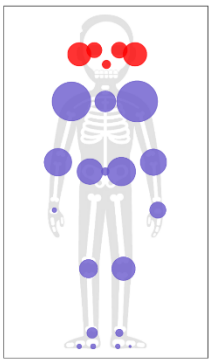
Corresponding Author: Dr David White, david.white@unsw.edu.au

Individual visualisation of participants' body maps during the navigation task

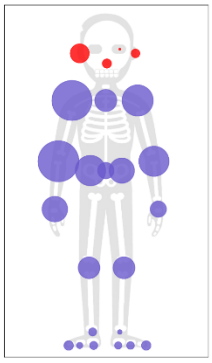
We observed high variability in the strategies engaged by participants when looking at people when navigating in the wild. Figure S1 shows proportion of fixations registered to each landmark separately for each participant in the navigation task.



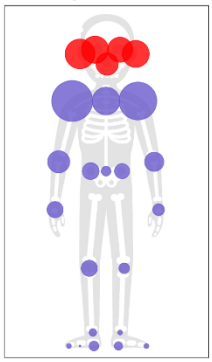
Participant: P9



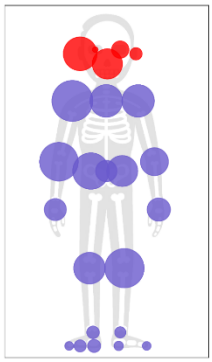
Participant: P10



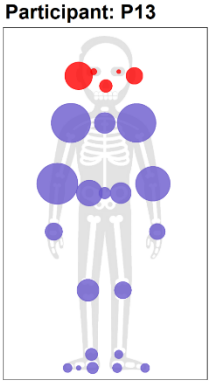
Participant: P11



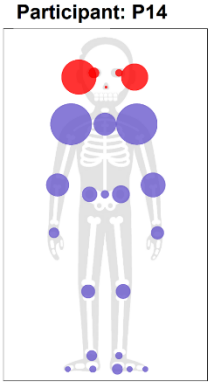
Participant: P12



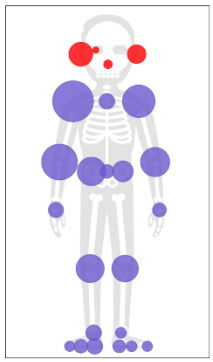
Participant: P13



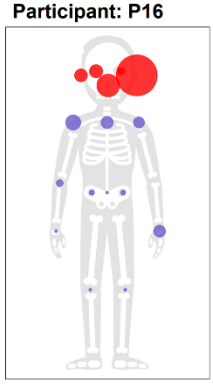
Participant: P14



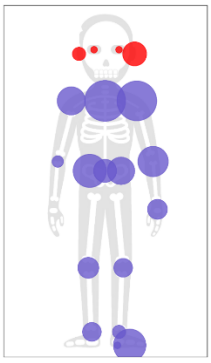
