



Other people's gaze encoded as implied motion in the human brain

Arvid Guterstam^{a,1} , Andrew I. Wilterson^a , Davis Wachtell^a , and Michael S. A. Graziano^a

^aDepartment of Psychology, Princeton University, Princeton, NJ 08544

Edited by Ranulfo Romo, National Autonomous University of Mexico, Mexico City, D.F., Mexico, and approved April 23, 2020 (received for review February 18, 2020)

Keeping track of other people's gaze is an essential task in social cognition and key for successfully reading other people's intentions and beliefs (theory of mind). Recent behavioral evidence suggests that we construct an implicit model of other people's gaze, which may incorporate physically incoherent attributes such as a construct of force-carrying beams that emanate from the eyes. Here, we used functional magnetic resonance imaging and multivoxel pattern analysis to test the prediction that the brain encodes gaze as implied motion streaming from an agent toward a gazed-upon object. We found that a classifier, trained to discriminate the direction of visual motion, significantly decoded the gaze direction in static images depicting a sighted face, but not a blindfolded one, from brain activity patterns in the human motion-sensitive middle temporal complex (MT+) and temporo-parietal junction (TPJ). Our results demonstrate a link between the visual motion system and social brain mechanisms, in which the TPJ, a key node in theory of mind, works in concert with MT+ to encode gaze as implied motion. This model may be a fundamental aspect of social cognition that allows us to efficiently connect agents with the objects of their attention. It is as if the brain draws a quick visual sketch with moving arrows to help keep track of who is attending to what. This implicit, fluid-flow model of other people's gaze may help explain culturally universal myths about the mind as an energy-like, flowing essence.

gaze | social cognition | motion perception | visual attention | theory of mind

Recent behavioral studies suggest that the brain, beyond registering low-level visual cues about the direction of other people's gaze (1–3), constructs a model of other people's active visual attention (4–7). This model may be simplified and schematic, involving the attribution of beams that emanate from the eyes toward the object of attention (5). The model is constructed at an implicit level—people are generally not aware they are doing it. The adaptive benefit for the brain to model other people's attentive gaze in such a schematic manner may be related to computational efficiency in processing complex social stimuli with multiple sources and targets of visual attention (4–6, 8). In the present experiment, we used functional magnetic resonance imaging (fMRI) to study the brain regions involved when people process actual motion and when people process the gaze of others. We hypothesized that processing gaze partly engages brain regions that process motion in a direction-specific manner. To test our hypothesis, we used multivoxel pattern analysis to decode the gaze direction in static images of faces looking at objects, with a classifier trained on discriminating the direction of actual visual motion. In addition to a whole-brain search, we focused on two regions of interest (ROIs): the motion-sensitive middle temporal complex (MT+) and the temporo-parietal junction (TPJ). The MT+ is a subregion of the extrastriate visual cortex specialized for visual motion perception (9, 10). The TPJ is consistently activated in tasks requiring theory of mind (11, 12). Moreover, in at least some experiments, modeling the active attention of others was associated with activity in the TPJ (7, 13). The TPJ also overlaps the caudal superior temporal sulcus (STS), which, in both humans and monkeys, is active in

association with processing visual cues about the gaze direction of others (14–18). We therefore hypothesized that these two regions in particular, the TPJ and MT+, would be involved in the present experiment.

Results

In an fMRI experiment involving 32 healthy human subjects (18 females; mean age, 26 y; range 18 to 52 y; see *Materials and Methods* for details), we used a slow, event-related design to estimate the brain activity associated with either viewing actual visual motion or viewing another person gazing at an object. As shown in Fig. 1A, the visual motion stimulus consisted of a square-shaped (5° × 5°) random dot motion display (19) where 100% of the dots moved coherently in the same direction, creating a strong sense of either leftward (dot motion left condition) or rightward motion (dot motion right condition). In the gaze trials, subjects observed a static image of a cartoon face on one side of the screen, gazing at an object (a tree) on the other side of the screen. The face and the tree were spatially aligned such that the empty space in between, where the hypothesized implied motion should occur, corresponded to the size and location of the random dot motion field (5° wide). The face appeared either on the right side gazing leftward (eyes open left condition), or on the left side gazing rightward (eyes open right condition). In two control conditions (eyes covered left and eyes covered right), the eyes of the face were covered by a blindfold, keeping all other aspects of the stimulus identical. We reasoned that the blindfold should prevent any gaze-induced implied motion in the experiment.

To make sure that subjects followed instructions and paid attention to the visual stimuli, we also included catch trials in

Significance

This study used fMRI scans of people to show that the brain processes the gaze of others, and visual flow, in a similar manner. Visual motion brain areas and social cognition areas were involved. It is as if the brain draws a quick visual sketch with moving arrows to help keep track of who is attending to what. We propose that this implicit, fluid-flow model of other people's gaze may help explain culturally universal myths about the mind as an energy-like, flowing essence.

Author contributions: A.G. and M.S.A.G. designed research; A.G., A.I.W., and D.W. performed research; A.G. and A.I.W. analyzed data; and A.G., A.I.W., D.W., and M.S.A.G. wrote the paper.

The authors declare no competing interest.

This article is a PNAS Direct Submission.

This open access article is distributed under [Creative Commons Attribution-NonCommercial-NoDerivatives License 4.0 \(CC BY-NC-ND\)](https://creativecommons.org/licenses/by-nc-nd/4.0/).

Data deposition: The data that support the findings of this study are available at https://figshare.com/articles/Other_People_s_Gaze_Encoded_as_Implied_Motion_in_the_Human_Brain/12184848/1.

