Supporting Information

Gallup et al. 10.1073/pnas.1116141109

SI Methods

First Experiment. ***Multiple analysis of covariance results with sex as a factor.*** For identifying the sex of each pedestrian, a sequence was extracted from the video by taking a 160 × 160-pixel image centered on the person's head in each frame. Because of both the positioning and resolution of our overhead camera, the identification was not completely precise, leaving 121 pedestrians unsexed. Therefore, this analysis included 2,701 pedestrians (1,326 male and 1,375 female).

Gaze-following response. Across all experiments, 27.5% (742/2,701) of passersby adopted the gaze direction of the stimulus group, and of these people, 14.3% (106/742) stopped to look. Stimulus group size was significantly associated with both the proportion of time spent looking and stopping to look [all looking: $F(6{,}2{,}700) = 36.647$, $P < 0.001$; stopping and looking: $F(6{,}2{,}700) = 10.021$, $P < 0.001$]. Posthoc tests with a Bonferroni correction show that the proportion of time spent gaze-following was lower for stimuli consisting of one and three members compared with all other group sizes (P values < 0.01). Groups of 12 and 15 members also drew a greater gaze-following response than groups of 5, 7, and 9 members (P values < 0.01). As for the proportion of time spent stopped and looking, posthoc analyses show that stimulus groups of 15 members drew a larger response than all other stimulus group sizes (P values < 0.05). No other comparisons were significant.

Pedestrian sex was also a significant factor, indicating that males spent a higher proportion of total time looking [0.084 vs. 0.066; $F(1{,}2{,}700) = 9.142$, $P < 0.01$] but not more time stopped while looking [0.010 vs. 0.006; $F(1{,}2{,}700) = 3.486$, $P = 0.062$]. There was also no difference in the proportion of pedestrians looking (male = 28.8%, female = 26.2%; $\chi^2 = 2.337$, $P = 0.126$). In addition, there was no interaction between stimulus group size and sex [all looking: $F(6{,}2{,}700) = 1.385$, $P = 0.217$; stopping and looking: $F(6{,}2{,}701) = 1.781$, $P = 0.099$].

The mean walking speed of pedestrians (i.e., not when stopped or gaze-following) was negatively associated with the proportion of time engaged in these behaviors [all looking: $F(1{,}2{,}700) = 26.542$, $P < 0.001$; stopped and looking: $F(1{,}2{,}700) = 48.767$, $P < 0.001$]. Similarly, crowd density was a significant negative predictor of the proportion of total time looking [$F(1{,}2{,}700) = 13.154$, $P < 0.001$] and the proportion of time stopped and looking [$F(1{,}2{,}700) = 4.592$, $P < 0.05$].

Second Experiment. ***Determining line of sight to the stimulus group.*** Eq. **S1** shows that a pedestrian was said to be directly facing the stimulus group members if the angle $A_{j,I}$ between the viewer's head direction and the vector pointing from the viewer to the observed person was less than a critical angle A_v. This critical angle depends on the size of the observed person (stimulus member) and their distance from the gazing pedestrian, D_v. We assume that a pedestrian looking within R_p m of the centroid of the stimulus group is observing those individuals (Eq. **S1**):

$$A_v = a\sin\left(R_p/D_v\right) \qquad \textbf{[S1]}$$

and (Eq. **S2**)

$$L\left(j,i\right) = 1, if\ A_{j,i} <= A_v, or\ 0\ if\ A_{j,i} > A_v. \qquad \textbf{[S2]}$$

Thus, a person farther away is less likely to be looked at than a closer person (Eq. **S2**).