Participant: P15



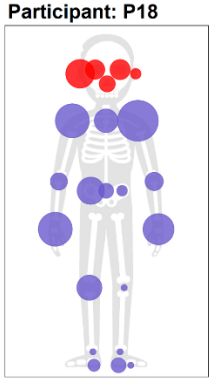
Participant: P16



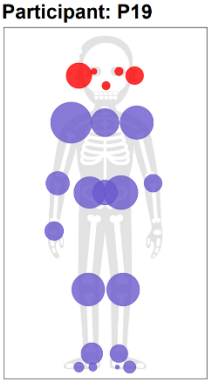
Participant: P17



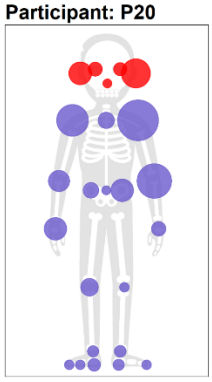
Participant: P18



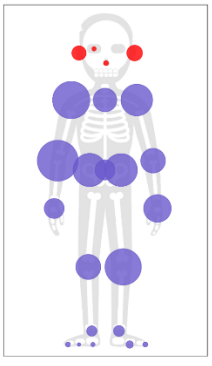
Participant: P19



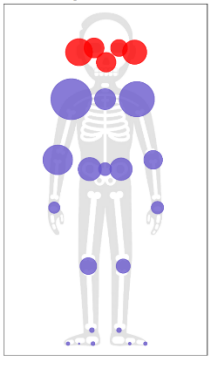
Participant: P20



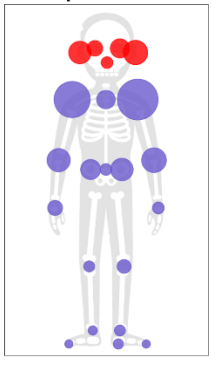
Participant: P21



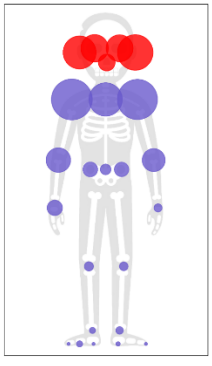
Participant: P22



Participant: P23



Participant: P24



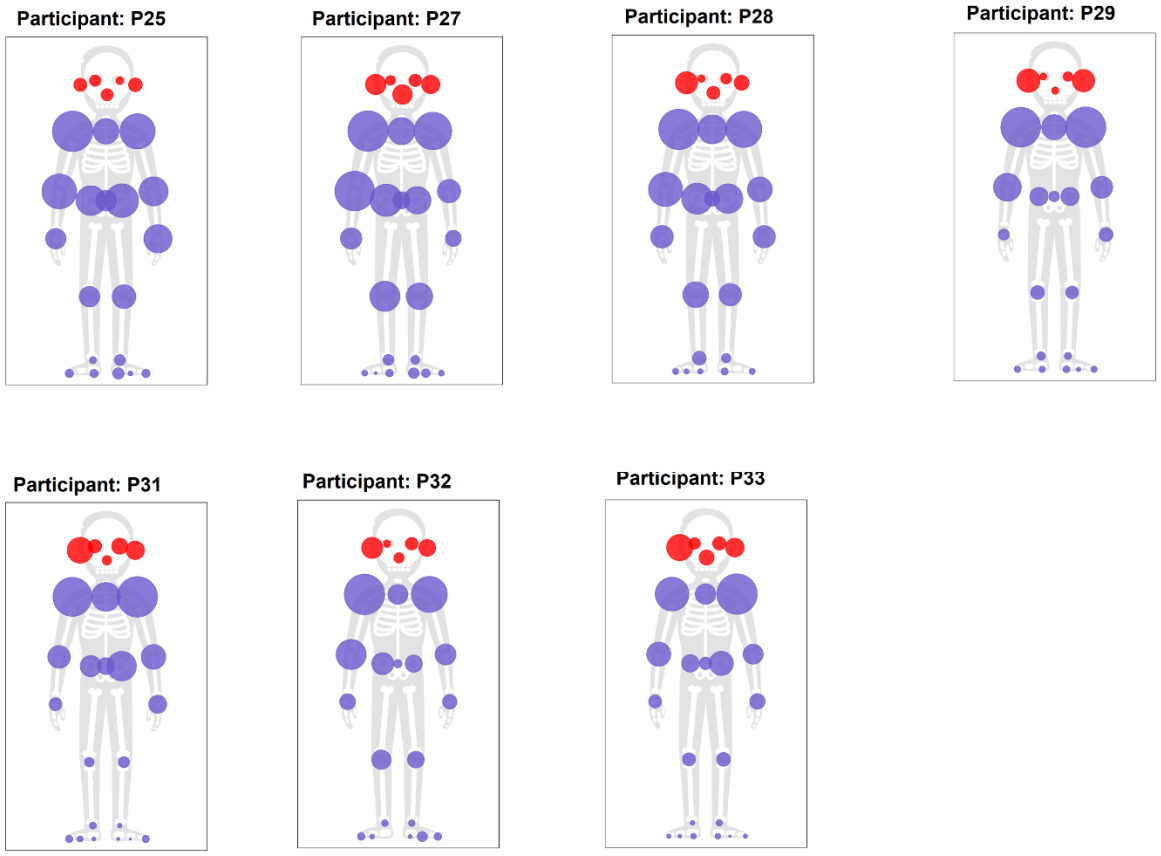


Figure S1. Individual participant data during navigation task. For each participant we show fixations on 25 dROI when viewing people during the navigation task. Face fixations are marked in red and body fixations in blue. The size of the circle for each dROI indicates the proportion of fixations participants made to that location.