¹To whom correspondence may be addressed. Email: arvidg@princeton.edu.

This article contains supporting information online at <https://www.pnas.org/lookup/suppl/doi:10.1073/pnas.2003110117/-DCSupplemental>.

First published May 26, 2020.

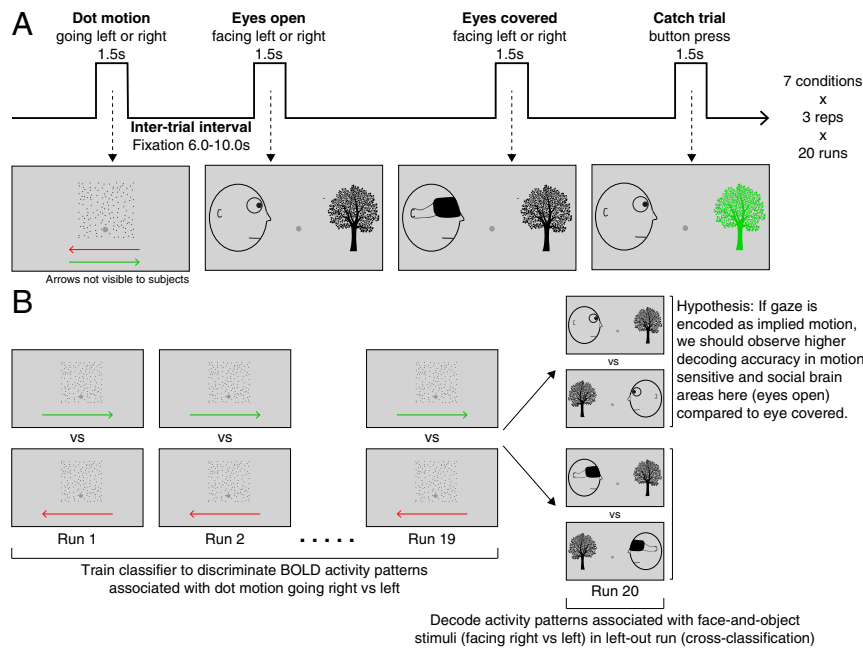


Fig. 1. Methods. (A) Schematic time line of the fMRI design. While subjects continuously fixated on a central spot, they were exposed to 1.5-s-long trials of either a random dot motion stimulus (going left or right), or a static image of a face gazing at a tree (facing left or right), or an image of a blindfolded face (facing left or right). In a catch trial condition, one of the image elements (head or tree) or the moving dots appeared bright green, in response to which the subjects pressed a button. Arrows shown here indicate dot motion directions and were not part of the actual stimuli. There were equal numbers of rightward and leftward facing face-and-tree trials, but only rightward facing images are shown here. (B) To test our hypothesis that gaze is encoded as implied motion in motion-sensitive and social brain areas, we used a locally multivariate (Searchlight), leave-one-run-out, cross-classification approach, using the runwise regression (beta) coefficients as model input. We trained a classifier to discriminate the BOLD activity patterns associated with visual motion going left versus right in 19 runs and then tested whether it could decode activity patterns associated with gaze direction (eyes open facing left versus right) significantly better than in the blindfolded condition (eyes covered facing left versus right) in the left-out twentieth run, repeated for all runs.

which either the moving dots, the face, or the tree appeared bright green, in response to which subjects pressed a button using their right index finger. Subjects detected these targets on a mean of 98% of the catch trials. All seven trial types were of 1.5 s duration and presented in a randomized order with a jittered intertrial interval of 6.0 to 10.0 s, divided into 20 runs featuring three repetitions per condition per run. To prevent systematic differences in eye movements across conditions, subjects were instructed to fixate on a central fixation spot throughout the experiment, and their eye movements were recorded using an MRI-compatible infrared eye tracker (*Materials and Methods*). The results of the eye tracking showed that subjects stayed on fixation on average 97% of the time and that fixation and saccade data alone were not sufficient to successfully decode the experimental conditions of interest (see *SI Appendix, Fig. S1* for details), suggesting that differences in eye movement dynamics can be excluded as a confound in any decoding results based on the fMRI data.

To test our hypothesis that other people's gaze is encoded as implied motion in MT+ and TPJ, or any other brain region outside our ROIs, we used a whole-brain, locally multivariate (Searchlight) (20), leave-one-run-out, cross-classification approach. We first employed a conventional general linear model (GLM) to estimate regression (beta) coefficients for the six main conditions in each run (one additional regressor of no interest modeled all of the catch trials) and then submitted these runwise beta coefficients to multivariate analyses (21, 22). We trained a support vector machine (SVM) classifier to discriminate dot motion left versus dot motion right in 19 of 20 runs and then tested whether it could decode gaze direction (i.e., eyes open left versus eyes open right) in the left-out 20th run (so-called cross-classification) (Fig. 1B), based on the blood oxygen level-dependent (BOLD) response pattern within 12-mm-radius

Searchlight spheres centered on each voxel in the entire brain. This procedure was repeated 20 times so that each run was left-out once, and a run-average decoding accuracy was calculated. The same classifier was also tested on the blindfolded control condition (eyes covered left versus eyes covered right). The key analysis was the comparison in cross-classification decoding accuracy in the eyes open versus eyes covered conditions because any area revealed by this contrast must contain brain activity patterns that are driven by low-level visual motion and specifically decode gaze direction, but only when the eyes of the face have unobscured vision. To make sure this contrast did not identify any area in which the performance of the classifier in the eyes open conditions was below the level of chance (50%), we only searched for voxels in which eyes open left versus eyes open right was decoded significantly ($P < 0.05$, uncorrected) better than chance. To define our ROIs, we delineated the MT+ bilaterally using a visual motion localizer that we ran on each subject, based on previously published localizer tasks (23–25), consisting of two 5-min runs of viewing moving or static dots (see *Materials and Methods* and *SI Appendix, Fig. S2* for details). The TPJ ROIs were defined as 10-mm-radius spheres centered on the left and right TPJ activation peaks in a previous landmark fMRI study using a theory of mind task (11). In our analysis, we corrected for multiple comparisons both within our ROIs (small-volume corrections), as well as at the whole-brain level, to reveal any potential significant decoding activity in the rest of the brain.

