ELSEVIER

Original Research Article

Implied motion as a possible mechanism for encoding other people's attention

Arvid Guterstam*, Michael S.A. Graziano

Department of Psychology, Princeton University, Princeton, NJ 08540, United States

ARTICLE INFO

ABSTRACT

Recent evidence suggests that the human brain automatically constructs a rich model of other people's attention, beyond registering low-level cues such as someone else's gaze direction. This model is not a physically accurate representation of attention, but instead appears to contain simplifying and physically incoherent features. For example, without explicitly realizing it, people treat the attentive gaze of others as though it exerts a gentle force pushing on objects. Here we specify another aspect of that implicit model of attention. People treat the attentive gaze of an agent as though it were travelling through space, with an implied motion encoded literally enough that it causes a perceptual motion adaptation effect. This implicit model of other people's attention may facilitate the process of keeping track of who is attending to what, which is essential for reading and predicting the minds and behavior of social agents. This implicit model of attention may also have shaped culturally widespread ideas about mind and spirit.

1. Introduction

Building a model of the attention of another agent is arguably one of the most crucial tasks in social cognition (Baron-Cohen, 1997; Graziano and Kastner, 2011). Attention has a profound effect on behavior, influencing how signals in the brain impact output systems, and therefore to predict the behavior of another agent, it would be useful to model the attentional state of that agent. It is not sufficient to reconstruct a list of items that the agent is attending; one needs a model of the dynamics and consequences of attention. How exactly we model the attention of others, however, has been surprisingly understudied. Most previous work on the topic in psychology and neuroscience has concentrated on how we monitor the gaze direction of others. The reconstruction of gaze direction is considered essential for reading other people's intentions, beliefs, and other components of theory of mind (Baron-Cohen, 1997; Calder et al., 2002; Carlin and Calder, 2013; Friesen and Kingstone, 1998; Frischen et al., 2007; Hoffman and Haxby, 2000; Kelly et al., 2014; Perrett et al., 1985; Symons et al., 2004; Wicker et al., 1998). Recent evidence, however, has begun to reveal a much richer model of other people's attention that goes beyond noting the direction of gaze (Guterstam et al., 2019; Pesquita et al., 2016). For example, people may implicitly encode whether someone else's attention was drawn to an object exogenously (by bottom-up, stimulus-driven salience) or endogenously (by internal choice) (Pesquita et al., 2016). Moreover, we

recently found that in modeling the attention of others, people may incorporate physically incorrect attributes such as a perceived gentle force that the attending agent applies to the attended object (Guterstam et al., 2019). Results like these begin to build a picture of an automatically constructed, schematic and quirky, partly implicit model of the attention of other agents.

The purpose of the present set of seven experiments was to test another aspect of this implicit model of attention. We used a visual motion adaptation paradigm to show that participants implicitly treated attention as though it were a flow moving invisibly through space from an agent to an object. The participants reported no explicit knowledge of this bias. It therefore appeared to reflect an implicit understanding or model. We speculate that this model of other people's attention, as a fluid-like substance that is generated inside of an agent and flows out toward targets, may be a useful simplification for keeping track of who is attending to what and by how much. It is possible that during the evolution of social brain mechanisms for tracking others' attention, it was adaptive for the brain to make use of already existing visual motion mechanisms. The present data suggest that people construct a simplified model of the attention of others, much of the model is constructed at an implicit level, and at least some aspects of the model are schematic and extremely physically inaccurate. Even though this model is automatic and partly implicit, we suggest it may have played a significant role in biasing intuitions and therefore in shaping

---

* Corresponding author.
  E-mail address: arvidg@princeton.edu (A. Guterstam).

culturally common ideas about mind and consciousness.

2. Materials and methods

The logic of the motion adaptation paradigm is that two motion stimuli are presented in sequence, within a single trial. The first causes a small amount of adaptation, affecting the response to the second. Subjects indicate the direction of the second stimulus in a speeded response. If the second stimulus has the same direction as the first, subjects are slower to respond due to the adaptation interference. If the second stimulus has the opposite direction as the first, subjects are relatively faster to respond. In this manner, one can confirm that subjects processed the motion of the first stimulus because of its impact on the response to the motion of the second stimulus. A similar, within-trial adaptation effect of one motion on detecting a second motion was observed previously (Levinson and Sekuler, 1980). Motion adaptation can occur with very brief stimulus exposures (Glasser et al., 2011), even shorter than those used in the present experiment, and therefore some measurable adaptation was expected here. The hypothesis that a motion adaptation effect might be observed for implied rather that actual motion, tested in our subsequent experiments, is lent some credence by previous studies in which viewing photographs with an implied motion (e.g., a running animal) activates motion-sensitive areas of the brain (Kourtzi and Kanwisher, 2000; Krekelberg et al., 2003) and is associated with direction-specific adaptation assessed with real-motion test probes (Winawer et al., 2008). The purpose of experiment 1 was to confirm the basic motion adaptation effect with real motion, so that it could be used as the basis for experiments 2–7.

The methods for experiment 1 are described in detail here. The modifications for experiments 2–7 are described briefly as needed throughout the Results. All subjects provided informed consent and all procedures were approved by the Princeton Institutional Review Board. In total, 37 subjects were recruited for experiment 1. Subjects sat stabilized by a chinrest 54 cm from a computer screen and used key presses on a standard keyboard for behavioral responses. The screen was a 38-cm-wide CRT monitor operating at 80 Hz and 1600 × 1200 pixel resolution. Visual stimuli were presented on the screen using MATLAB (MathWorks) and the Psychophysics Toolbox (Brainard, 1997). All experiments were carried out in a darkened, quiet room.

Eye movements were recorded via a desktop-mounted eye tracker (SR Research EyeLink 1000 Plus) sampling at 1000 Hz. Before each experiment, a calibration routine on nine screen locations was used and repeated until the maximum error for any point was less than 1°. Before analyzing the obtained eye position data, it was cleaned of artifacts related to blink events and smoothed using a 20-ms moving average.

Fig. 1A shows the behavioral paradigm for experiment 1. After a variable $1-2$ s inter-trial interval in which a neutral gray field covered the screen, a black central fixation point (0.5° diameter) appeared. Subjects were instructed to fixate on the point and try to maintain fixation in that area of the screen throughout the trial. After 1.5 s, the point disappeared and a gray-scale sinusoidal grating (one grating period = 0.8°, overall size of grating 14.7° wide x 5.7° high) was presented for 1.5 s. While the rectangular boundary of the grating remained stationary, the grating itself within the boundary moved to the left or right at 0.8°/s.