Extended ANOVA analysis for 'Faces of passersby do not capture attention in live natural settings'

The main manuscript reports that participants were more likely to fixate on people in the navigation task when their faces were in full view. But we found no evidence that faces captured this attention any more than other body regions.

This conclusion was supported by an ANOVA analysis of the data shown in Figure 2 comparing the proportion of fixations to heads and bodies when faces were fully visible in a video frame versus when only particlly visible due to head rotation or other occlusions. A 2 (Face type: part face, full face detected) X 3 (Fixation type: Head, Body, Not Person fixations) ANOVA revealed a significant interaction between face and fixation type, $F(2,60) = 9.76$, $p < 0.001$, $\eta^2_p=0.246$. Analysis of simple main effects showed a significant reduction of non-person fixations, $F(1,30) = 12.86$, $p < 0.001$, $\eta^2_p=0.300$, and an increase in both head and body fixations [Head: $F(1,30) = 7.035$, $p = 0.013$, $\eta^2_p=0.190$; Body: $F(1,30) = 6.64$, $p < 0.015$, $\eta^2_p=0.181$].

Extended analysis for 'Individual differences in naturalistic social attention' (analysis of residuals)

Figure S2 shows scatterplots illustrating the individual differences analysis of residuals reported in the main text. In Figure S2A, we calculated the linear regression model predicting the probability of fixating people (head and body) as a function of the average number of people detected in video frames, separately for the two route segments (First and Second). This analysis allowed us to calculate a residual value for each participant for each route segment. Figure S2B shows the correlation between these residuals for the two route segments (Spearman's rho(29)= 0.532, p= 0.002), indicating that some participants tend to fixate more on people than others, regardless of the number of people they encountered on the walk.

