Visual motion assists in social cognition

Arvid Guterstam[a,b,1] and Michael S. A. Graziano[a]

[a]Department of Psychology, Princeton University, Princeton, NJ 08544; and [b]Department of Clinical Neuroscience, Karolinska Institutet, 171 77 Solna, Stockholm, Sweden

Recent evidence suggests a link between visual motion processing and social cognition. When person A watches person B, the brain of A apparently generates a fictitious, subthreshold motion signal streaming from B to the object of B's attention. These previous studies, being correlative, were unable to establish any functional role for the false motion signals. Here, we directly tested whether subthreshold motion processing plays a role in judging the attention of others. We asked, if we contaminate people's visual input with a subthreshold motion signal streaming from an agent to an object, can we manipulate people's judgments about that agent's attention? Participants viewed a display including faces, objects, and a subthreshold motion hidden in the background. Participants' judgments of the attentional state of the faces was significantly altered by the hidden motion signal. Faces from which subthreshold motion was streaming toward an object were judged as paying more attention to the object. Control experiments showed the effect was specific to the agent-to-object motion direction and to judging attention, not action or spatial orientation. These results suggest that when the brain models other minds, it uses a subthreshold motion signal, streaming from an individual to an object, to help represent attentional state. This type of social-cognitive model, tapping perceptual mechanisms that evolved to process physical events in the real world, may help to explain the extraordinary cultural persistence of beliefs in mind processes having physical manifestation. These findings, therefore, may have larger implications for human psychology and cultural belief.

social cognition | attention | motion | theory of mind

A recent series of reports suggests a link between visual motion processing and social cognition. The human brain appears to fabricate a subtle, false motion signal when looking at another person. The fictitious motion streams in a beam from the other person toward the object of that person's attention (1–3). The signal can be observed directly with functional MRI in motion-processing cortical areas, and it can be observed indirectly by how it causes a motion aftereffect in the area of a scene between a face and the object of the face's attention. The signal, however, is perceptually subthreshold—people are not explicitly aware of it. The functional purpose, if any, of this subthreshold false motion signal is not known, although we speculated it is part of the social toolkit for modeling the attention of others. Because previous studies were correlative—showing a correlation between social cognition and an internally generated motion signal—the causal relationship is not known (4, 5). To establish this new subfield of study in which social cognition taps into preexisting perceptual machinery to model the mind states of others, a direct causal experiment is needed. Here, we provide that test. We asked, if we contaminate a participant's visual world with a subthreshold motion that streams from another person toward an object, can we manipulate the participant's perception of that other person's attention? The results demonstrated a behaviorally meaningful impact of subthreshold motion on social judgments. It explains why the human brain fabricates a motion signal during social cognition. Modeling the attention state of others is a crucial part of social cognition (6–9), and recruiting the motion-processing system evidently contributes to that model. It may have proved adaptive to co-opt the brain's existing motion-processing mechanism to encode

sources and targets of attention, in essence drawing a quick visual sketch with moving arrows to help keep track of who is attending to what in a complex environment. We suggest that the beam of motion represents, instantaneously, the relationship between an agent and the target of its attention. In this interpretation, the subthreshold motion signal balances two adaptive pressures: it is strong enough to influence social cognition in a meaningful direction while at the same time not so strong that it materially interferes with the normal motion perception of real objects. This type of social-cognitive model, borrowing low-level perceptual mechanisms that evolved to process physical events in the real world, may help to explain the extraordinary cultural persistence of beliefs in mind processes having physical manifestation. It is a common belief across time and cultures that attentive gaze comes with a palpable outward flow and that other properties of the mind are linked to specific physical auras and flows. The present findings, therefore, may have larger implications for human psychology and cultural belief.