After 1.5 s, the grating disappeared and a random dot motion stimulus was presented (Britten et al., 1992). This type of stimulus is standard for psychophysical tests of motion direction judgment. The black dots on the gray background were presented within a central, 5° x 5° area. Dot density was 50 dots per square visual degree. Each dot was 0.05° in diameter, had a velocity of 2°/s and a lifetime of 200 ms. Dot direction was random for 60 % of dots and coherent for 40 % of dots;

on the keyboard as quickly as possible. As soon as a response was given, the dot motion stimulus disappeared and the inter-trial interval began. If the subject did not respond within a 2 s time window, the dot motion stimulus disappeared, the text "Too Slow!" was presented for 5 s, and then the inter-trial interval began. Subjects responded within 2 s on most trials (97.1 %). If a subject did not respond during this time window, the same configuration of trial was automatically repeated later in the test sequence so that all subjects completed a balanced number of trial types.

On each trial, the grating could move to the left or right, and the dot motion stimulus could be congruent (same direction as the grating) or incongruent. Subjects were instructed that their only task was to determine the direction of the dot motion, and that the grating movement was task irrelevant. All trial types were balanced and presented in a random order. For analysis, trial types were collapsed into two major conditions: congruent and incongruent. Subjects performed 60 trials in three blocks of 20 trials each, thus 30 trials per major condition.

The primary analysis was focused on reaction time rather than accuracy, though we report both measures. In general, reaction time is a more sensitive measure than accuracy for many reasons, including that reaction time varies on a continuous scale whereas accuracy is binary – either yes or no on each trial. Given the limited response window and therefore speeded response in our paradigm, we hoped that reaction time would be a sensitive enough measure to register evidence of implied motion. As reported in the Results, reaction time did prove to be a more sensitive measure than accuracy. This emphasis on reaction-time is common in visual behavioral experiments. For each subject, we computed the average reaction time (RT) for congruent and incongruent trials. A difference score was then computed: ΔRT = [average RT in congruent trials] - [average RT in incongruent trials]. A positive score indicated that subjects were impaired (slower) at discriminating the motion of the dot stimulus when it followed an adapting motion of the same direction, compared to when it followed an adapting motion of the opposite direction. Thus, a positive score would indicate that some degree of motion adaptation had occurred. Below we also report the mean reaction times for the congruent and incongruent conditions separately. In addition, we report a difference score for accuracy, ΔA, for which a negative score would indicate adaptation, although, as expected, the accuracy measure proved to be less sensitive than the reaction time measure.

People show substantial individual differences in ability to detect random dot motion direction (Pilly and Seitz, 2009). Because our experiment would not be meaningful if subjects were unable to perceive the direction of the dot stimulus, each subject began with a practice session consisting of 10 trials, after which feedback regarding performance accuracy was displayed on the screen. If a subject did not reach at least 80 % accuracy, the practice was repeated up to four times. Subjects who were unable to achieve 80 % accuracy within four practice sessions were excluded and did not undergo further testing. Furthermore, subjects who successfully completed the practice but whose overall accuracy in the main experiment was at chance level (not significantly greater than 50 % as determined by permutation testing with 10,000 iterations) were excluded from analyses. Of 37 subjects recruited, 13 were excluded for inability to perform the dot motion task to criterion, leaving 24 subjects in the final analysis (15 females, 18–23 y old, normal or corrected-to-normal vision). Among the 24, discrimination accuracy was 91.8 % on average.

Based on the sample sizes used in a previous, related study (Guterstam et al., 2019) as well as on pilot data, we aimed to include about 24 subjects for each of the 7 experiments reported here, and tested more to compensate for exclusions. The primary hypothesis tested in this series of experiments was whether a static image of a face
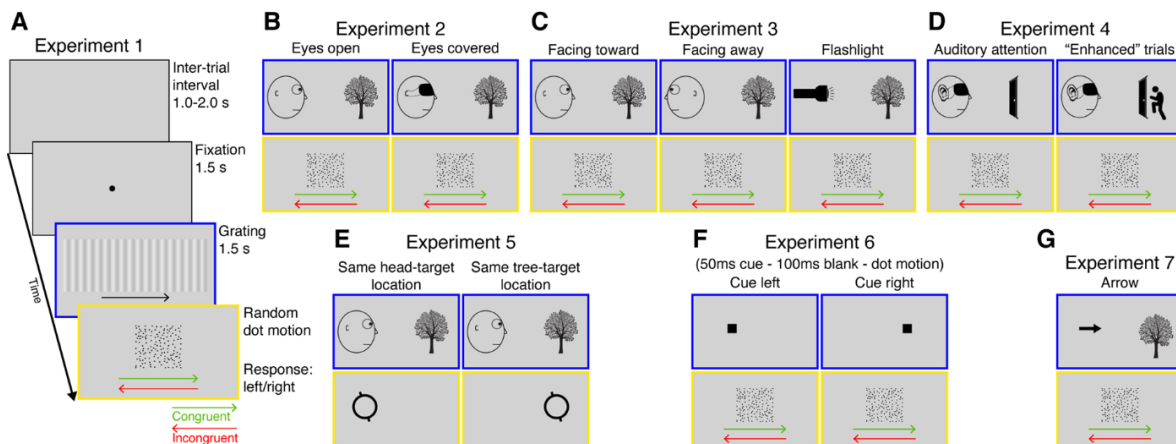
Fig. 1. Behavioral paradigms for experiments 1-7. **A.** Experiment 1. After a 1-2 s inter-trial interval, subjects fixated on a central spot for 1.5 s, then saw a grating for 1.5 s drifting left or right, the right motion illustrated here (arrow not visible to subjects). Subjects then saw a random dot motion stimulus for up to 2 s also drifting left or right. Arrows (not visible to subjects) illustrate dot motion congruent (green arrow) or incongruent (red arrow) to grating direction. Subjects reported the direction of motion of the dot stimulus by key press as quickly as possible. Blue outline and yellow outline (not visible to subjects) indicate the two key phases of the paradigm altered in experiments 2-7. **B.** Experiment 2. Similar to experiment 1 except the first stimulus was a static, line-drawing image of a head facing a tree. The head could be on the left facing the right (shown here), or on the right facing the left; and it could be sighted or blindfolded. The second stimulus, the dot motion, could be congruent (same direction that the head was facing) or incongruent. **C.** Experiment 3. The first stimulus could be one of three items: a head facing a tree, a head facing away from the tree, or a flashlight icon facing the tree. Whether the tree was to the right of the screen (shown here) or left was counterbalanced. The second stimulus, the dot motion, could be congruent (streaming toward the tree) or incongruent. Subjects discriminated the direction of the second stimulus as in experiment 1. **D.** Experiment 4. The first stimulus was a blindfolded head facing a door. Whether the head was on the left of the screen (shown here) or right was counterbalanced. Subjects were told that "Kevin" was attentively listening for anyone sneaking up to the other side of the door. On a small proportion of "enhanced" trials, subjects were reminded of the story by an icon of a person approaching the door. The second stimulus, the dot motion, could be congruent (streaming toward the door) or incongruent. **E.** Experiment 5. The first stimulus was a head facing a tree. Whether the head was to the left of the tree (shown here) or to the right was counterbalanced. The second stimulus was a target object at the location of the head or tree. Subjects indicated whether the target tilted to the left or right. **F.** Experiment 6. The first stimulus was a black square presented to the left or to the right for 50 ms, followed by a blank screen for 100 ms. The second stimulus, the dot motion, could be congruent (direction toward the location of the cue) or incongruent. **G.** Experiment 7. The first stimulus was an arrow pointing at the tree for 1.5 s. The arrow could be on the left pointing right (shown here), or on the right pointing left. The second stimulus, the dot motion, could be congruent (same direction as the arrow) or incongruent. Subjects discriminated the direction of the second stimulus as in experiment 1.