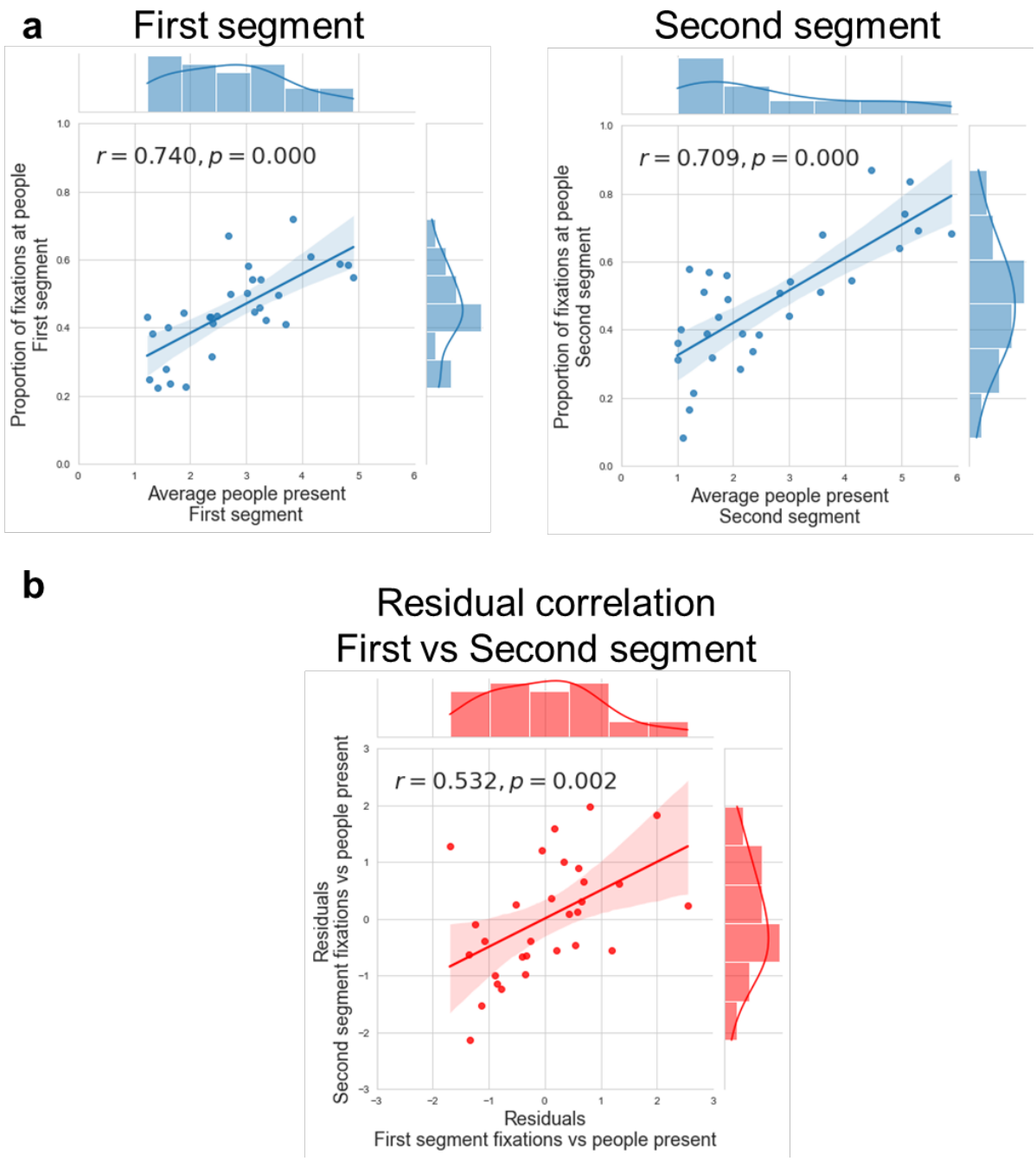