Results

In experiment 1, subjects were asked to indicate which of two faces seemed to be attending more to an object (Fig. 1A and see *Materials and Methods* for details). Though subjects were not told, one face was an exact mirror reversal of the other. The faces were above and to either side of the central object, with eyeballs tilted slightly downward but not pointing directly at the object. The background was filled with random dot motion featuring 0% coherence (100% random noise). A weak motion signal was added in a beam-shaped area streaming from one face (yellow-highlighted area in Fig. 1A, yellow coloring not visible to subjects), where 30% of the dots moved coherently toward the object. The motion signal was subtle enough that only 7 of 657 subjects across all experiments

Significance

Recent evidence suggests that our brains may generate subtle, fictitious motion signals streaming from other people to the objects of their attention. However, the functional significance of this internally-generated motion signal remains unclear. Here, we tested whether subthreshold motion processing plays a causal role in judging others' attention. Participants viewed a display including faces, objects, and a subthreshold motion hidden in the background. Judgments of the attentional state of the faces were significantly altered by the motion signal. Control experiments indicated the effect was specific to judging attention. These findings suggest that a crucial aspect of social cognition, monitoring others' attention, draws on useful but highly physically inaccurate models of social agents in the world, rooted in low-level perceptual mechanisms.

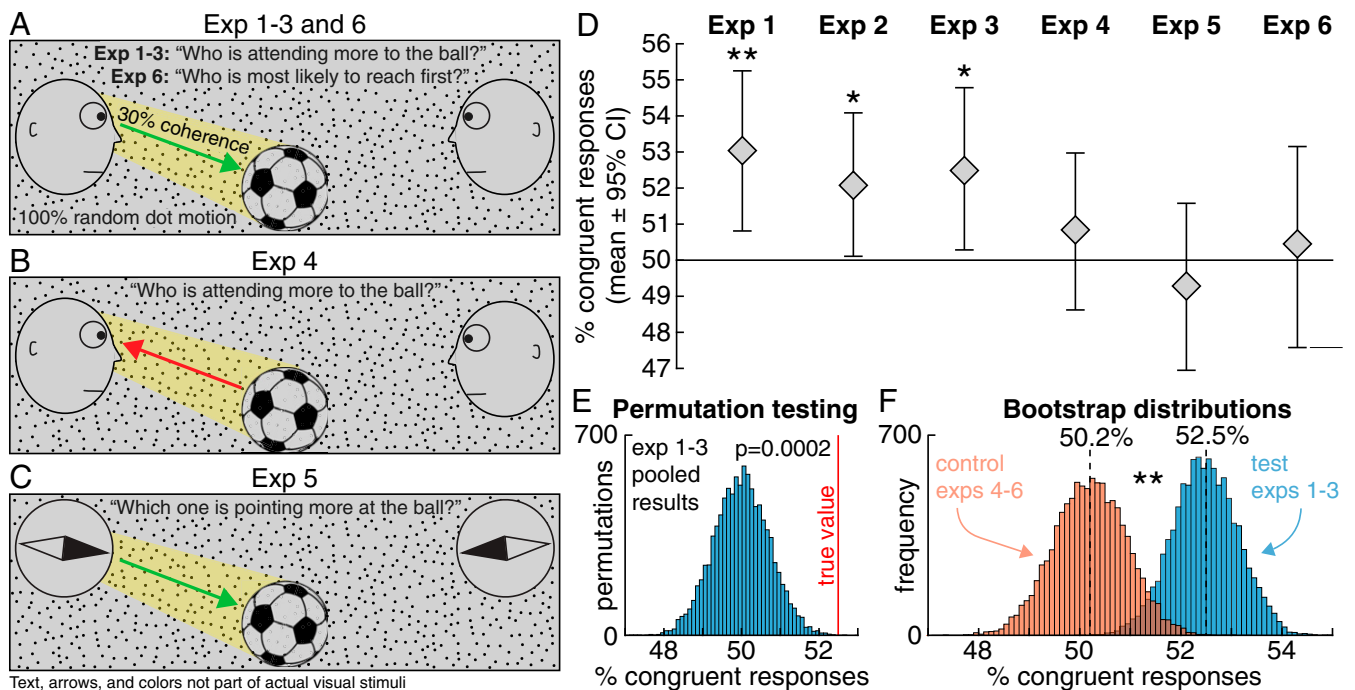[1]To whom correspondence may be addressed. Email: arvidg@princeton.edu.