To consider if a person at O had a line of sight to a person at C, we calculated the critical angle A_v with $R_p = 0.29$ m. This calculation defined a field of view $F_{OG} = [\theta - A_v, \theta + A_v]$ in which the O could see C, where θ is the angle from O to C (Fig. S4 and Eq. **S3**). It also defined the region. Any individual within one body radius of this region obstructs the field of view. OABCDE bounds the set of points x such that if there is a person at position x, modelled as a circle with radius Rp, a person at C will be (partially) occluded as seen by a person at point O. For every individual i within the region T = OABCDE, the critical angle A_{vi} to it is calculated with $R_p = 0.29$ m. This calculation defined another field of view $F_{Oi} = [\theta_i - A_{vi}, \theta_i + A_{vi}]$, where θ_i is the angle from O to i. Each F_{Oi} is subtracted from F_{OG} to determine the remaining (if any) field of view in which O has a line of sight to C (Eq. **S3**):

$$F'_{OG} = F_{OG} \setminus \bigcup_{i \in T} F_{Oi}. \qquad \textbf{[S3]}$$

If the field of view F'_{OG} is not empty, then we say that O has a line of sight to C.

Spatial effects in the commuter train station. For the experiments in the commuter station, our overhead camera was placed above and behind the announcement board for train arrivals/departures. Because of the tendency for pedestrians to frequently look up at this board during our experiments, we tested for differences in the visual attention directed to the stimulus group as a function of spatial positioning. We predicted that pedestrians in front of the stimulus group and therefore, closer to the announcement board would be less likely to look back at the stimulus members.

As expected, the proportion of pedestrians who looked at the stimulus was significantly lower for pedestrians with average spatial positions that were closer to the announcement board [19.6% vs. 55.7%; $\chi^2(1) = 26.905$, $P < 0.001$]. The proportion of time spent looking by pedestrians when average positions in front of the stimulus group and therefore, closer to the announcement board was also significantly lower compared with those pedestrians with average positions behind the stimulus group [0.037 vs. 0.185; $t(124.891) = 5.034$, $P < 0.001$]. However, there was no difference in the density of pedestrians in front or behind the stimulus group per experimental condition [$\chi^2(1) = 2.495$, $P > 0.05$]. There was also no sex difference in spatial positioning [$\chi^2(1) = 0.071$, $P > 0.05$].

Temporal cluster analysis. To identify significantly nonrandom clustering of gazes to the stimulus group, a runs test was performed for each individual experiment in both the experimental and control conditions within the shopping street and the commuter train station. Each experiment was broken into 20 3-s bins (although results are not dependent on binning category), and the number of pedestrians gazing to the stimulus group in each bin was calculated. A run was composed of consecutive bins identified as having either at least two pedestrians looking (one) or one or no pedestrians looking (zero). For instance, 1110000 contains two runs; the first run represents 9 s, during which time there were multiple pedestrians gazing in the direction of the stimulus group. The second run represents 12 s, where either only one or no pedestrians were looking. Thus, there was no distinction between bins with two or more pedestrians (one) or between one or no pedestrians (zero).

Negative Z scores indicate greater clumping rather than dispersion. Trials with a Z score of zero represent a case in which there is neither clumping or dispersion (i.e., random distribution) or a case where there were either (*i*) one or fewer pedestrians or (*ii*) two or more pedestrians looking the entire trial (all zeros or ones). Our results show that the suspicious condition in the train station is the only condition in which we can see significant clustering of directed gazes to the stimulus group (2/9 P values < 0.05; mean Z score ± SD = -0.729 ± 1.192). There was no evidence of temporal clustering in the shopping thoroughfare (mean Z score ± SD = 0.008 ± 0.033).

Analysis of variance results with sex as a factor. The identification of pedestrian sex was performed the same way as in the first experiment, leaving 16 pedestrians unsexed. Therefore, this analysis included 487 pedestrians (245 male and 242 female).

Pedestrians in the shopping street spent a lower proportion of time looking in the direction of the stimulus group [$F(1,487) = 19.583$, $P < 0.001$]. Distance was also a significant main factor [$F(1,487) = 14.702$, $P < 0.001$], indicating that pedestrians spent more time looking in the direction of the stimulus group the closer that they were to them. There was also a significant interaction between location and distance [$F(1,487) = 8.243$, $P < 0.001$], indicating that the proportion of time spent looking at the stimulus group increased for the suspicious condition within the shopping street, whereas the opposite was true within the commuter station. Although the proportion of time spent looking at the stimulus group did not differ between experimental conditions across trials [$F(1,487) = 0.027$, $P > 0.05$], there was a significant interaction between experimental condition and distance [$F(1,487) = 4.516$, $P < 0.05$], with both closer (<1.500 m) and farther distances (1.750–2.000 m) drawing a greater response in the train station (P values < 0.01).