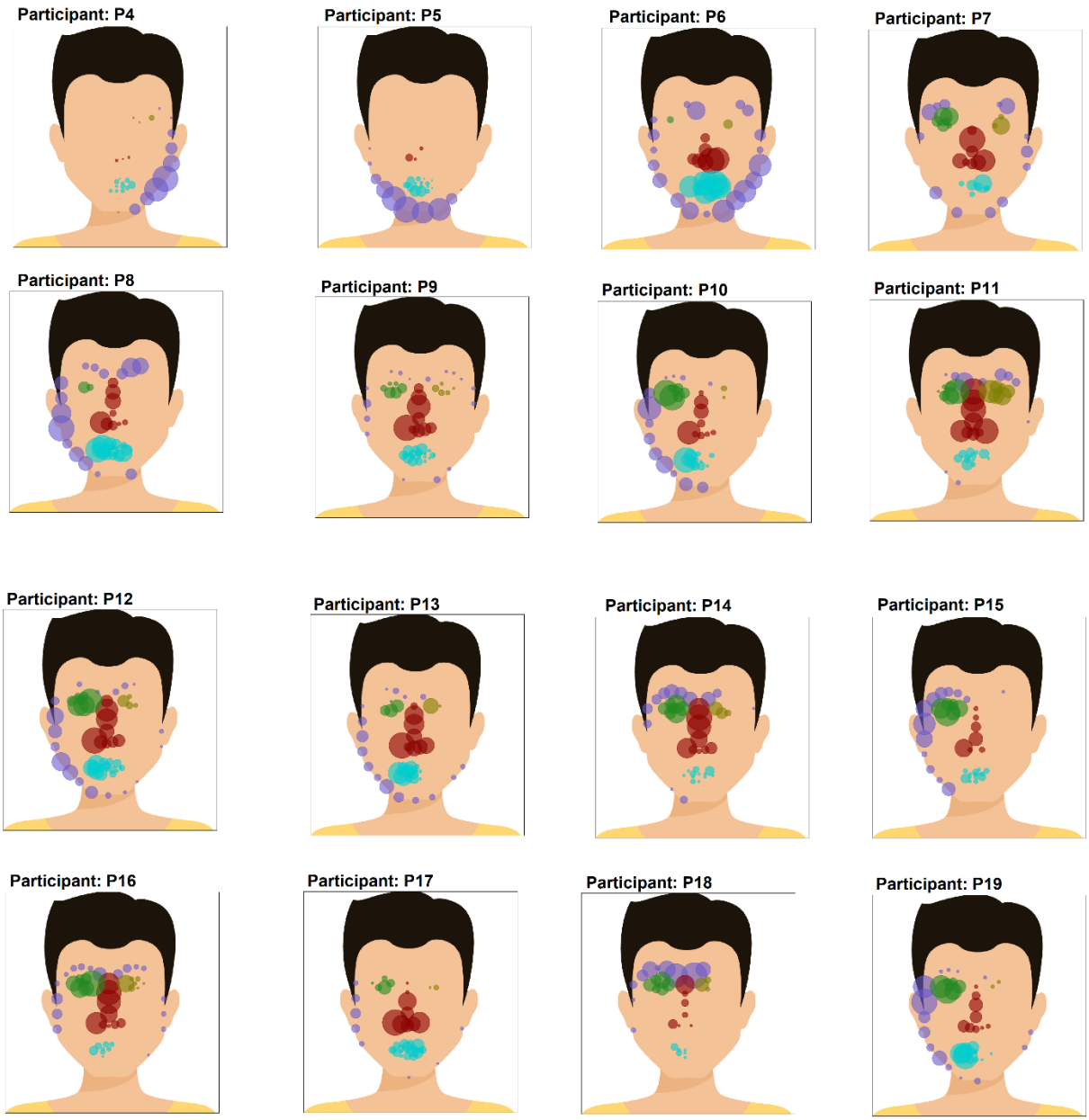