Fig. 1. The effect of motion on social cognition. (*A*) Subjects viewed two faces turned toward an object, presented on a background of random dot motion noise (0% dot coherence). In a region connecting one face to the object (yellow highlight, color not seen by subjects), 30% of dots moved coherently toward the object, constituting a subtle, subthreshold motion signal (98.9% of subjects reported being unaware of it). The beam extended from the left or right face on interleaved trials. In experiments 1 to 3, on each trial, subjects indicated which face they perceived as attending more to the object. In experiment 6, subjects indicated which person they perceived as more likely to reach for the object first. (*B*) The same except that motion was from object to face. (*C*) The same except compass needles replaced the faces and subjects reported which needle appeared to point more toward the object. (*D*) The proportion of congruent responses (responses in which the subjects selected the face or compass needle adjacent to the motion stream). Error bars represent 95% CI based on a bootstrap distribution. *$P < 0.05$, **$P < 0.01$. (*E*) Meta-analysis. An analysis of all data from experiments 1 to 3 combined is shown. A null distribution of group means was generated based on permutation testing (10,000 iterations) with shuffled condition labels. The red vertical line represents the true mean using the correct condition labels. (*F*) Between-group meta-analysis. An analysis of combined data from all test experiments (1 to 3) and all control experiments (4 to 6) created by generating two bootstrap distributions (10,000 iterations) is shown. exp, experiment.

(1.1%) reported being aware of it in a postexperiment survey (those 7 subjects were excluded from analysis). On half of the randomly interleaved trials, the beam emanated from the left head, and on the other half, it emanated from the right head. Subjects were tested using an online, remote platform (Prolific) (10) due to restrictions on research imposed by the coronavirus epidemic (see *Materials and Methods* for details of sample sizes and exclusion criteria).

Fig. 1*D* summarizes the results. In experiment 1 ($n = 88$), subjects were more likely to choose the face from which the motion streamed toward the object (termed the "congruent face") than the face without motion streaming from it (the "incongruent face") (mean proportion of congruent responses, 53.0%; mean proportion of incongruent responses, 47.0%; $P = 0.006$ by permutation testing). Thus, the motion beam connecting the congruent face to the object, though not overtly noticed, made subjects 6% more likely to choose the congruent face rather than the incongruent face as directing more attention to the object. Experiment 2 was a replication that showed a similar result ($n = 123$; mean proportion of congruent responses, 52.1%; $P = 0.028$ by permutation testing). Experiment 3 was another replication ($n = 68$; mean proportion of congruent responses, 52.5%; $P = 0.040$ by permutation testing). Fig. 1*E* shows an analysis combining experiments 1 to 3 ($n = 279$). On average, subjects chose the congruent face 2.5% more often than the expected chance level of 50%, or, equivalently, subjects were 5.0% more likely to choose the congruent face over the incongruent face as directing more attention to the object (mean proportion of congruent responses, 52.5%; $P = 0.0002$ by permutation testing). The subthreshold motion signal had a significant effect on social decisions about other people's attention.

Experiment 4 was identical to experiments 1 through 3 except that the subthreshold motion was reversed, streaming from the central object to the face (Fig. 1*B*). The effect on social judgment disappeared (Fig. 1*D*; $n = 87$; mean proportion of congruent responses, 50.8%; $P = 0.257$ by permutation testing). This result suggests that the findings in experiments 1 through 3 were unlikely to be caused by low-level features of the motion stream or by the presence of motion on one side of the screen drawing subjects' attention or gaze more to one face.