There was no main effect for sex [$F(1,487) = 1.181$, $P = 0.278$], but there was an interaction between experimental condition and sex [$F(2,487) = 3.876$, $P < 0.05$]. Males spent a greater proportion of time looking in the direction of stimulus group members in the control condition, whereas females spent more time orienting to the suspicious behavior. However, this effect was context-dependent and present only within the train station. This effect was observed by a significant interaction between location, experimental condition, and sex [$F(2,487) = 3.887$, $P < 0.05$]. In the shopping street, both males and females spent a greater proportion of time looking at the suspicious-acting stimulus groups, and this finding was also true for females in control trials within the train station. However, males spent a lower proportion of time directing visual attention to the suspicious-acting stimulus groups in the train station (Fig. S9). That being said, there was no significant difference in the proportion of males and females looking at the suspicious stimulus in this location [36.4% male vs. 42.9% female; $\chi^2(1) = 0.421$, $P = 0.5162$].

Although there was only a marginally significant interaction between location, condition, and distance when including sex as a factor [$F(2,487) = 2.639$, $P = 0.072$], logistic regression analyses showed that pedestrians who looked in the direction of the stimulus group were closer to these members in both control conditions and the suspicious condition in the shopping thoroughfare (P values < 0.01); however, distance was not associated with looking behavior in the suspicious condition within the train station ($P > 0.05$).

Experimental Setup for the First Experiment. Experiments were conducted in a busy thoroughfare in Oxford, United Kingdom, and they occurred during 1000 and 1600 h. Crowds were filmed from above using a Canon XM2 miniDV camcorder (25 fsp) atop a roof of one of the buildings on the street (Fig. S1) positioned 14 m above the ground at an angle of 24° to the vertical. For simplicity, we also used this placed camera as the target of the gaze direction for the stimulus group members. A group of confederates (our stimulus crowd) entered the street from a hidden location in a small side street, and in the middle of our observation area, they stopped and looked up to the camera for 60 s. At the end of this period, the group was signaled subtly to disperse by remote alarm (one of the confederates wore a countdown timer on a wristwatch that had an audible alarm). After the stimulus group left, we waited 3 min before another stimulus group entered. We hired up to 15 people each day (both males and females), and stimulus group sizes were chosen in random order during 1-h blocks.

We tracked all pedestrians who entered the filming region during our trials, and we also recorded, for every frame, whether individuals were looking in the same direction as the stimulus group and how long the look occurred. This tracking was done manually using point and click software that we developed in Matlab. Although this procedure was very time-consuming, we found it to be the only viable method that ensured accurate detection of gaze-following. We calculated each pedestrian's position, speed, and tortuosity of path, and we determined when, where, and for how long they copied the gaze of the stimulus group (i.e., looked up to the camera) (Fig. S1). We categorized an individual as stopped if his/her speed fell below 0.2 ms^{-1}. We also discretized the region of filming to allow us to analyze the spatial relationship between crowd density, flow patterns, and where individuals tended to follow the gaze direction of the stimulus crowd.

Experimental Setup for the Second Experiment. Experiments were carried out in the same shopping street (Oxford) and a city commuter train station in London. In Oxford, all experiments occurred during the hours of 1000 and 1400 h, and like in the first experiment, each experiment was spaced by at least 3 min. In London, experiments occurred between 1000 and 1500 h, allowing at least an 11-min break between each one to ensure that a new population of passengers was within the concourse. The experiments were filmed using an overhead HD video camera (25 fps, HG21 HD; Canon), and from this film, we obtained accurate positions and head poses of all of the pedestrians in the scene (Fig. 4). In the shopping street, this camera was positioned 14 m above the ground at an angle of 24° to the vertical. In the train station, this camera was 9 m above the ground at an angle of 45° to the vertical. The average total filming region for each location was 213 and 290 m^2, respectively. Similar to the first experiment, each experiment lasted for 60 s, and treatments were distributed randomly within 1-h blocks. We used between seven and nine people each day (all male) who were randomly assigned to act in our stimulus groups.