Figure S2. Individual differences analysis of residuals. Panel A shows the proportion of fixations to people as a function of the average number of people present for each route segment. Panel B shows the correlation between the residuals found in panel A.

Individual visualisation of participants' facial maps during the face-to-face interaction task

We observed high variability in the fixation patterns shown by participants when engaging in a conversation with the experimenter. We filtered participants' recordings to analyse only fixation frames that contained the experimenter's face looking straight at the participant, by using only frames where nose landmarks were

detected by OpenPose (Cao et al., 2019). The landmark registration method (see the main manuscript) detected 70 possible dynamic regions of interest (dROI) participants attended. Figure S3 shows individual participant gaze patterns registered to facial landmarks.



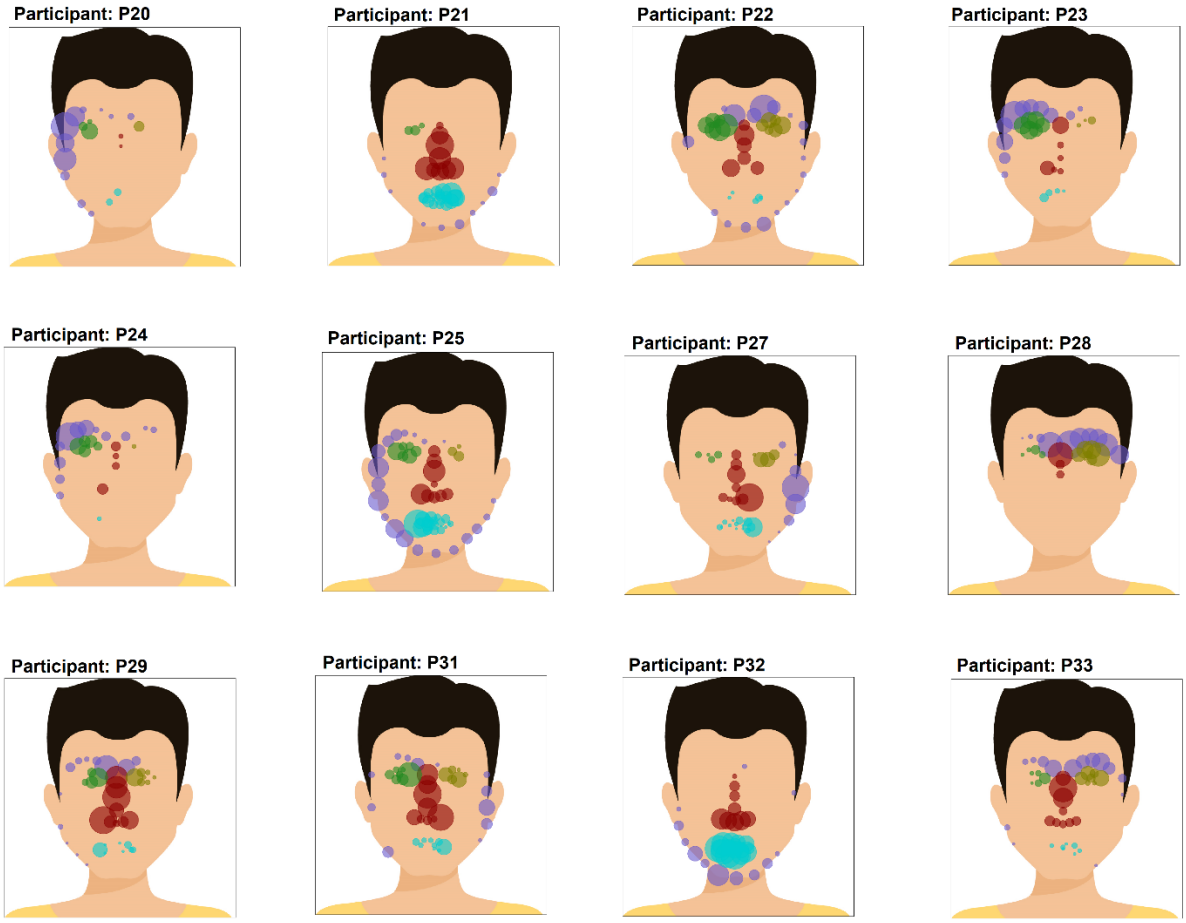
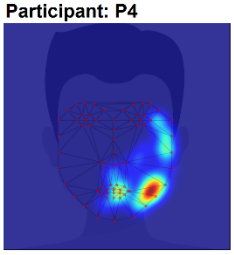