Fig. 2 and Table 1 show the results. A classifier, trained on discriminating the direction of low-level visual motion, could decode the gaze direction in the eyes open condition significantly better than in the eyes covered condition in the right MT+ (decoding accuracy difference: 53.5% versus 49.7%, $t = 3.64$, $P = 0.027$, small-volume corrected), right posterior STS (belonging to

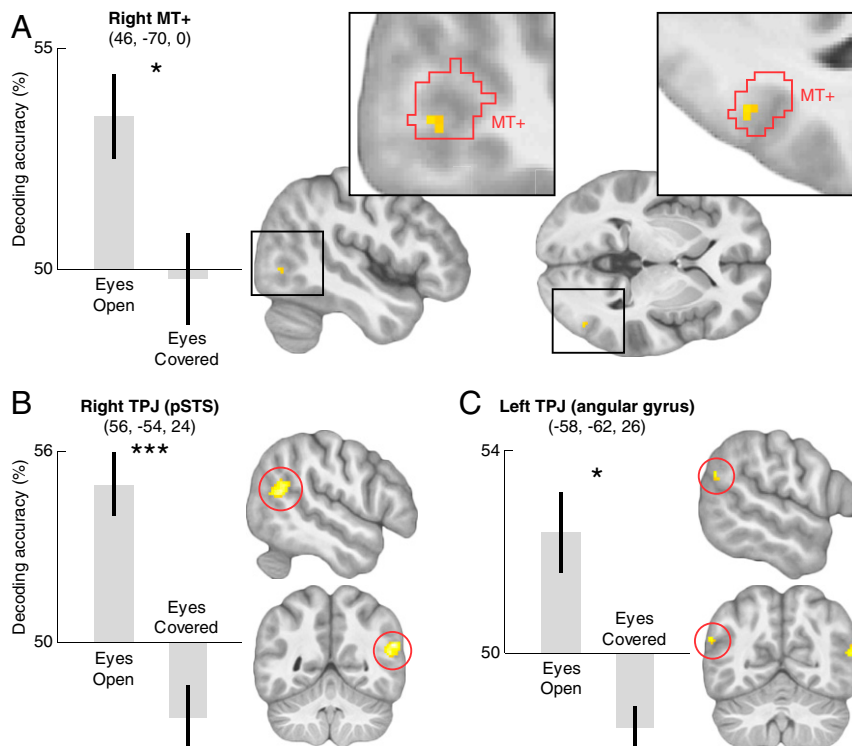


Fig. 2. Results. Brain areas in which a classifier, trained on discriminating the direction of dot motion, significantly decoded the direction of gaze in static images of a face looking at an object (eyes open), using a blindfolded face as control (eyes covered). These results suggest that gaze is encoded as implied motion, in a specific direction, in the motion-sensitive middle temporal cortical complex (MT+, outlined in red) on the right side (A), and in the temporoparietal junction (TPJ, red circles) bilaterally (B and C). Errors bars show SE, significance shown by * $P < 0.05$ and *** $P < 0.001$, corrected for multiple comparisons. See text for statistical details. The decoding maps are thresholded at $P < 0.001$ (uncorrected), for visualization purposes. pSTS, posterior superior temporal sulcus.

the TPJ) (decoding accuracy difference: 55.0% versus 47.6%, $t = 5.82$, $P < 0.001$, small-volume corrected), and left angular gyrus (within the TPJ) (decoding accuracy difference: 52.4% versus 48.5%, $t = 4.03$, $P = 0.019$, small-volume corrected). No significant voxels were found in the MT+ ROI in the left hemisphere. As shown in Table 1, by far the largest cluster of voxels and strongest decoding peak in the brain was found within the pre-defined ROI in the right TPJ, which also survived correction for multiple comparisons at the whole-brain level ($P = 0.040$). These findings suggest that a set of areas involving the MT+ and TPJ, primarily on the right side, encode the gaze of social agents as implied motion flowing from the eyes across the empty space to the gazed-at object. In addition to the peaks found within the ROIs, we also found decoding peaks ($P < 0.001$, uncorrected), albeit not whole-brain significant after correcting for multiple comparisons, in the right fusiform gyrus, left midinsula, left putamen, and the right ventral striatum (Table 1 and *SI Appendix*, Fig. S3).