Experiment 5 was identical to experiments 1 through 3 except that instead of faces, subjects saw two compass needles (Fig. 1*C*). The needles pointed downward to the same degree that the eyeballs had in the previous experiments. Subjects were asked which compass appeared to be pointed more directly to the object. The motion beam had no significant effect on responses ($n = 80$; mean proportion of congruent responses, 49.3%; $P = 0.661$ by permutation testing). This result shows that the effect observed in experiments 1 through 3 was not the result of a low-level visual illusion in which the motion beam caused a general distortion in the perception of tilt angles.

Experiment 6 was identical to experiments 1 through 3 except for the instructions. Rather than judging attention, subjects were asked to judge which person seemed more likely to reach for the object first. The motion beam had no significant effect on responses ($n = 80$; mean proportion of congruent responses, 50.5%; $P = 0.379$ by permutation testing). The result obtained in experiments 1 through 3 was therefore specific to judging the attention of agents, not the actions of agents, even potential actions that involve a motion from the agent toward the object, such as

reaching. Because the stimuli were identical to those in experiment 1, the result also suggests that the effect cannot be explained by the motion beam biasing the subject's attention in any simple manner or by any other simple or low-level feature of the visual display.

Fig. 1*F* shows an analysis of all six experiments, comparing the three control conditions (experiments 4 to 6, $n = 247$) to the three attention beam conditions (experiments 1 to 3, $n = 279$). The mean proportion of congruent responses was significantly greater for the attention beam distribution than for the control distribution (52.5% versus 50.2%, $P = 0.008$ by bootstrap testing). In the control experiments, subjects were 0.4% more likely to pick the congruent choice than the incongruent choice, whereas in the attention beam experiments, subjects were 5.0% more likely to pick the congruent choice than the incongruent choice.

Discussion

Our results demonstrate that it is possible to manipulate people's perception of an agent's attentional state in a predictable manner by artificially introducing a subthreshold motion signal emanating from that agent and streaming toward an object. The control experiments suggest that the effect cannot be explained by factors such as low-level features of the motion stream (experiments 4 and 6), a low-level visual illusion in which the motion stream distorts the general perception of angles (experiment 5), the subjects covertly thinking about any active relationship between the face and the object (experiment 6), and the subject's attention being biased toward one choice by the motion beam (experiments 4 to 6). Instead, the effect appears to be specific to judging a face, when the motion flows specifically from the face to the object, and when the subjects judge the attention state of the face. The finding suggests that motion plays a causal role in coding other people's attention by enhancing or highlighting the directional connection between agent and object. The finding helps explain why the human brain generates a motion signal when processing other people's attentive gaze (2, 3). The link between social cognition and motion processing is not an epiphenomenal trait but rather a useful mechanism that facilitates the tracking of other people's attention. It is an example of how the brain can construct models of the world that serve some adaptive behavioral function, even if those models are not physically accurate depictions of the world. A prediction to be tested in future experiments is that the brain will rely more heavily on this motion-highlighting mechanism in complex social environments that require simultaneous tracking of multiple sources and targets of attention.

Materials and Methods

Participants. For each experiment, participants were recruited through the online behavioral testing platform Prolific (10). Using the tools available on the Prolific platform, we restricted participation such that no subject could take part in more than one of our experiments. Thus, all subjects were naive to the paradigm when tested. All participants indicated normal or corrected-to-normal vision, English as a first language, and no history of mental illness or cognitive impairment. All experimental methods and procedures were approved by the Princeton University Institutional Review Board, and all participants confirmed that they had read and understood a consent form outlining their risks, benefits, compensation, and confidentiality and that they agreed to participate in the experiment. Each subject completed a single experiment in a 3- to 5-min session in exchange for monetary compensation. As is standard for online experiments, because of greater expected variation than for in-laboratory experiments, relatively large numbers of subjects were tested. A target sample size of 100 subjects per experiment was chosen arbitrarily before data collection began. Because of stringent criteria for eliminating those who did not follow all instructions (see below), initial total sample sizes were larger than 100, and final sample sizes for those included in the analysis varied between experiments (experiment 1, $n_{total} = 107$, $n_{included} = 88$, 36 female, mean age 29 [SD, 11]; experiment 2, $n_{total} = 153$, $n_{included} = 123$, 58 female, mean age 34 [SD, 14];

experiment 3, $n_{total} = 103$, $n_{included} = 68$, 22 female, mean age 27 [SD, 9]; experiment 4, $n_{total} = 106$, $n_{included} = 87$, 36 female, mean age 29 [SD, 11]; experiment 5, $n_{total} = 108$, $n_{included} = 80$, 22 female, mean age 25 [SD, 8]; experiment 6, $n_{total} = 102$, $n_{included} = 80$, 35 female, mean age 27 [SD, 9]).