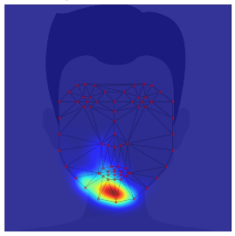
Figure S3. Individual participant data during face-to-face interaction task. We show fixations to the 70 dROI identified by OpenPose while participants attended to the experimenter during the face-to-face interaction task. The colours delimit facial regions for this visualisation, and the size of thecuircles at each dROI indicate the number of fixations at that point.

The landmark registration in Figure S3 is constrained to record 70 possible dROI positions on a face. However, the heatmap registration method allows more fine-grained analysis because it uses the relation between these landmarks to determine the exact location of where a fixation landed on a face (see main text). Figure S4 shows individual heatmaps for each participant as they focused on the experimenter's faceduring the face-to-face task. Comparing the patterns in Figures S3 and S4 there is some indication of the advantage of using triangulation of spatial location. For example, when analysed using landmark registration, P32 appears to have a relatively diffuse gaze pattern, but this is revealed as a more focal pattern when using triangulation.

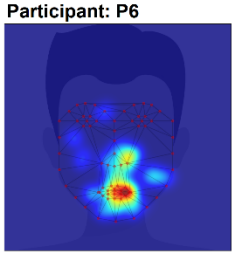
Participant: P4



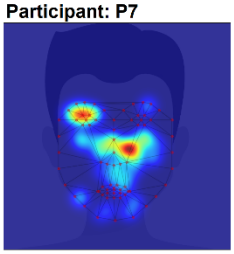
Participant: P5



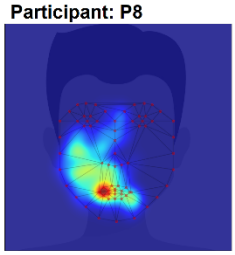
Participant: P6



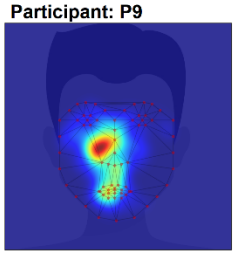
Participant: P7



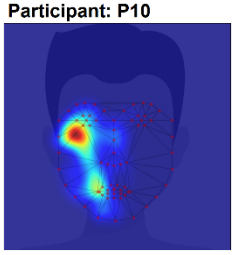
Participant: P8



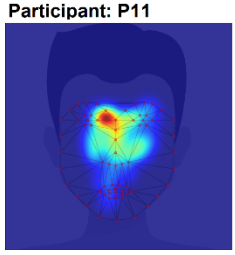
Participant: P9



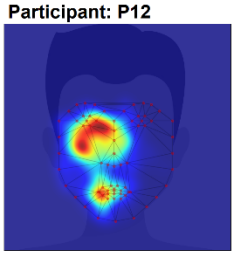
Participant: P10



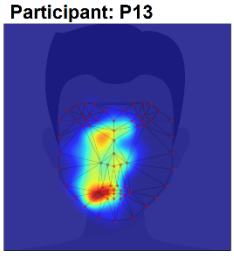
Participant: P11



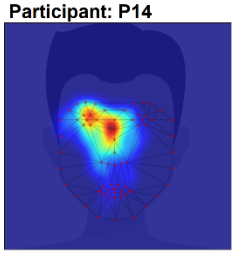
Participant: P12



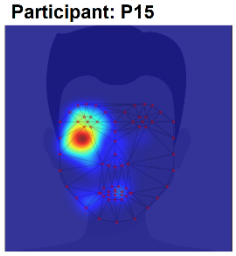
Participant: P13