Discussion

These results strongly suggest that, when people view a face looking at an object, the brain treats that gaze as though a movement were present, passing from the face to the object. That movement encoding was observed in area MT+, known to be involved in visual motion processing, and in the TPJ, known to be involved in social cognition. Gaze is arguably the most relevant cue to the state of someone else's visual attention, and having an efficient neural machinery for keeping track of gaze is thus essential for reading and predicting other people's minds and behavior (14, 26, 27). The present findings demonstrate that this process involves more than simple registration of low-level visual cues about other people's eyes. Specific regions of the

human brain appear to encode other people's gaze as an active motion streaming through the empty space from the agent to the gazed-upon object. These findings are consistent with previous behavioral work showing that people implicitly treat other people's eyes as though they emanated a weak force, gently "pushing" on objects in the external world (5, 6). We propose that this implicit, fluid-flow model of other people's gaze may help keep track of visual attention in a complex social environment. The model may be a part of theory of mind, modeling another mind actively focusing on content, a precondition for reconstructing that mind's intentions, beliefs, emotions, and so on (4, 11, 26, 28, 29). It is well-known that, during the course of evolution, it is not uncommon that ancient biological mechanisms are reused in a different role, a phenomenon called "exaptation" (30). We speculate that the visual motion system may have been used during the evolution of social brain mechanisms for tracking the attention of others. It may have simply proved adaptive to coopt the brain's motion system to keep track of sources and targets of visual attention. It is as if the brain draws a quick visual sketch with moving arrows to help keep track of who is attending to what.

The MT+ and TPJ are well-situated for constructing a simplified, fluid-flow model of other people's gaze. The MT+ is highly specialized in visual motion perception (31) and is activated when people view static images featuring conventional implied motion stimuli (e.g., a running animal) (32). The TPJ has been implicated in a range of social cognition tasks (33, 34), is a key node in the theory of mind network (11, 12), and has been implicated in processing the awareness or attention states of others (7, 13). Our finding that a classifier trained to discriminate the direction of actual visual motion successfully generalizes to decode gaze direction in independent left-out data, and that this

Table 1. Decoding results

| Anatomical region | MNI x, y, z | Peak T | P value (FWE-corr) | Cluster size |
|---|--------------|--------|--------------------|--------------|
| Temporal lobe | | | | |
| R. posterior STS (TPJ) | 56, -54, 24 | 5.82 | <0.001 | 201 |
| R. parieto-temporo-occipital cortex (MT+) | 46, -70, 0 | 3.64 | 0.027 | 4 |
| R. fusiform gyrus | 22, -46, -16 | 3.62 | — | 11 |
| R. superior temporal gyrus | 72, -28, 6 | 3.71 | — | 5 |
| Parietal lobe | | | | |
| R. supramarginal gyrus | 56, -40, 36 | 4.18 | — | 14 |
| R. supramarginal gyrus | 58, -38, 44 | 3.64 | — | 4 |
| L. angular gyrus (TPJ) | -58, -62, 26 | 4.03 | 0.019 | 10 |
| Frontal lobe | | | | |
| R. precentral gyrus (premotor cortex) | 58, -4, 42 | 4.35 | — | 32 |
| L. inferior frontal gyrus | -46, 10, 20 | 3.62 | — | 7 |
| Insular cortex | | | | |
| L. midinsula | -36, -2, 16 | 4.45 | — | 20 |
| Subcortical structures | | | | |
| R. ventral striatum | 10, 4, -10 | 4.79 | — | 25 |
| L. putamen | -20, 6, 4 | 3.90 | — | 13 |

All brain regions (peaks) in which a classifier, trained on discriminating dot motion direction, decoded gaze direction at a threshold of $P < 0.001$, uncorrected for multiple comparisons, better in the eyes open than in the eyes covered condition. All listed regions also decoded gaze direction in the eyes open condition significantly ($P < 0.05$, uncorrected) better than chance (50%). FWE rate-corrected (corr) P values are reported for regions that survived the correction for multiple comparisons in our predefined ROIs (small-volume correction), consisting of the activation cluster from the MT+ visual motion localizer (*SI Appendix, Fig. S2 and Table S1*), or 10-mm-radius spheres around the TPJ activation peaks in a previous fMRI study on theory of mind (11). The right TPJ peak in the posterior STS also survived correction for multiple comparisons using the whole brain as search space ($P = 0.040$). L., left; R., right.

decoding performance is found in the hypothesized brain regions MT+ and TPJ, supports our proposal. Although we cannot exclude the possibility that the observed MT+ and TPJ decoding reflects the participants anticipating actions of the agent, such as reaching, rather than encoding the agent's gaze, we consider this alternative explanation less likely. First, the decoding is significantly stronger in the sighted than in the blind agent. Second, as a target of gaze, we chose a tree rather than a more commonly grasped item such as a cup. Third, prior behavioral experiments using similar stimuli suggested that there is a flow-field effect of the eyes linked specifically to participants reconstructing an agent's attention (5, 6). The results therefore provide evidence that the visual motion system is used to facilitate social brain mechanism for tracking the gaze of others.

The recruitment of motion systems in social cognition, representing gaze as implied motion emanating from the eyes, might help explain several culturally universal folk beliefs. These beliefs include the extramission myth that vision involves energy flowing out of the eyes (5, 35–39) and beliefs in the mind as a kind of energy that can flow out of the body and affect external objects. It is possible that basic theory-of-mind mechanisms have provided people with highly inaccurate intuitions and biases about the properties of the mind, leading to common myths and folk beliefs that have been intuitively compelling to humans across cultures and time periods.

Materials and Methods

Participants. Thirty-two human volunteers (18 females, 28 righthanded), aged 18 to 52 y (mean age, 26 y; SD = 8) participated in the study. Subjects were recruited either from a paid subject pool, receiving 50 USD for participation, or among Princeton undergraduate students, who received course credits as compensation. All subjects provided informed consent, and all procedures were approved by the Princeton Institutional Review Board.