Exclusion Criteria. Subjects were excluded based on a set of predefined criteria. On average, across all six experiments, the exclusion rate was 23% (experiment 1, $n_{excluded} = 19$ [18%]; experiment 2, 30 [20%], experiment 3, 35 [34%]; experiment 4, 19 [18%]; experiment 5, 28 [26%]; experiment 6, 22 [22%]). The most common reason for exclusion was failure to pass the instructional manipulation check (IMC) (experiment 1, 18; experiment 2, 25; experiment 3, 34; experiment 4, 17; experiment 5, 28; experiment 6, 18). The IMC was adapted from ref. 11 and was included to ensure that subjects read and adequately understood the instructions. It consisted of the following sentence inserted at the end of the instructions page: "In order to demonstrate that you have read these instructions carefully, please ignore the 'Continue' button below, and click on the 'x' to start the practice session." Two buttons were presented at the bottom of the screen, "Continue" and "x," and clicking on "Continue" resulted in a failed IMC. Because the intent was for subjects to judge the two faces carefully, we also excluded people with fast (<500 ms) average response times (experiment 1, 1; experiment 2, 0; experiment 3, 0; experiment 4, 1; experiment 5, 0; experiment 6, 0) and subjects who missed (failed to respond within 5 s) more than 30% of trials (experiment 1, 0; experiment 2, 2; experiment 3, 1; experiment 4, 0; experiment 5, 0; experiment 6, 0). One subject in experiment 2 was excluded due to software failure during testing. Finally, subjects who, in the postexperiment survey, reported awareness of the coherent dot motion were excluded from analysis (experiment 1, 0; experiment 2, 2; experiment 3, 0; experiment 4, 1; experiment 5, 0; experiment 6, 4).

Apparatus. After agreeing to participate, subjects were redirected to a website where stimulus presentation and data collection were controlled by custom software based on Hypertext Markup Language, Cascading Style Sheets, JavaScript (using the jsPsych javascript library (12)), and PHP: Hypertext Processor. Subjects were required to complete the experiment in full-screen mode. Exiting full screen resulted in the termination of the experiment and no payment. Because the visual stimuli were rendered on participants' own web browsers, viewing distance, screen size, and display resolutions varied. We therefore report stimulus dimensions below using pixel (px) values.

Visual Stimuli. The visual stimuli (1280 px wide × 370 px high) were created using MATLAB (MathWorks) and then converted to high-resolution, high–frame-rate video files for presentation online. Images of two identical, mirror-reversed cartoon faces (250 px wide × 273 px high), located equidistant on either side of a central circular object (200 px diameter), were presented on top of a random dot motion stimulus and a gray (red: 210, green: 210, blue: 210) background (Fig. 1*A*). The faces were located slightly (90 px) above the object, making the angle between the eyes of the faces and the object 21°. The pupils of the eyeballs were tilted down only 6°, creating an ambiguous impression of whether the faces were looking at the object or not. To increase the variety of the visual stimuli across trials (to prevent subject boredom), we used one of four different objects in each trial: a soccer ball, a pizza, a beach ball, or a dartboard.