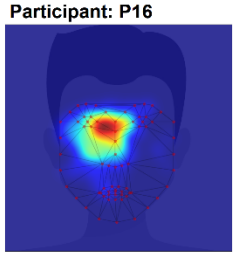
Participant: P14



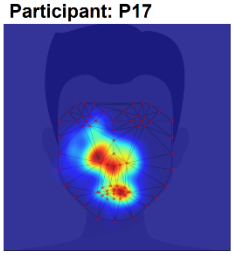
Participant: P15



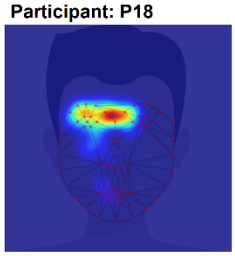
Participant: P16



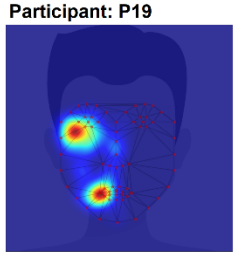
Participant: P17



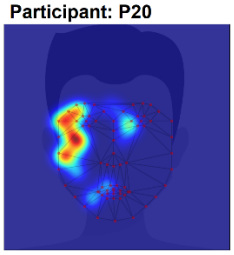
Participant: P18



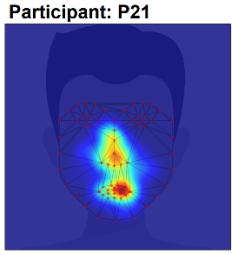
Participant: P19



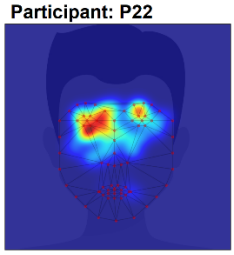
Participant: P20



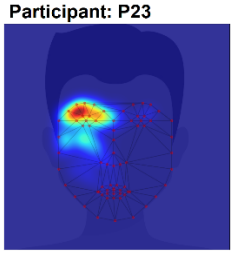
Participant: P21



Participant: P22



Participant: P23



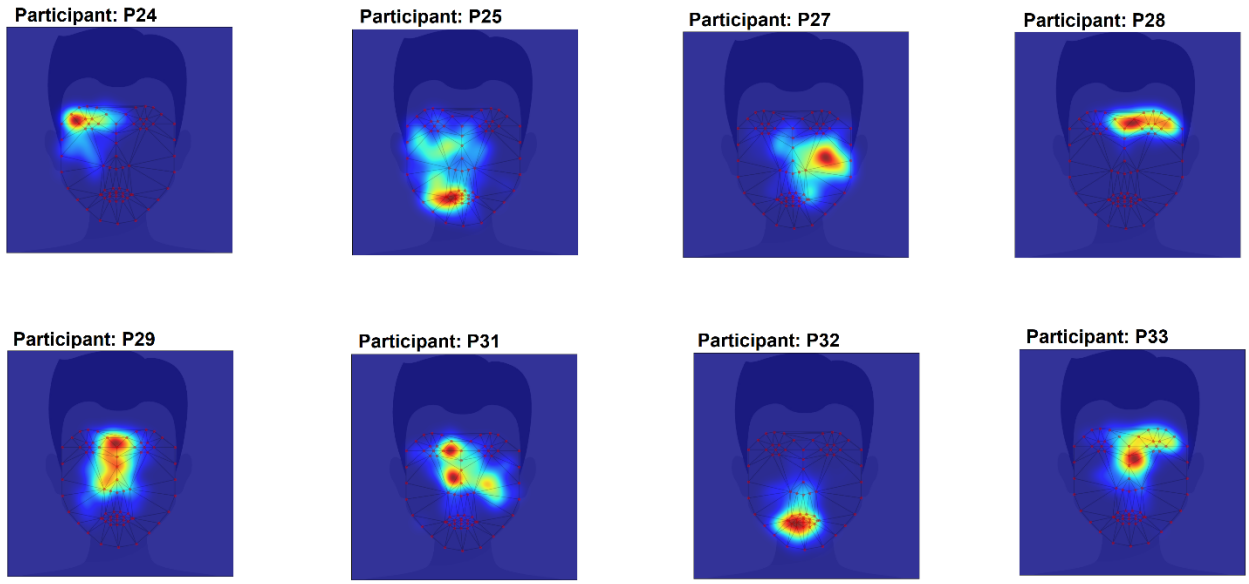


Figure S4. Individual participant data during face-to-face interaction task. We show the heatmap registration method results to where participants attended when viewing a face during the face-to-face interaction task.

Comparing automatic versus manual coding

To validate the use of OpenPose to detect the presence of a person in a video frame, we randomly sampled 560 frames from participants' navigation task recordings in which a fixation was recorded. Figure S6 shows an example of a frame image. Four naïve volunteers then manually count the number of people in each of the 560 frames. Figure S7 shows the significant positive correlation between the average manual coding values with the automatic values provided by OpenPose ($r(559)= 0.89$, $p<0.001$).

Figure S6. Example of a randomly selected video frame used to validate the OpenPose system.

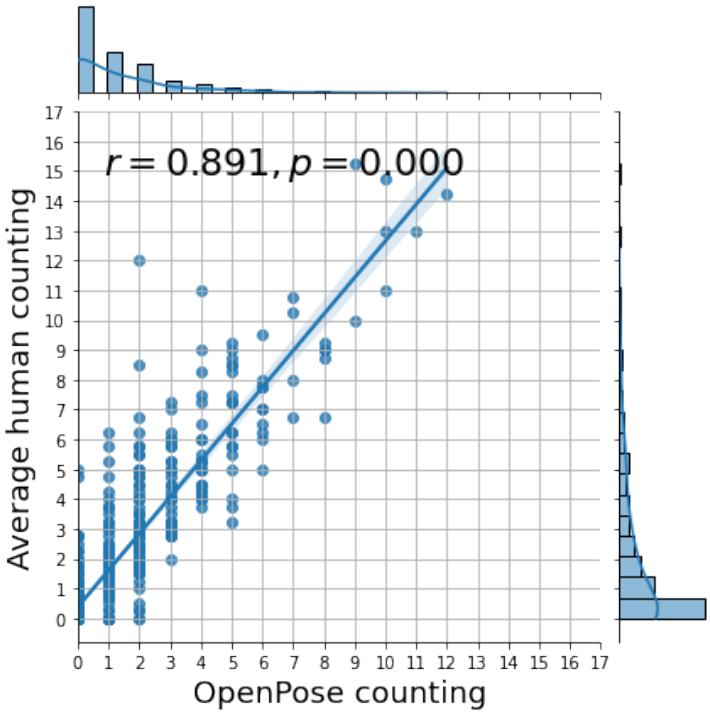


Figure S7. Correlation between human vs algorithm estimates of number of people in the scene

References

Cao, Z., Hidalgo, G., Simon, T., Wei, S. E., & Sheikh, Y. (2019). OpenPose: realtime multi-person 2D pose estimation using Part Affinity Fields. *IEEE transactions on pattern analysis and machine intelligence*, *43*(1), 172-186.