Experimental Setup. Before commencing the scanning session, subjects were shown six sample trials on a laptop computer screen and given the instructions related to fixation and button press responses. During scanning, the subjects lay comfortably in a supine position on the MRI bed. Through an angled mirror mounted on top of the head coil, they viewed a translucent

screen (56 × 30 cm) positioned ~80 cm from the eyes (viewable area: 39° × 21°), on which visual stimuli were projected from a Hyperion MRI Digital Projection System (Psychology Software Tools, Sharpsburg, PA) with a resolution of 1,920 × 1,080 pixels. A computer running MATLAB (MathWorks, Natick, MA) and the Psychophysics Toolbox (40) was used to present the visual stimuli. A right hand five-button response unit (Psychology Software Tools Celeritas, Sharpsburg, PA) was strapped onto the subjects' right wrist, and they used the right index finger button to indicate responses during the catch trials.

Experimental Conditions and Visual Stimuli. The experiment comprised seven conditions, featuring either random dot motion or static images of a face and a neutral object (a tree). The motion stimulus consisted of a 5°-wide × 5°-high field of randomly distributed, short-lived, moving black dots. Each dot had a diameter of 0.05°, a velocity of 2°/s, and a lifetime of 200 ms. The density was 50 dots per square visual degree. One hundred percent of the dots moved coherently in the same direction, creating an unambiguous sense of motion to the left (dot motion left condition) or right (dot motion right condition). The stimulus was presented for 1,500 ms on each trial.

In the trials featuring static images, participants saw a cartoon face in profile on one side of the display, facing a tree on the other side. In the eyes open conditions, the eyes on the face were open, implying that the head was gazing at the tree. The face was 5.2° wide × 5.7° high, and the tree was 4.5° wide × 5.7° high. The face and the tree appeared on opposite sides of the fixation point, and the edge of each image was distanced 2.5° from the midline (thus, the total distance between the tree and the face was 5°). The face appeared either on the right side looking left (eyes open left condition) or on the left side looking right (eyes open right condition). In the two control conditions, eyes covered left and eyes covered right, the eyes of the cartoon face were covered with a blindfold, while all other visual features were kept identical. Again, the stimulus was presented for 1,500 ms on each trial.

The purpose of the catch trials was to ensure that the subjects paid attention to the visual stimuli. On each catch trial, some part of the visual stimulus was colored bright green, and subjects were instructed to press a button as soon as they discovered that something was green. The catch trials consisted of dot motion featuring bright green dots (instead of black ones), or the static face-and-tree image where either the face (eyes open or blindfolded) or the tree was colored green. Within each run, there was exactly one trial of each of the three (dots, face, object) catch trial types.

Throughout each run—in all seven conditions, as well as during the intertrial intervals—a light gray fixation point (0.5° diameter) of a shade (red [R], 198; green [G], 198; blue [B], 198) just slightly darker than the gray background (R, 210; G, 210; B, 210) was positioned slightly below the center of the screen (note that the fixation point shown in Fig. 1 is darker than the one used in the stimulus, for visualization purposes). The purpose of using a subtle gray color and placing the fixation below the line of sight of the cartoon face was to avoid the visual impression that the face was gazing at the fixation point instead of at the tree. Subjects were instructed to maintain fixation as continuously as possible throughout the run.

The experiment consisted of 20 runs. In each run, the seven conditions were repeated three times, yielding a total of 21 trials per run. The trial order was fully randomized, with the limitation that two consecutive trials could not belong to the same condition. We included 10 s of baseline before the onset of the first trial, and 16 s of baseline after the offset of the last trial, in each run. The run-average intertrial interval (ITI) was 8.0 s (individual ITIs were jittered between 6.0 and 10.0 s), yielding a total run duration of 3 min 38 s.

fMRI Acquisition. Functional imaging data were collected using a Siemens Prisma 3T scanner equipped with a 64-channel head coil. Gradient-echo T2*-weighted echo-planar images (EPIs) with BOLD contrast were used as an index of brain activity (41). Functional image volumes were composed of 54 near-axial slices with a thickness of 2.5 mm (with no interslice gap), which ensured that the entire brain, excluding cerebellum, was within the field-of-view in all subjects (54 × 78 matrix, 2.5 mm × 2.5 mm in-plane resolution, echo time [TE] = 30 ms, flip angle = 80°). Simultaneous multislice (SMS) imaging was used (SMS factor = 2). One complete volume was collected every 2 s (repetition time [TR] = 2,000 ms). A total of 2,180 functional volumes were collected for each participant in the main experiment, divided into 20 runs (109 volumes per run). In the subsequent area MT+ functional localizer experiment, 316 functional volumes were collected (158 volumes × two runs). The first three volumes of each run were discarded to account for non-steady-state magnetization. A high-resolution structural image was acquired for each participant at the end of the experiment (three dimensional magnetization prepared rapid gradient echo sequence, voxel size = 1 mm isotropic, field of view = 256 mm, 176 slices, TR = 2,300 ms, TE = 2.96 ms, inversion time = 1,000 ms, flip angle = 9°, integrated parallel acquisition technique generalized autocalibrating partially parallel acquisition = 2). At the end of each scanning session, matching spin echo EPI pairs (anterior-to-posterior and posterior-to-anterior) were acquired for blip-up/blip-down field map correction.

fMRI Preprocessing. Preprocessing was carried out using the FMRIprep version v1.2.3 (42) pipeline. See *SI Appendix, Supplementary Text* for details.

fMRI Analysis. The fMRI data from all participants were analyzed with the Statistical Parametric Mapping software (SPM12) (Wellcome Department of Cognitive Neurology, London, UK) (43). We first used a conventional GLM to estimate regression (beta) coefficients for the six main conditions in each run: dot motion left, dot motion right, eyes open (facing) left, eyes open right, eyes covered left, and eyes covered right. One regressor of no interest was created to model all of the catch trials. Each condition was modeled with a boxcar function of duration 1.5 s and convolved with the standard SPM12 hemodynamic response function. The runwise beta coefficients for the six main conditions were then submitted to subsequent multivariate analyses (21, 22).