As for the commuter station, permissions were obtained from the owners, who asked that we brief the Transport Police and the security staff that patrols the station. Local authorities were contacted and provided full disclosure regarding the details and rationale behind these experiments. However, there were no cases in which a pedestrian either approached or tried to communicate with the stimulus group, members of the research staff, or the local authorities during any point during or after our experiments.

For each pedestrian, a video sequence was extracted from the video by taking a 160 × 160-pixel image centered on the person's head in each frame, and the head pose of each person was recorded manually to depict the direction that his or her face was pointing in each frame. An operator then played back these images and overlaid an arrow starting at the crown of the head and pointing in the direction that the face was pointing (Fig. 4). Therefore, for each person, an accurate head position and direction was obtained in every frame for each person. The majority of the labeling was distributed to workers over the internet using Amazon Mechanical Turk, which is an online marketplace where businesses, developers, etc., can post human intelligence tasks for pay (https://www.mturk.com/mturk/welcome). In no cases were the identities of the pedestrians or the context of the environment provided. For the Oxford data, a trained researcher entered the data for 1% of all images, and these preentered labels were provided to ensure that users on Amazon Mechanical Turk entered data to a required standard. For the London data, a trained researcher visually inspected ∼10% of the labeling returned for each tracked person and then either approved or rejected it. The labeling of the head position was required to be within 14 pixels, and the head pose angle was required to be within 29°.

Tracking. Because of the facts that our cameras were not directly above the crowd and that the lens has inherent curvature, the tracked positions were initially distorted. We, therefore, needed to determine the internal camera geometric and optical characteristics for each experiment to remap pixels in the image to their corresponding real-world positions. We used the four-step method in ref. 1, which involves replacing the physical parameters with nonphysical implicit parameters that are used to interpolate between known tie points in the scene. To do this task, we take into consideration the focal length, skew, and distortions to correct our geometry and recalibrate trajectories, and therefore, they appear as if we filmed from directly above the crowd. An overview of this technique and links to useful technical papers on this topic can be found online (2). The tracking of pedestrians was performed by estimating, for each frame, the point on the ground directly below the body (usually between the feet). Camera calibration techniques enabled the determination of the height of the individual, which was then used to project the position of the head onto the ground plane. Calibrated trajectories were analyzed in Matlab using our own custom-written code.

1. Heikkila J, Silven O (1997) A four-step camera calibration procedure with implicit image correction. *Proc CVPR IEEE* 10:1106–1112.

2. Bouguet JY (2007) *Camera Calibration Toolbox for Matlab.* Available at http://www.vision.caltech.edu/bouguetj/calib_doc/. Accessed April 2, 2007.