The random dot motion stimulus featured black dots. Dot size was 2 px. Dot density was 30 per 100 px × 100 px square. Dot speed was 95 px/s. Dot lifetime was 200 ms. Motion direction of dots was 100% random across the entire display, except for an area forming the "beam" connecting one face to the object (yellow area, Fig. 1*A*). Within the beam area, 30% of the dots moved in a coherent direction toward the object (21° downward motion in experiments 1 to 3 and 5 to 6) or toward the face (21° upward motion in experiment 4), while 70% of the dots moved in random directions. The 30% coherence level was chosen based on the results of a pilot experiment ($n = 7$) conducted in the laboratory, showing that this motion signal was below the threshold of awareness (none out of seven subjects reported being aware of the coherently moving dots while performing the task).

The visual stimuli generated in MATLAB were converted to video files using on-screen recording software capturing 60 frames per second. The videos were edited in Final Cut Pro (Apple Inc.) to generate one video clip per trial type, the size of which was ~20 MB per video. The videos used in all experiments are available at https://figshare.com/articles/dataset/Visual_motion_assists_in_social_cognition/12665078/1 (13). During playback, the video encompassed 90% of the width of the subject's computer screen (which ran the experiment in full-screen mode).

Experimental Design.

Experiment 1. The experiment consisted of an initial instructions page, followed by three practice trials, the experimental session, and a postexperiment survey. The written instructions were as follows: "Humans are extremely sensitive to where other people are attending. In this experiment, you will view images of two people located on either side of an object. Your task is to determine who is attending more to the object. In each trial, indicate the person that you think is paying more attention to the object by hitting Q for the person on the left, or P for the person on the right. Respond as quickly as you can. You must respond within 5 secs."

When the subject pressed the response key to indicate one of the two faces, the trial terminated and the next trial began. If a response was not made within 5 s, the trial terminated and the warning "Too Slow!" was displayed.

There were eight randomly interleaved trial types (motion beam on the left or right side × four different objects), each repeated twice, yielding a total of 16 trials per subject. In online experiments, for better subject compliance, it is advantageous to use few trials per subject, thus minimizing each subject's time commitment while testing a larger sample of subject. For analysis, all trial types were collapsed, and we calculated, for each subject, the proportion of trials in which the subject's response (left or right) was congruent with the location of the motion beam (left or right). At the group level, we used permutation testing for statistical inference. To create a null distribution, we randomly shuffled the motion beam condition labels (left or right) 10,000 times and calculated the group mean proportion of congruent responses for each iteration. We then calculated a P value by comparing the true group mean proportion of congruent responses with the null distribution using the following equation: $P = (1 + \text{number of permuted group mean values} > \text{true value})/(1 + \text{total number of permutations})$. A 95% CI was calculated using bootstrapping, where individual means were resampled with replacement generating a bootstrap distribution of group means (10,000 iterations).

After the experiment, the subjects completed a survey. They were first asked two open-ended questions: "What do you think the hypothesis of the experiment was?" and "How did you perceive the background motion during the experiment?" They were then given the binary yes-or-no question, "Did you perceive the background motion as completely random?" Subjects who responded "no" were asked, "Please describe in what way the background motion wasn't random." Subjects who in any way reported that they had perceived coherently moving dots or streams of motion were excluded from the analysis.

Experiment 2. Experiment 2 was a replication of experiment 1 and used the same procedures.

Experiment 3. Experiment 3 was a replication of experiment 1 with one modification to the instructions: a monetary incentive was added to motivate subjects to perform with greater focus. The written instructions contained the following addition: "The amount of payment will depend on your score, which is an indication of the effort you are putting into the task. If you are correct on at least 80% of the trials, you will receive a 20% bonus pay."

There were, of course, no formally correct answers, and at the end of the experiment, all subjects were awarded the bonus pay.

Experiment 4. The design, procedures, and statistical analysis of experiment 4 were identical to those of experiment 1, with one exception: the coherent dots within the motion beam moved from the object toward the face.