Within each participant, we used locally multivariate mapping (the Searchlight approach) (20) to identify multivoxel patterns. The analysis was carried out in The Decoding Toolbox (TDT) version 3.997 (44) for SPM. The brain was partitioned into overlapping voxel clusters, each of which was approximately spherical in shape with a radius of 12 mm (default value in TDT). In each of these clusters, we used linear SVMs (with the fixed regularization parameter of $C = 1$) to compute decoding accuracies. We used a cross-classification approach where an SVM was trained to discriminate dot motion left versus dot motion right and then tested on either the eyes open (left versus right) or eyes covered (left versus right). To ensure independent training and testing datasets, we used a 20-fold leave-one-run-out cross-validation approach. This process resulted in two decoding accuracy maps (eyes open_{left vs. right} and eyes covered_{left vs. right}) for each subject, in which the value of each voxel represents the average proportion of correctly classified runs. Because the goal of the analysis was to identify areas in which the decoding accuracy was higher in eyes open_{left vs. right} than in eyes covered_{left vs. right}, we calculated the voxel-wise decoding accuracy difference

(eyes open_{left vs. right} minus eyes covered_{left vs. right}). The resulting decoding maps were spatially normalized to the standard Montreal Neurological Institute (MNI) space and smoothed using a 3-mm full width at half maximum (FWHM) Gaussian kernel, and then entered into a second-level analysis using SPM12.

At the second-level, we employed a voxel-wise whole-brain approach. The whole-brain decoding difference map (eyes open_{left vs. right} minus eyes covered_{left vs. right}) was thresholded at $P < 0.001$ (uncorrected for multiple comparisons), using the decoding map representing eyes open_{left vs. right} versus chance level (50%) as an inclusive mask (thresholded at $P < 0.05$, uncorrected), and projected onto orthogonal sections of the average structural scan generated from the 32 subjects. For the statistical inference, we applied corrections for multiple comparisons within our ROIs using the family-wise error (FWE) rate correction implemented in SPM12 (“small-volume correction”). Our a priori defined ROIs consisted of the functionally localized left and right area MT+ (see MT Complex Visual Motion Localizer below), and the left and right TPJ defined as 10-mm-radius spheres centered on the peak MNI coordinates (left TPJ, −54, −60, 21; and right TPJ, 51, −54, 27) in a previous landmark study of theory of mind (11). For areas outside the hypothesized regions, we corrected for multiple comparisons using the whole brain as search space. For the whole-brain analysis, we used the permutation testing approach implemented in the SnPM13 toolbox (45), in which whole-brain significant voxels are identified as voxels whose t values are greater than 95% of the whole-brain maximum t values of a distribution of group-level statistical maps with permuted condition labels (10,000 iterations). Only one such whole-brain significant voxel was identified, and it was located within our a priori defined right TPJ ROI. In a purely descriptive manner, we also report the locations and t values of strong activations ($P < 0.001$, uncorrected) to illustrate the specificity of the significant activations (Table 1). In the figures, the activation maps had a significance threshold of $P < 0.001$ (uncorrected) (Fig. 2 and *SI Appendix, Fig. S3*). For all activations, the coordinates of the peak voxel are given in the MNI standard space (x, y, z).

MT Complex Visual Motion Localizer. To facilitate the localization of BOLD responses within the MT complex (MT+), we exposed subjects to a functional localizer of MT+ (23–25) after completing the 20 runs of the main experiment. The procedures and stimuli were identical to those described for the “Full-field hMT+ visual motion localizer” condition in Jiang et al. (23). Specifically, the localizer stimulus consisted of blocks of moving dots, static dots, and a fixation condition containing no dots. The dots were white on a black background and had a limited lifetime of 200 ms. In the moving condition, all of the dots moved coherently in one of eight directions (spaced evenly between 0° and 360°) with a speed of 8°/s. The direction of motion changed once per second (the same direction was prevented from appearing twice in a row). In the static condition, dots were presented without motion, and the positions of the dots were reset once per second. In the fixation condition, subjects were presented with only the fixation cross but no dots. Dots were presented within a circular aperture (radius 8°) with a central fixation cross-surrounded by a gap (radius 1.5°, to minimize motion-induced eye movements) in the dot field. The diameter of each dot was 0.3°, and dot density was one per square degree.

Participants were asked to fixate throughout the scan and performed no task. Each block lasted 10 s, during which one of the three visual stimulation conditions (motion, static, and fixation) was presented. The three conditions were cycled in a fixed order (motion, static, and fixation). Every participant performed two runs, each lasting ~5 min, and included 30 10-s blocks.

After preprocessing using FMRIprep as described above, we smoothed the spatially normalized functional volumes using a 6-mm FWHM Gaussian kernel. The smoothed data were then entered into a GLM. In the first-level analysis, we defined separate regressors for the moving and static conditions, modeling the 10-s epochs with a boxcar function convolved with the standard SPM12 hemodynamic response function. We defined a linear contrast (moving-static) in the GLM, and the contrast images from all subjects were entered into a random effects group analysis (second-level analysis). The resulting group-level activation map was thresholded at $P < 0.05$ (FWE-corrected using the entire brain as search space), and the distinct clusters of activation in the temporo-parieto-occipital junction on the left and right side defined our MT+ ROIs (*SI Appendix, Fig. S2*).