achieve a power of 0.8, suggesting that our experiments were adequately powered to detect the relevant ΔRT effects.

3. Results

3.1. Experiment 1

As shown in Fig. 2A, ΔRT was significantly greater than zero (congruent RT = 899 ms; incongruent RT = 841 ms; ΔRT = 57 ms, SE = 14, t_{23} = 4.17, p < 0.001; for accuracy data, ΔA = -5.0 %, SE = 1.8,

t_{23} = -2.79, p = 0.011). Subjects were slower to discriminate the direction of the dot motion stimulus when it followed an adapting stimulus moving in the same direction, than when it followed an adapting stimulus moving in the opposite direction. This result is consistent with a motion adaptation effect. The confirmation allowed us to use the same type of paradigm and analysis in experiments 2–7.

3.2. Experiment 2

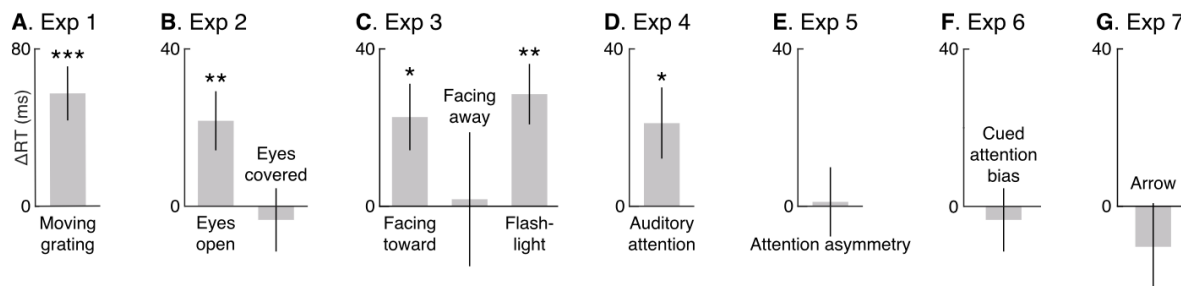Experiment 2 used the same paradigm as experiment 1, except that

Fig. 2. Results for experiments 1-7. **A.** Experiment 1. Effect of moving visual grating on subsequent dot motion discrimination. ΔRT = [average RT for congruent dot motion] - [average RT for incongruent dot motion]. Bar shows mean ΔRT across subjects. Error bar shows standard error. ΔRT significantly > 0 indicated with * (p < 0.05), ** (p < 0.01) or *** (p < 0.001). See text for statistical details. **B.** Experiment 2. Effect of sighted and blindfolded head on subsequent dot motion discrimination. ΔRT defined as in Experiment 1. **C.** Experiment 3. Effect of head gazing at tree, head facing away from tree, and flashlight aimed at tree, on dot motion discrimination. ΔRT defined as in Experiment 1. **D.** Experiment 4. Effect of blindfolded but listening head on dot motion discrimination. ΔRT defined as in Experiment 1. **E.** Experiment 5. Relative visual attention on head and tree stimuli, measured through target discrimination. ΔRT = [mean RT when target appeared

the first, adapting stimulus was not a moving grating, but instead a static image of a face gazing across the screen toward an arbitrary object, a tree (Fig. 1B). When this static image disappeared, the random-dot stimulus was presented in the space interposed between where the head and the tree had been. The purpose of this experiment was to determine whether the static image of a face gazing at an object could induce a motion adaptation effect. If so, the result would suggest that subjects encoded the static image as though something were moving across the empty space from the face to the tree.

All methods were the same as in experiment 1 except in the following ways. As noted, the moving grating stimulus was replaced by the image of a head gazing at a tree. The head could be on the left of the display gazing at the tree on the right, or on the right gazing at a tree on the left. In addition, the head could have uncovered eyes or a black blindfold covering the eyes. In the instructions, subjects were not told anything about the head except that it was irrelevant to their motion discrimination task. All other aspects of the paradigm, including the dot motion stimulus and the direction discrimination task, were the same as in experiment 1. For analysis, trials were grouped into four major conditions: congruent (head pointing in the same direction that the dot motion stimulus moved) and eyes open; incongruent and eyes open; congruent and eyes closed; and incongruent and eyes closed. Subjects performed 120 trials in 6 blocks of 20 trials each, thus 30 trials per condition.

Thirty-two entirely naïve subjects, untested in any of the other experiments, were recruited for experiment 2, of which 8 were excluded for poor performance on the dot motion task (see **Materials and Methods** for exclusion criteria), leaving 24 in the analysis (15 females, 18–31 y old).

After all trials were completed, subjects were given a questionnaire asking what they thought the purpose of the experiment might be, and whether they were explicitly aware of any influence of the head-and-tree stimulus on their ability to respond to the dot motion stimulus. Though subjects offered guesses about the purpose of the experiment, none indicated anything close to a correct understanding. All subjects also insisted that, as far as they were aware, the head-and-tree stimulus had no impact on their response to the second stimulus. These questionnaire results suggest that any motion adaptation effects observed here probably occurred at an implicit level.

Fig. 2B shows that when the eyes were uncovered, ΔRT was significantly greater than zero (congruent RT = 751 ms; incongruent RT = 729 ms; ΔRT = 22 ms, SE = 8, t_{23} = 2.89, p = 0.008; ΔA = -2.4 %, SE = 1.6 %, t_{23} = -1.52, p = 0.141). The subjects' treatment of the static image of a head gazing at an object was therefore similar to their treatment of the moving grating stimulus from experiment 1. Both stimuli directionally affected how subjects responded to the random dot stimulus. The effect was consistent with subjects implicitly encoding the gaze stimulus as though something were moving from the head to the tree, passing through the intervening space, inducing a measurable motion adaptation effect. The actual visual movement tested in experiment 1, however, had a stronger effect (compare Fig. 2A with B).