Experiment 5. To exclude low-level visual effects of subthreshold motion on perceived visual angles, in experiment 5, we substituted the two faces with two compass needles (Fig. 1*C*). The compasses matched the face images in position and size. The compass needles matched the eyes in angle (6° downward tilt) and location (the tip of the compass needle was in the same location as the pupil). Subjects were asked to indicate which compass (left or right) pointed more directly to the object: "Humans are extremely sensitive to where things are pointing. In this experiment, you will view images of two compasses located on either side of an object. Your task is to determine which compass is pointing more at the object. In each trial, indicate the compass that you think is pointing more at the object by hitting Q for the compass on the left, or P for the compass on the right. Respond as quickly as you can. You must respond within 5 secs."

Experiment 6. The visual stimuli in experiment 6 were identical to those in experiment 1. The procedures were the same in all respects except for the instructions, which read as follows: "Humans are extremely good at predicting when other people are about to reach for something. In this experiment, you will view images of two faces located on either side of an object. The faces may be subtly different in ways that you do not consciously notice. The two people are both planning to reach for the object. Your task is to decide, based on your intuition about the faces, which person is most likely to reach for the object first."

Between-Group Meta-Analysis. In a between-group meta-analysis (Fig. 1*F*), we combined the responses across all test experiments ($n = 279$; experiments 1 to 3) and compared them to the responses across all control experiments ($n = 247$; experiments 4 to 6). We first generated two bootstrap distributions (resampling with replacement, 10,000 iterations) centered around the mean proportion of congruent responses for each group. We then calculated a P value by comparing the two bootstrap distributions (iteration by iteration in random order) using the following equation: $P = (1 + \text{number of bootstrap iterations for which group mean}_{\text{exp 4 to 6}} > \text{group mean}_{\text{exp 1 to 3}})/(1 + \text{total number of bootstrap iterations})$.

Data Availability. The data that support the findings of this study are available on Figshare at https://figshare.com/articles/dataset/Visual_motion_assists_in_social_cognition/12665078/1 (13).

1. 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).
2. A. Guterstam, A. I. Wilterson, D. Wachtell, M. S. A. Graziano, Other people's gaze encoded as implied motion in the human brain. *Proc. Natl. Acad. Sci. U.S.A.* **117**, 13162–13167 (2020).
3. A. Guterstam, M. S. A. Graziano, Implied motion as a possible mechanism for encoding other people's attention. *Prog. Neurobiol.* **190**, 101797 (2020).
4. M. Görner, H. Ramezanpour, I. Chong, P. Thier, Does the brain encode the gaze of others as beams emitted by their eyes? *Proc. Natl. Acad. Sci. U.S.A.* **117**, 20375–20376 (2020).
5. A. Guterstam, M. S. A. Graziano, Reply to Görner et al.: Encoding gaze as implied motion. *Proc. Natl. Acad. Sci. U.S.A.* **117**, 20377 (2020).
6. S. Baron-Cohen, *Mindblindness: An Essay on Autism and Theory of Mind* (MIT Press, 1997).
7. M. S. A. Graziano, S. Kastner, Human consciousness and its relationship to social neuroscience: A novel hypothesis. *Cogn. Neurosci.* **2**, 98–113 (2011).
8. A. Frischen, A. P. Bayliss, S. P. Tipper, Gaze cueing of attention: Visual attention, social cognition, and individual differences. *Psychol. Bull.* **133**, 694–724 (2007).
9. A. J. Calder *et al.*, Reading the mind from eye gaze. *Neuropsychologia* **40**, 1129–1138 (2002).
10. S. Palan, C. Schitter, Prolific.ac—A subject pool for online experiments. *J. Behav. Exp. Finance* **17**, 22–27 (2018).
11. D. M. Oppenheimer, T. Meyvis, N. Davidenko, Instructional manipulation checks: Detecting satisficing to increase statistical power. *J. Exp. Soc. Psychol.* **45**, 867–872 (2009).
12. J. R. de Leeuw, jsPsych: A JavaScript library for creating behavioral experiments in a Web browser. *Behav. Res. Methods* **47**, 1–12 (2015).
13. A. Guterstam, Visual motion assists in social cognition. *Figshare.* Dataset. https://doi.org/10.6084/m9.figshare.12665078.v1. Deposited 30 October 2020.