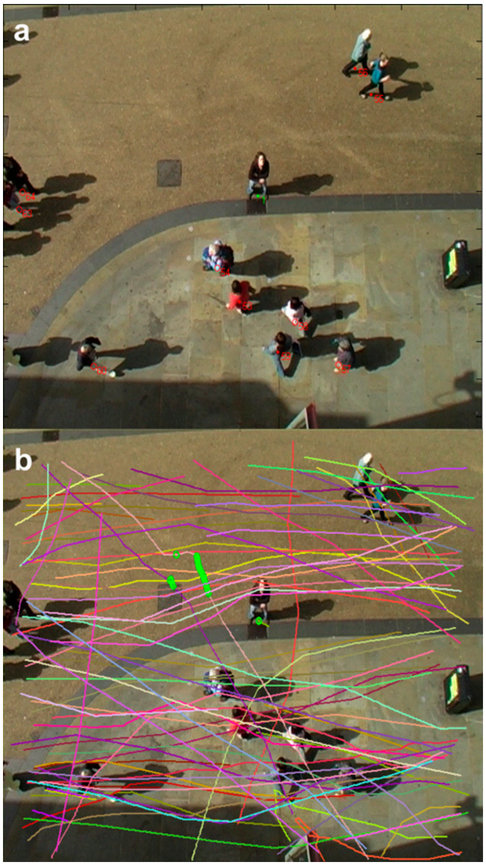

Fig. S1. Tracking the looks and trajectory of passersby. (*A*) Individual pedestrians are each assigned unique but anonymous identities. (*B*) A 1-min sequence of tracks showing the motions of pedestrians as colored lines projected onto the ground beneath their feet and sections of trajectories where the individual was estimated to have copied the stimulus gaze shown in bold green. Note that a stimulus individual is in the center of the frame.

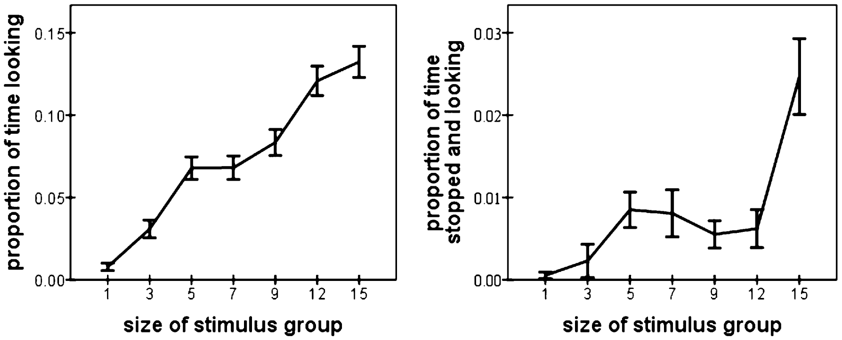

Fig. S2. The proportion of time spent looking up (*Left*) and time stopped while looking up (*Right*). Both were positively associated with the size of the stimulus group (mean ± SE shown).

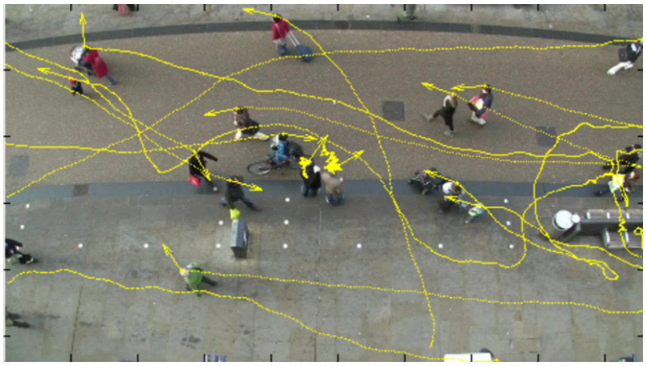

Fig. S3. Tracking the visual orientation and trajectory of passersby. A partial sequence of tracks showing the trajectories and estimated gaze direction of pedestrians in the shopping street. Note that a stimulus group is in the center of the frame.

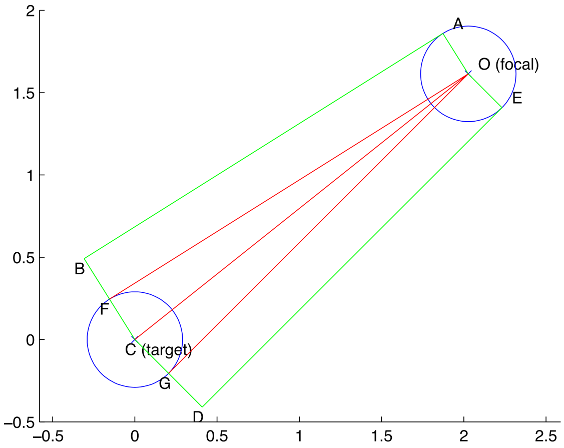

Fig. S4. The regions of interest for considering whether a person at O has a line of sight to a person at C. The red lines indicate the region corresponding to the field of that O could see C. The angle COF = COG is the critical angle A_v. The people in the region 1 body radius around that field of view will partially (or fully) obstruct O's view of C (this region is represented by the green lines).