When the head was blindfolded, ΔRT was not significantly greater than zero (congruent RT = 728 ms; incongruent RT = 732 ms; ΔRT = -3 ms, SE = 8, t_{23} = -0.43, p = 0.670; ΔA = -2.6 %, SE = 1.5 %, t_{23} = -1.78, p = 0.088). Even though the blindfolded head still had an obvious directionality pointed toward the tree, directionality by itself, without open eyes gazing at the tree, was evidently not enough to induce a measurable motion adaptation effect.

3.3. Eye movement analysis

One possibility is that an asymmetric distribution of the subjects'

distribution of subjects' eye position did not differ significantly between the eyes-open and eyes-covered conditions, and therefore could not easily explain why the eyes-open condition showed a motion adaptation effect and the eyes-covered condition did not. As expected, subjects tended not to look at the head or the tree, since these stimuli were irrelevant to the behavioral task. Instead they tended to look toward the center as instructed, with little evidence of asymmetry.

To determine whether the eyes-open condition caused a shift in horizontal eye position relative to the eyes-covered condition, we first coded eye position relative to the head shown in the display (flipping the x axis when the head was to the right of the display). We then computed a difference score for each subject: ΔX = [mean horizontal eye position for the eyes-open trials] – [mean horizontal eye position for the eyes-covered trials]. We then performed a t-test among the 24 subjects and found that ΔX was not significantly different from zero (mean ΔX = 0.032°, SE = 0.036, t_{23} = 0.90, p = 0.380). We found a similar result when we analyzed eye position during the first phase of the trial, when the head-and-tree stimulus was present, before the dot-motion stimulus appeared (mean ΔX = 0.104°, SE = 0.075, t_{23} = 1.38, p = 0.180). Furthermore, the mean horizontal eye position did not differ significantly between the first (head-and-tree) and second (dot motion) phases of the trial in the eyes open (t_{23} = 0.47, p = 0.644) and eyes covered (t_{23} = 1.53, p = 0.140) conditions. Again, this result was expected, since subjects were instructed to fixate centrally and the head and tree images were not relevant to the task as far as the subjects knew.

We also analyzed saccades during the random-dot discrimination phase of the trial. Saccades were defined as eye movements exceeding the velocity of 22°/s. Each saccade in the data was assigned the value -1 (if toward the face in the display) or +1 (if toward the tree in the display). The duration of the saccade in milliseconds was also recorded. For each trial, we computed three variables: total number of saccades, the sum of the values coding saccade direction (<0 if more saccades toward the face, >0 if more saccades toward the tree, = 0 if equal number of saccades in both directions), and the total duration in ms, relative to x direction (<0 if longer total duration of saccades toward the face, >0 if longer total duration of saccades toward the tree, = 0 if equal total duration of saccades in both directions). At the group level, we used two-tailed t tests to contrast the eyes-open and eyes-covered conditions. We found no significant differences with respect to the average number of saccades per trial (ΔN = 0.005, SE = 0.058, t_{23} = 0.093, p = 0.926), the relative number of saccades per trial directed toward versus away from the location of the tree (ΔN = -0.05, SE = 0.05, t_{23} = -1.01, p = 0.322), or the relative durations of the saccades directed toward versus away from the tree (ΔDuration = -0.2 ms, SE = 1.2, t_{23} = -0.13, p = 0.895). Again, we found a similar result when we analyzed eye position during the first phase of the trial, when the head-and-tree stimulus was present, before the dot-motion stimulus appeared (average number of saccades per trial, ΔN = -0.10, SE = 0.13, t_{23} = -0.77, p = 0.448; relative number of saccades per trial directed toward versus away from the location of the tree, ΔN = -0.05, SE = 0.05, t_{23} = -0.98, p = 0.337; relative durations of saccades directed toward versus away from the tree, ΔDuration = -0.7 ms, SE = 1.5, t_{23} = -0.46, p = 0.653).

We therefore found no evidence that the image of a head with open eyes, relative to closed eyes, affected the performance on the random-dot discrimination task by causing any horizontal shift in the subjects' eye position or saccades, such as might occur if subjects were biased to look more at the head or more at the tree. It is possible, however, that covert attention, dissociated from overt eye position, may have become asymmetric, an issue addressed below in experiments 5 - 7.
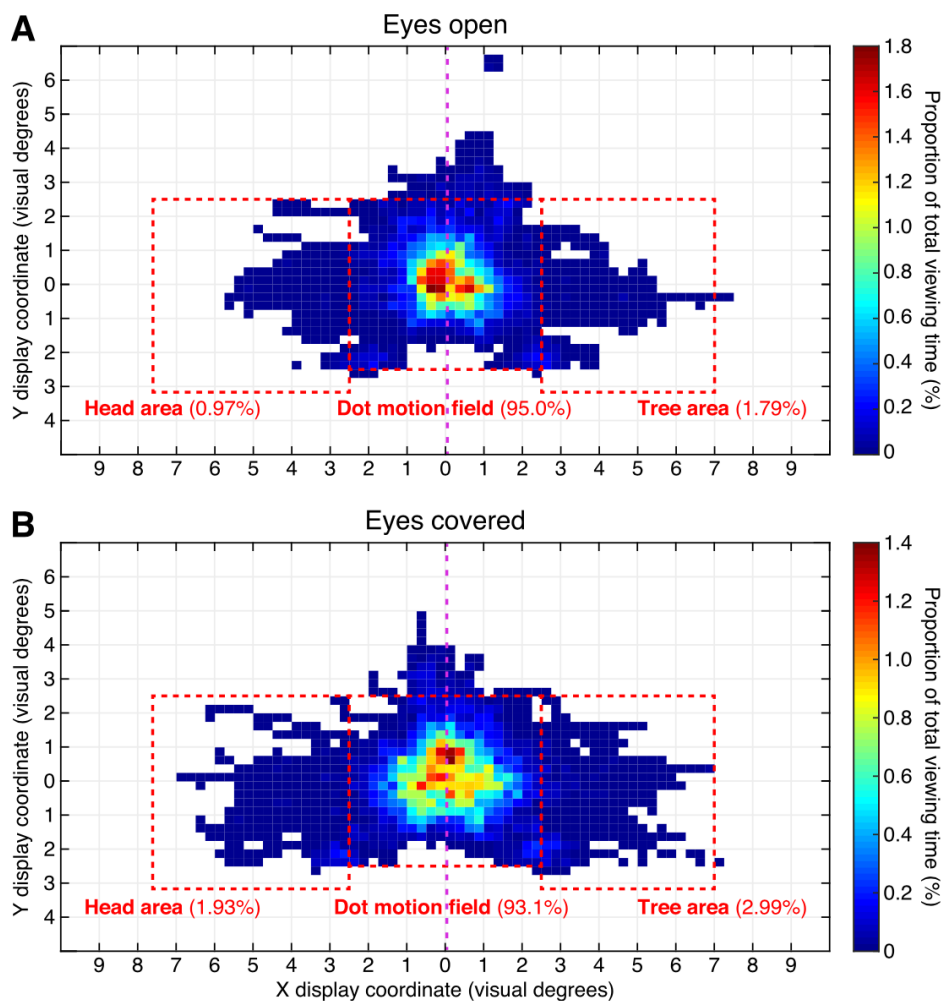
Fig. 3. Distribution of eye position in experiment 2. For trials in which the head appeared on the right side, the X coordinate of the data was flipped such that every trial was coded as though the head were to the left. Data is shown from the task phase in which the random-dot motion stimulus was present. **A.** Distribution of eye positions when the head was not blindfolded. The percentage of time that subjects' eye position fell within each $0.25 \times 0.25°$ grid square is indicated with a color scale. Dotted red lines indicate the display regions in which the random-dot field, the head, and the tree were displayed. Percentage of time spent in each of these three regions is indicated. The vertical dotted line in magenta shows the average horizontal eye position. **B.** Distribution of eye positions when the head was blindfolded.