Eye Tracking. Eye movements were recorded via an MRI-compatible infrared eye tracker (SR Research EyeLink 1000 Plus), mounted just below the projector screen, sampling at 1,000 Hz. Before each scanning session, a

calibration routine on five screen locations was used and repeated until the maximum error for any point was less than 1°. The obtained eye position data were cleaned of artifacts related to blink events and smoothed using a 20-ms moving average. It was then analyzed in two ways. First, we calculated the proportion of time subjects stayed on fixation (defined as an eye position within the display area 2.5° surrounding the fixation point), to estimate how well subjects followed the instructions. Second, we built an SVM decoding model analog to the cross-classification approach used for the fMRI data, but here based purely on eye-tracking data, to test whether eye movement dynamics alone are sufficient to decode the conditions of interest. In keeping with a previous study (46), we organized the data in the following way. The part of the display within which the stimuli appear was divided into a 3 × 9 grid of 27 equally sized (1.6°) squares. For each trial, the proportion of time that the subject fixated within each square (27 features) and the saccades between those regions (27 × 27 = 729 features) were calculated. These 756 features, representing information about both where people were looking as well as saccade dynamics, were then averaged across repetitions for each of the six conditions within each of the 20 runs, yielding one eye movement feature vector per condition per run (per subject). The feature vectors were submitted to an SVM classifier (C = 1). Using a 20-fold leave-one-run-out approach, the SVM model was trained on dot motion left versus dot

motion right and tested on eyes open (left versus right) or eyes covered (left versus right) in the left-out run. At the group level, the decoding accuracies were tested against chance level using *t* tests. The results showed that gaze direction could not be decoded significantly ($P > 0.05$) better than chance in either the eyes open or the eyes covered conditions (see the *SI Appendix*, Fig. S1 legend for statistical details).

Postscan Questionnaire. After the scanning session was completed, subjects were given a questionnaire asking what they thought the purpose of the experiment might be. Though subjects offered guesses about the purpose of the experiment, none indicated anything close to a correct understanding.

Data Availability. The data that support the findings of this study are available at https://figshare.com/articles/Other_People_s_Gaze_Encoded_as_Implied_Motion_in_the_Human_Brain/12184848/1 (47).

ACKNOWLEDGMENTS. This work was supported by the Princeton Neuroscience Institute Innovation Fund. A.G. was supported by the Wenner-Gren Foundation, the Swedish Society of Medicine, and the Foundation Blanceflor. We thank Sam Nastase for valuable input regarding the multivoxel pattern analysis.