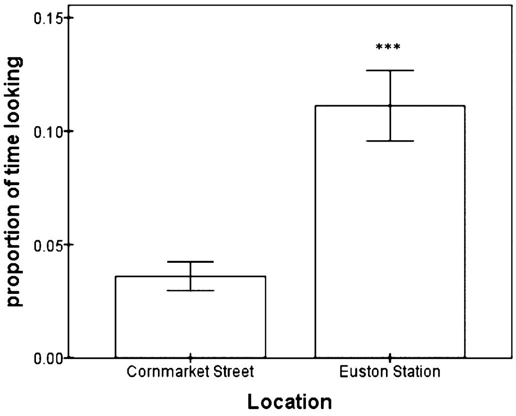

Fig. S5. Main effect for location. Pedestrians in the commuter station spent a larger proportion of time looking in the direction of the stimulus group members compared with those pedestrians walking in the shopping street (***$P < 0.001$; mean ± SE shown).

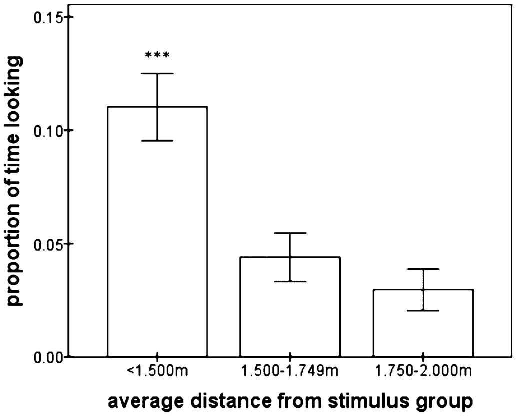

Fig. S6. Main effect for distance. As expected, pedestrians at closer distances (<1.500 m) spent a larger proportion of time orienting their attention to the stimulus group members compared with both the intermediate and farthest distance categories (***$P < 0.001$). There was no difference in this response between the 1.500–1.749 and 1.750–2.000-m distances (mean ± SE shown).