2 except in the following ways. As shown in Fig. 1C, three image types were randomly interleaved. First, the head was displayed gazing at the tree as in experiment 2. Second, the head was displayed facing away from the tree. The facing-away condition served as a control for the mere presence of eyes in the stimulus. Because the hypothesized eye-beams project from the eyes in the direction of an agent's attention, the facing-away condition should be associated with implied motion directed away from the tree. The implied motion should thus fall outside the central area of the screen within which the subsequent random-dots are displayed, preventing motion adaptation mechanisms from affecting the subjects' performance in the motion discrimination task. Therefore, we predicted that ΔRT should not be significantly different from zero. The third interleaved image type in experiment 3 was a picture of a flashlight aimed at the tree. The flashlight image was used because it implied energy travelling out of the flashlight toward the tree, and therefore, we hypothesized, it should have the same effect on performance as the head gazing at the tree. Just as in experiment 2, the static image was followed by the dot-motion stimulus, and subjects performed the direction discrimination task. On each trial, the object opposed to the tree could be a head facing the tree, a head facing away from the tree, or a flashlight facing the tree; the object could be on the left with the tree on the right, or on the right with the tree on the left; and the dot motion could be flowing to the left or to the right. Thus 12

presented in a random order. Subjects performed 120 trials in six blocks of 20 trials each, thus 20 trials per major condition. For each of the three image types, a ΔRT was computed to test whether the image induced a motion adaptation effect. Thirty-two entirely naïve subjects, untested in any of the other experiments, were recruited for experiment 3, of which 8 were excluded for poor performance on the dot motion task, leaving 24 in the analysis (17 females, 18–27 y old).

Fig. 2C shows the results. When the head gazed at the tree, ΔRT was significantly greater than zero, thus replicating the results of experiment 2 (congruent RT = 788 ms; incongruent RT = 765 ms; $\Delta RT = 24$ ms, SE = 9, $t_{23} = 2.68$, p = 0.013; $\Delta A = -1.0$ %, SE = 1.1 %, $t_{23} = -0.93$, p = 0.364). When the eyes were open but facing away from the tree, ΔRT was not significantly greater than zero (congruent RT = 779 ms; incongruent RT = 777 ms; $\Delta RT = 2$ ms, SE = 18, $t_{23} = 0.10$, p = 0.918; $\Delta A = 4.4$ %, SE = 2.1 %, $t_{23} = 2.11$, p = 0.046). Finally, as predicted, when the flashlight was aimed at the tree, ΔRT was significantly greater than zero (congruent RT = 782 ms; incongruent RT = 752 ms; $\Delta RT = 30$ ms, SE = 8, $t_{23} = 3.70$, p = 0.001; $\Delta A = -1.3$ %, SE = 1.5 %, $t_{23} = -0.84$, p = 0.408). These results suggest that people literally treat an agent gazing at an object in the same way that they treat light beaming at the object. Both are implicitly encoded as though something were streaming from a source to a target.

another agent (Baron-Cohen, 1997; Calder et al., 2002; Friesen and Kingstone, 1998; Frischen et al., 2007; Kelly et al., 2014; Symons et al., 2004). However, we hypothesized that when people build a model of the attention of other agents, the model is deeper than just a reconstruction of gaze direction or a reconstruction of visual attention. Though other cues to attention may be less potent, it is possible that people can use contextual cues to model attention, and also possible that people model a general, supramodal attention rather than specifically visual attention. The purpose of experiment 4 was to test whether the implicit motion effect measured here could indeed generalize beyond the gaze cue and beyond vision. Subjects saw a blind-folded face turned toward an object. We already know from experiment 2 that no motion adaptation effect is expected in this configuration. However, in experiment 4, we told subjects that the head in the display was listening intently – directing auditory attention. We hypothesized that the addition of the contextual information about auditory attention would cause the motion adaptation effect to re-emerge despite the presence of the blindfold.

The methods were the same as in the previous experiment except in the following ways. As shown in Fig. 1D, the static image included a picture of a blindfolded face to one side of the display and a door to the other side. During the instruction period, subjects were told that Kevin, the character depicted in the display, could not see but was listening intently for anyone who might sneak up on the opposite side of the door. The face was named only in the present experiment, to make the instructions clearer. To remind subjects of Kevin's supposed state of attention, in a minority of trials (12/72), a cartoon figure was displayed on the opposite side of the door, sneaking toward it and reaching one hand as though about to open it. These "enhanced" trials were not included in the analysis because of their potential for visual distraction. Only trials without the added figure – trials that included only two objects, the head and the door – were analyzed. It is important to note that, in these analyzed trials, the contextual information provided to the subjects did not suggest that there was any actual sound at the door, which might have caused subjects to perceive an implied motion of sound waves from the door to Kevin's head. Instead, in the absence of anyone actually at the door making a sound, Kevin was directing auditory attention toward the door. On each trial, Kevin could be to the left of the display or to the right; and the subsequent dot motion stimulus could move toward the left or the right. All trial types were balanced and presented in a random order. Trials were collapsed for analysis into two major conditions: congruent (the dot motion was toward the location of the door) and incongruent (the dot motion was away from the location of the door). Subjects performed 72 trials in 6 separate blocks, thus 30 trials per major condition plus 12 unanalyzed trials with enhanced illustration. Twenty-nine entirely naïve subjects, untested in any of the other experiments, were recruited for experiment 4, of which 2 were excluded for poor performance on the dot motion task, leaving 27 in the analysis (11 females, 18–22 y old).