1. A. L. Yarbus, "Eye movements during perception of complex objects" in *Eye Movements and Vision*, A. L. Yarbus, Ed. (Springer, 1967), pp. 171–211.
2. L. A. Symons, K. Lee, C. C. Cedrone, M. Nishimura, What are you looking at? Acuity for triadic eye gaze. *J. Gen. Psychol.* **131**, 451–469 (2004).
3. H. Kobayashi, S. Kohshima, Unique morphology of the human eye. *Nature* **387**, 767–768 (1997).
4. M. S. Graziano, S. Kastner, Human consciousness and its relationship to social neuroscience: A novel hypothesis. *Cogn. Neurosci.* **2**, 98–113 (2011).
5. A. Guterstam, H. H. Kean, T. W. Webb, F. S. Kean, M. S. A. Graziano, Implicit model of other people's visual attention as an invisible, force-carrying beam projecting from the eyes. *Proc. Natl. Acad. Sci. U.S.A.* **116**, 328–333 (2019).
6. A. Guterstam, M. S. A. Graziano, Implied motion as a possible mechanism for encoding other people's attention. *Prog. Neurobiol.*, 10.1016/j.pneurobio.2020.101797 (2020).
7. Y. T. Kelly, T. W. Webb, J. D. Meier, M. J. Arcaro, M. S. Graziano, Attributing awareness to oneself and to others. *Proc. Natl. Acad. Sci. U.S.A.* **111**, 5012–5017 (2014).
8. M. S. A. Graziano, A. Guterstam, B. J. Bio, A. I. Wilterson, Toward a standard model of consciousness: Reconciling the attention schema, global workspace, higher-order thought, and illusionist theories. *Cogn. Neuropsychol.*, 1–18 (2019).
9. W. T. Newsome, E. B. Paré, A selective impairment of motion perception following lesions of the middle temporal visual area (MT). *J. Neurosci.* **8**, 2201–2211 (1988).
10. R. B. Tootell et al., Visual motion aftereffect in human cortical area MT revealed by functional magnetic resonance imaging. *Nature* **375**, 139–141 (1995).
11. R. Saxe, N. Kanwisher, People thinking about thinking people. The role of the temporo-parietal junction in "theory of mind". *Neuroimage* **19**, 1835–1842 (2003).
12. H. L. Gallagher et al., Reading the mind in cartoons and stories: An fMRI study of "theory of mind" in verbal and nonverbal tasks. *Neuropsychologia* **38**, 11–21 (2000).
13. K. M. Igelström, T. W. Webb, Y. T. Kelly, M. S. A. Graziano, Topographical organization of attentional, social, and memory processes in the human temporoparietal cortex. *eNeuro* **3**, ENEURO.0060-16.2016 (2016).
14. A. J. Calder et al., Reading the mind from eye gaze. *Neuropsychologia* **40**, 1129–1138 (2002).
15. A. Puce, T. Allison, S. Bentin, J. C. Gore, G. McCarthy, Temporal cortex activation in humans viewing eye and mouth movements. *J. Neurosci.* **18**, 2188–2199 (1998).
16. D. I. Perrett et al., Visual cells in the temporal cortex sensitive to face view and gaze direction. *Proc. R. Soc. Lond. B Biol. Sci.* **223**, 293–317 (1985).
17. E. A. Hoffman, J. V. Haxby, Distinct representations of eye gaze and identity in the distributed human neural system for face perception. *Nat. Neurosci.* **3**, 80–84 (2000).
18. B. Wicker, F. Michel, M.-A. Henaff, J. Decety, Brain regions involved in the perception of gaze: A PET study. *Neuroimage* **8**, 221–227 (1998).
19. B. Julesz, *Foundations of Cyclopean Perception*, (University of Chicago Press, 1971).
20. N. Kriegeskorte, R. Goebel, P. Bandettini, Information-based functional brain mapping. *Proc. Natl. Acad. Sci. U.S.A.* **103**, 3863–3868 (2006).
21. J. V. Haxby et al., Distributed and overlapping representations of faces and objects in ventral temporal cortex. *Science* **293**, 2425–2430 (2001).
22. M. Misaki, Y. Kim, P. A. Bandettini, N. Kriegeskorte, Comparison of multivariate classifiers and response normalizations for pattern-information fMRI. *Neuroimage* **53**, 103–118 (2010).
23. F. Jiang, M. S. Beauchamp, I. Fine, Re-examining overlap between tactile and visual motion responses within hMT+ and STS. *Neuroimage* **119**, 187–196 (2015).
24. A. C. Huk, R. F. Dougherty, D. J. Heeger, Retinotopy and functional subdivision of human areas MT and MST. *J. Neurosci.* **22**, 7195–7205 (2002).
25. M. S. Beauchamp, N. E. Yasar, N. Kishan, T. Ro, Human MST but not MT responds to tactile stimulation. *J. Neurosci.* **27**, 8261–8267 (2007).
26. S. Baron-Cohen, *Mindblindness: An Essay on Autism and Theory of Mind*, (MIT Press, 1997).
27. A. Frischi, A. P. Bayliss, S. P. Tipper, Gaze cueing of attention: Visual attention, social cognition, and individual differences. *Psychol. Bull.* **133**, 694–724 (2007).
28. R. Saxe, Uniquely human social cognition. *Curr. Opin. Neurobiol.* **16**, 235–239 (2006).
29. S. Baron-Cohen, Precursors to a theory of mind: Understanding attention in others. *Nat. Theor. Mind Evol. Dev. Simul. Everyday Mindreading* **1**, 233–251 (1991).
30. S. J. Gould, E. S. Vrba, Exaptation—A missing term in the science of form. *Paleobiology* **8**, 4–15 (1982).
31. S. Zeki et al., A direct demonstration of functional specialization in human visual cortex. *J. Neurosci.* **11**, 641–649 (1991).
32. Z. Kourtzi, N. Kanwisher, Activation in human MT/MST by static images with implied motion. *J. Cogn. Neurosci.* **12**, 48–55 (2000).
33. T. Allison, A. Puce, G. McCarthy, Social perception from visual cues: Role of the STS region. *Trends Cogn. Sci.* **4**, 267–278 (2000).
34. J. Greene, J. Haidt, How (and where) does moral judgment work? *Trends Cogn. Sci.* **6**, 517–523 (2002).
35. V. A. Benassi, P. D. Sweeney, G. E. Drevno, Mind over matter: Perceived success at psychokinesis. *J. Pers. Soc. Psychol.* **37**, 1377–1386 (1979).
36. A. Dundes, *The Evil Eye: A Casebook*, (University of Wisconsin Press, 1992).
37. C. G. Gross, The fire that comes from the eye. *Neuroscientist* **5**, 58–64 (1999).
38. H. Sidky, *The Origins of Shamanism, Spirit Beliefs, and Religiosity: A Cognitive Anthropological Perspective*, (Lexington Books, 2017).
39. G. A. Winer, J. E. Cottrell, V. Gregg, J. S. Fournier, L. A. Bica, Fundamentally misunderstanding visual perception. Adults' belief in visual emissions. *Am. Psychol.* **57**, 417–424 (2002).
40. D. H. Brainard, The psychophysics toolbox. *Spat. Vis.* **10**, 433–436 (1997).
41. N. K. Logothetis, J. Pauls, M. Augath, T. Trinath, A. Oeltermann, Neurophysiological investigation of the basis of the fMRI signal. *Nature* **412**, 150–157 (2001).
42. O. Esteban et al., fMRIPrep: A robust preprocessing pipeline for functional MRI. *Nat. Methods* **16**, 111–116 (2019).
43. K. J. Friston et al., Statistical parametric maps in functional imaging: A general linear approach. *Hum. Brain Mapp.* **2**, 189–210 (1994).
44. M. N. Hebart, K. Gørgen, J.-D. Haynes, The Decoding Toolbox (TDT): A versatile software package for multivariate analyses of functional imaging data. *Front. Neuroinformatics* **8**, 88 (2015).
45. T. E. Nichols, A. P. Holmes, Nonparametric permutation tests for functional neuroimaging: A primer with examples. *Hum. Brain Mapp.* **15**, 1–25 (2002).
46. B. Schneider, Y. Pao, R. D. Pea, "Predicting students' learning outcomes using eye-tracking data" in *LAK '13: Proceedings of the Third International Conference on Learning Analytics and Knowledge*, D. Suthers, K. Verbert, E. Duval, Eds. (Association for Computing Machinery, New York, 2013).
47. A. Guterstam, Other People's Gaze Encoded as Implied Motion in the Human Brain (2020). <https://doi.org/10.6084/m9.figshare.12184848.v1> (23 April 2020).