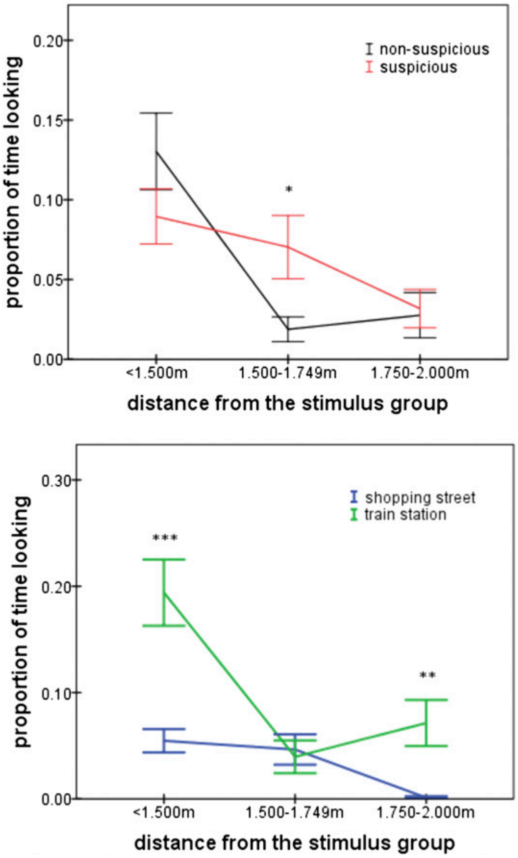

Fig. S7. Interactions between condition and distance and location and distance. There was a significant interaction in the response to look in the direction of the stimulus group members between experimental condition and average distance from the centroid of the group (*Upper*). There was a gradual increase in the proportion of visual attention directed to these members with decreasing distance in the suspicious condition, whereas this response was more pronounced for the shorter distance within the shopping street. This finding resulted in a significant difference at the intermediate distance (1.500–1.749 m; $^*P < 0.05$). There was also a significant interaction in the response to look in the direction of the stimulus group members between location site and average distance from the centroid of the group (*Lower*). Both closer (<1.500 m) and farther distances (1.750–2.000 m) drew a greater response in the train station compared with the shopping street ($^{**}P < 0.01$, $^{***}P < 0.001$; mean ± SE shown).

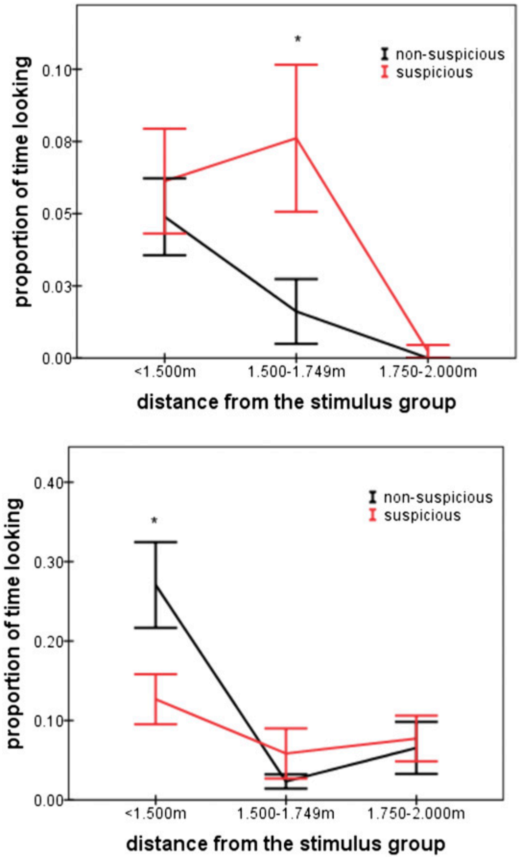

Fig. S8. Three-way interaction of main effects. The proportion of time spent looking at the stimulus group members as a function distance and location (mean ± SE shown). In the shopping street (*Upper*), the experimental condition drew more attention from passersby at intermediate distances (1.500–1.749 m). In the train station (*Lower*), the stimulus group in the control condition drew more attention than the experimental group at close distances (<1.500 m; $^*P < 0.05$).

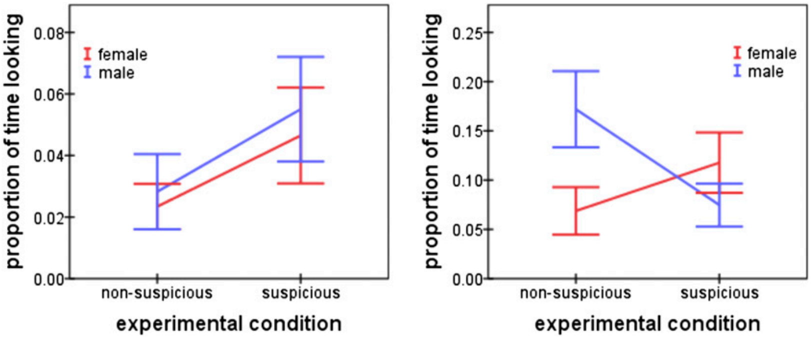

Fig. S9. Interaction between location, condition, and sex. The proportion of gaze that was directed to the stimulus group in each location as a function of the sex of the pedestrians (mean ± SE shown). There was a significant interaction between location, condition, and pedestrian sex ($P < 0.001$). In the shopping street (*A*), both males and females showed an increase in the proportion of time looking at the stimulus group in the suspicious trials. In the train station (*B*), however, females showed an increase in the time spent looking at the stimulus group in the suspicious-acting condition, whereas the opposite was true for males.