Fig. 2D shows that ΔRT was significantly greater than zero (congruent RT = 725 ms; incongruent RT = 704 ms; ΔRT = 21 ms, SE = 9, t_{26} = 2.34, p = 0.027; ΔA = 1.0 %, SE = 1.3 %, t_{26} = 0.79, p = 0.439). Subjects treated the image of the blindfolded head as though something were moving from the head, across the intervening space, to the door, the object of attention. This result supports the hypothesis that the motion adaptation effect in the present studies reflects a deeper process in which subjects model an agent's attention, whether visual or auditory attention, rather than reflecting a superficial process in which subjects model the literal gaze direction of an agent. Gaze is likely to be a dominant cue for reconstructing the attention of an agent, and visual attention may still be the most common kind of attention that people model. But even when gaze is blocked with a blindfold, if subjects are

3.6. Experiment 5

It is now well established that seeing a face gaze at an object can redirect one's own attention to that object (Calder et al., 2002; Friesen and Kingstone, 1998; Frischen et al., 2007; Symons et al., 2004). It is therefore possible that in experiment 2, when the subjects saw a head gazing at a tree, their covert attention was automatically drawn to the tree. That bias in attention might have somehow asymmetrically affected their ability to discriminate the direction of the subsequent dot motion stimulus. This explanation is unlikely at the outset for the following reason. The image of the head gazing at the tree was presented for 1.5 s, longer than typical of experiments that use gaze as an attention cue. In previous experiments, the attention effect of a gazing head tends to fade at longer time intervals (Friesen and Kingstone, 1998). Therefore, in the present experiment, a shift of attention caused by the head's gaze is unlikely to extend to the subsequent dot motion stimulus and therefore unlikely to explain the subjects' response to that motion stimulus. We nonetheless tested whether, after 1.5 s of exposure to the head-and-tree stimulus, subjects' attention was spatially biased to one side or the other, thereby potentially affecting responses in the second phase of the trial.

Experiment 5 used the same methods as experiment 2 except in the following ways. After the 1.5 s presentation of the head-and-tree stimulus, instead of a random dot stimulus, a target stimulus appeared either in the position that had just been occupied by the head or by the tree (Fig. 1E). The target was 3.3° wide x 4.0° high and consisted of a black circle with two attached line segments that formed an implied line through the circle (Webb et al., 2016). The line was tilted by 18° from vertical toward the left (half of trials) or right (half of trials). Subjects were instructed to discriminate the orientation of the tilt by pressing the left or right arrow key on a keyboard as quickly as possible. Subjects were allowed a 2 s window in which to respond, and all subjects responded within that window on all trials. As soon as the subject responded, the target disappeared and the inter-trial interval began. On each trial, the head could be on the left or right; the target could be on the left or right; and the target line could be tilted clockwise or counterclockwise. All trial types were balanced and presented in a random order. For analysis, trial types were collapsed into two conditions: target matching the head location, and target matching the tree location. Subjects performed 64 trials in 4 blocks of 16 trials each, thus 32 trials per condition. Twenty-four entirely naïve subjects, untested in any of the other experiments, were tested here (18 females, 18–25 y old). None were excluded from analysis.

If attention at the time of target onset was greater at one location, either at the location of the head or of the tree, then mean RT should have been faster for targets at that location (Webb et al., 2016). A difference score was computed: ΔRT = [average RT to target at the location of the head] - [average RT to target at the location of the tree]. Fig. 2E shows the result. The ΔRT was not significantly different from zero (target at face location RT = 497 ms; target at tree location RT = 496 ms; ΔRT = 1 ms, SE = 9, t_{23} = 0.13, p = 0.901; ΔA = 1.2 %, SE = 1.3, t_{23} = 0.92, p = 0.367). The result shows no evidence that attention was biased toward the head or the tree at the time the head-and-tree stimulus disappeared and a subsequent, task-relevant stimulus appeared. This result strongly suggests that, in experiments 2, 3, and 4, an asymmetric distribution of covert attention as subjects viewed the random dot motion stimulus is not a likely explanation for their pattern of response to the motion stimulus.

3.7. Experiment 6

The goal of experiment 6 was to further rule out the possibility that

used an exogenous attention cue to attract the subjects' attention to one side of the display, and then presented the dot motion stimulus, to determine whether a shift in exogenous attention would affect performance on the dot motion task.

The experiment used the same methods as experiment 2 except in the following ways. Instead of the 1.5 s presentation of the head-and-tree stimulus, a visual cue consisting of a 1.8° by 1.8° black square was presented briefly (50 ms) on either the left or right side, to attract covert exogenous attention (Fig. 1F). After the presentation of the cue, the screen remained blank for 100 ms before the motion discrimination stimulus appeared. Thus, 150 ms was interposed between cue onset and target onset, a time period known to capture consistent exogenous attention effects (Webb et al., 2016). The dot motion stimulus and discrimination task were the same as in experiments 1 and 2. On each trial the cue could be on the left or right, and the dot motion could be toward the left or right. All trial types were balanced and presented in a random order. For analysis, trial types were collapsed into two major conditions: dot motion toward the location where the cue was flashed, and dot motion away from the cue location. Subjects performed 60 trials in 4 blocks of 15 trials each, thus 30 trials per condition. Twenty-five entirely naïve subjects, untested in any of the other experiments, were recruited, out of which one was excluded because of poor performance on the dot motion task (see **Materials and Methods** for exclusion criteria), leaving 24 subjects (11 females, 18–30 y old).

If spatially biasing attention to one or the other side does differentially affect the ability to detect motion in one versus the other direction, then the RT for motion toward the location of the cue should differ significantly from the RT for motion away from the location of the cue. A difference score was computed: ΔRT = [average RT to motion toward the location of the cue] - [average RT to motion away from the location of the cue]. Fig. 2F shows the result. The ΔRT was not significantly different from zero (motion toward cue RT = 687 ms; motion away from cue RT = 691 ms; ΔRT = -4 ms, SE = 8, t_{23} = -0.53, p = 0.604; ΔA = 2.8 %, SE = 1.5, t_{23} = 1.81, p = 0.084). This result corroborates the conclusion that an asymmetric distribution of the subjects' attention cannot easily explain the results of experiments 2 - 4.

3.8. Experiment 7

One possible explanation for the pattern of results in experiment 2 is that participants are subject to a simple directional congruency effect, rather than an implied motion effect. In this alternative explanation, when the subject sees a sighted face looking at the tree, the stimulus does not induce an implied motion from the eyes to the tree, but instead acts as a simple directional cue like an arrow. When a rightward cue is followed by a rightward dot motion, reaction times are somehow increased, and when a rightward cue is followed by a leftward dot motion, reaction times are somehow reduced. In this alternate explanation, the congruency effect is unrelated to implied motion. This explanation is unlikely for two reasons. First, a directional cue ought to have had the opposite effect. If subjects are primed with a rightward cue, for example, they should respond more quickly to a rightward dot motion, not a leftward dot motion. Second, in experiment 2, when the face in the display was blindfolded, no directional congruency effects were obtained, even though a blindfolded face still has a definite directionality, its nose pointed at the tree. To test the directional congruency hypothesis more directly, in experiment 7 we replaced the sighted head of experiment 2 with the image of a black, thick, easily visible arrow pointing to the tree (Fig. 1G), while keeping all other experimental factors the same. We hypothesized that an arrow would provide a clear spatial cue but that people would not perceive implied motion projecting from its tip and flowing across the center of the screen toward

on one side of the display pointing to a tree on the other side. To ensure that it was clearly visible, the arrow was large (4.3° wide x 1.8° high), pointing in an obvious direction and presented in high contrast black against the light gray background. The arrow was on screen for the same duration (1.5 s) and presented at the same visual eccentricity (2.5° from display midline) as the face stimulus in experiment 2. Sometimes the arrow was on the left, pointing to a tree on the right; and sometimes it was on the right, pointing to a tree on the left. In phase 2 of the trial, the arrow-and-tree disappeared and subjects saw a random dot stimulus as in experiment 2, drifting either to the right or to the left. All trial types were balanced and presented in a random order. For analysis, the trial types were collapsed into 2 major conditions: dot motion congruent or incongruent with arrow direction. Subjects performed 60 trials in three blocks of 20 trials each, thus 30 trials per condition. Twenty-five entirely naïve subjects, untested in any of the other experiments, were recruited for experiment 7, of which 1 was excluded for poor performance on the dot motion task (see **Materials and Methods** for exclusion criteria), leaving 24 in the analysis (12 females, 18–22 y old).

Fig. 2G shows the result. When the sighted head was replaced by an arrow pointing at the tree, ΔRT was not significantly greater than zero (congruent RT = 849 ms; incongruent RT = 859 ms; ΔRT = -10 ms, SE = 11, t_{23} = -0.93, p = 0.363; ΔA = 3.2 %, SE = 1.7, t_{23} = 1.87, p = 0.074). The effect was slightly negative, consistent with a directional cueing effect and inconsistent with the adaptation effects observed in experiment 2. This finding suggests that simple spatial compatibility effects cannot explain the pattern of results in experiment 2. The effect of a directional arrow is not the same as the effect of a face gazing at the tree.

3.9. Meta-analysis

We pooled data from all experiments involving conditions in which the attention-induced motion adaption effect should, by hypothesis, be present (eyes open in experiment 2, facing toward tree in experiment 3, and auditory attention in experiment 4). Among these 75 subjects, ΔRT in the attention beam conditions was significantly greater than 0 (ΔRT = 22 ms, SE = 5, t_{74} = 4.54, p = 0.00002). We also pooled all data from experiments involving conditions in which a face should, by hypothesis, not induce a motion adaptation effect (eyes covered in experiment 2, facing away from tree in experiment 3). Among the 48 participants, ΔRT in these control conditions was not significantly different from 0 (ΔRT = -1 ms, SE = 10, t_{47} = -0.09, p = 0.933). Finally, among the 48 participants tested with both an attention beam condition (eyes open in experiment 2, facing toward tree in experiment 3) and a control condition (eyes covered in experiment 2, facing away from tree in experiment 3), we found that the ΔRT in the eyebeam condition (mean = 23 ms, SE = 4) was significantly greater than in the control condition (-1 ms, SE = 10) (t_{47} = 2.29, p = 0.026). These results confirm that the motion adaptation effect is robust and only present in conditions where the social agent is directly attending to the object.

4. Discussion

The most commonly studied aspect of theory of mind is the process by which people reconstruct the contents of other people's minds, such as beliefs, intentions, and emotions (Baron-Cohen, 1997; Wellman, 2018; Wimmer and Perner, 1983). However, theory of mind may also include a less well studied feature, a model of what it means for a mind to actively grasp content – not just a model of the information inside the vessel, but a model of the nature of the vessel itself. The present results suggest that people implicitly model attention – the feature of an agent

motion adaptation paradigm. This finding is consistent with previous studies showing that viewing photographs with an implied motion (e.g., a running animal) activates motion-sensitive areas of the brain (Kourtzi and Kanwisher, 2000; Krekelberg et al., 2003) and is associated with direction-specific adaptation assessed with real-motion test probes (Winawer et al., 2008). The present study shows that the perceived attention of others belongs to the category of stimuli that can cause an apparent motion effect. The subjects in our study were unaware of their response bias, yet it was consistent and replicable.

Other recent studies have suggested a similar general principle that people model not only the contents within each other's minds, but also construct a model of the essential thing that takes hold of information. Pesquita et al. (2016) found that when subjects watched a video of an actor attending to an object, the subjects implicitly encoded whether the actor's attention was drawn to the object exogenously (by the salience of the object) or was directed endogenously (by the actor's own choice). Here again people were constructing an implicit description of an agent's attention – not just information about the object of attention, but information about the attention itself. A recent study of ours (Guterstam et al., 2019) found that people implicitly assumed that the attention of another agent applied a gentle, physical force to the object of attention.

The fact that these studies use very different behavioral paradigms and yet converge on a similar general principle lends some confidence to the findings. The emerging picture is that people construct a descriptive model of the attention of others, that model is constructed implicitly and automatically, sometimes at odds with people's explicit intellectual beliefs, and the model contains some physically incorrect and schematic features. In our interpretation of the data, in the model, other agents are represented as a source of an energy-like essence that attends to information about the world; the essence radiates invisibly through space where the agent directs it; and the essence touches and even physically affects the object of attention. As bizarre as this implicit perception may seem to modern intellectual sensibilities, we suggest that it may represent a computationally easy way for the brain to keep track of sources and targets of attention in a complex social world.

It is of adaptive benefit for the human brain (and perhaps the brains of other species) to construct a model of the attentional state of other agents. The reason is that attention greatly influences behavior, and thus a model of someone else's attention allows for better prediction of that other person's behavior, which, in turn, allows one to coordinate a more adaptive behavioral response to the other person. However, it is important to remember that attention is not simply the direction of someone else's gaze. Attention – physically real attention – is an extraordinarily complex process made up of billions of neurons interacting in the brain. A literally accurate model of someone else's attention would be impossible. First, we do not have direct information on other people's internal physical brain processes, and second, modeling the competitive interactions between billions of someone else's neurons would be prohibitively computationally expensive. We do not look at another person and intuitively understand, "The neurons in her visual system are engaged in a signal competition via lateral inhibitory interactions, resulting, across hierarchical layers, in one particular visual signal being enhanced at the expense of others, allowing the winning signal to dominate cognitive networks." Instead, to model another agent's attention, it is necessary to construct a simpler, more schematic model, one that can be computed on the basis of minimal cues and without impossibly large computational resources. Our data suggest that attention is modeled at least partly like an invisible fluid that flows from the agent to the object of attention. The adaptive benefit may be that the model specifies who is attending, what is attended, and the directional valence between agent and object, in a

physically incorrect – there is no beam coming out of someone else's head – but the inaccuracy of the model is not evolutionarily important. Evolution trends toward models in the brain that are of adaptive benefit to the animal, not models that meet an artificial modern standard of scientific accuracy.

The present findings may also provide clues about the possible neural correlates of this schematic model of attention. 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" (Gould and Vrba, 1982). We speculate that, given the present findings, the preexisting visual motion system may have been used during the evolution of social brain mechanisms for tracking the attention of others. A testable prediction of this hypothesis is that brain activity patterns in motion sensitive areas, such as area MT (Newsome and Pare, 1988; Tootell et al., 1995), should contain directional information not only for visual motion but also for other people's attention.

Finally, we note that this fictional, energy-like attention-essence that people may reflexively attribute to each other resembles the most common human mythologies, across cultures and millennia, about spirit and mind, including beliefs in energy flowing out of the eyes, the evil eye, telekinesis, and a plasma-like spirit that can flow out of the body (Benassi et al., 1979; Dundes, 1992; Gross, 1999; Guterstam et al., 2019; Sidky, 2017; Winer et al., 2002). One possibility is that these beliefs are intuitively compelling to people, and ubiquitous across cultures, because they resonate with the natural, simplifying models that are automatic in our social cognitive machinery (Graziano, 2013).

Declaration of Competing Interest

The authors declare no competing financial interests.

Acknowledgments

References

Baron-Cohen, S., 1997. Mindblindness: An Essay on Autism and Theory of Mind. MIT Press.

Benassi, V.A., Sweeney, P.D., Drevno, G.E., 1979. Mind over matter: perceived success at psychokinesis. J. Pers. Soc. Psychol. 37, 1377–1386.

Brainard, D.H., 1997. The psychophysics toolbox. Spat. Vis. 10, 433–436.

Britten, K.H., Shadlen, M.N., Newsome, W.T., Movshon, J.A., 1992. The analysis of visual motion: a comparison of neuronal and psychophysical performance. J. Neurosci. 12, 4745–4765.

Calder, A.J., Lawrence, A.D., Keane, J., Scott, S.K., Owen, A.M., Christoffels, I., Young, A.W., 2002. Reading the mind from eye gaze. Neuropsychologia 40, 1129–1138.

Carlin, J.D., Calder, A.J., 2013. The neural basis of eye gaze processing. Curr. Opin. Neurobiol. 23, 450–455.

Dundes, A., 1992. The Evil Eye: A Casebook. University of Wisconsin Press.

Friesen, C.K., Kingstone, A., 1998. The eyes have it! Reflexive orienting is triggered by nonpredictive gaze. Psychon. Bull. Rev. 5, 490–495.

Frischen, A., Bayliss, A.P., Tipper, S.P., 2007. Gaze cueing of attention. Psychol. Bull. 133, 694–724.

Glasser, D.M., Tsui, J.M., Pack, C.C., Tadin, D., 2011. Perceptual and neural consequences of rapid motion adaptation. Proc. Natl. Acad. Sci. 108, E1080–E1088.

Gould, S.J., Vrba, E.S., 1982. Exaptation—a missing term in the science of form. Paleobiology 8, 4–15.

Graziano, M.S., 2013. Consciousness and the Social Brain. Oxford University Press.

Graziano, M.S., Kastner, S., 2011. Human consciousness and its relationship to social neuroscience: a novel hypothesis. Cogn. Neurosci. 2, 98–113.

Gross, C.G., 1999. The fire that comes from the eye. Neuroscientist 5, 58–64.

Guterstam, A., Kean, H.H., Webb, T.W., Kean, F.S., Graziano, M.S.A., 2019. Implicit model of other people's visual attention as an invisible, force-carrying beam pro-

Kourtzi, Z., Kanwisher, N., 2000. Activation in human MT/MST by static images with implied motion. J. Cogn. Neurosci. 12, 48–55.

Krekelberg, B., Dannenberg, S., Hoffmann, K.-P., Bremmer, F., Ross, J., 2003. Neural correlates of implied motion. Nature 424, 674–677.

Levinson, E., Sekuler, R., 1980. A two-dimensional analysis of direction-specific adaptation. Vision Res. 20, 103–107.

Newsome, W.T., Pare, E.B., 1988. A selective impairment of motion perception following lesions of the middle temporal visual area (MT). J. Neurosci. 8, 2201–2211.

Perrett, D.I., Smith, P.A.J., Potter, D.D., Mistlin, A.J., Head, A.S., Milner, A.D., Jeeves, M.A., 1985. Visual cells in the temporal cortex sensitive to face view and gaze direction. Proc. R. Soc. Lond., B, Biol. Sci. 223, 293–317.

Pesquita, A., Chapman, C.S., Enns, J.T., 2016. Humans are sensitive to attention control when predicting others' actions. Proc. Natl. Acad. Sci. 113, 8669–8674.

Pilly, P.K., Seitz, A.R., 2009. What a difference a parameter makes: a psychophysical comparison of random dot motion algorithms. Vision Res. 49, 1599–1612.

Sidky, H., 2017. The Origins of Shamanism, Spirit Beliefs, and Religiosity: A Cognitive Anthropological Perspective. Lexington Books.

Symons, L.A., Lee, K., Cedrone, C.C., Nishimura, M., 2004. What are you looking at? Acuity for triadic eye gaze. J. Gen. Psychol. 131, 451–469.

Tootell, R.B., Reppas, J.B., Dale, A.M., Look, R.B., Sereno, M.I., Malach, R., Brady, T.J., Rosen, B.R., 1995. Visual motion aftereffect in human cortical area MT revealed by functional magnetic resonance imaging. Nature 375, 139–141.

Webb, T.W., Kean, H.H., Graziano, M.S., 2016. Effects of awareness on the control of attention. J. Cogn. Neurosci. 28, 842–851.

Wellman, H.M., 2018. Theory of mind: the state of the art. Eur. J. Dev. Psychol. 15, 728–755.

Wicker, B., Michel, F., Henaff, M.-A., Decety, J., 1998. Brain regions involved in the perception of gaze: a PET study. NeuroImage 8, 221–227.

Wimmer, H., Perner, J., 1983. Beliefs about beliefs: representation and constraining function of wrong beliefs in young children's understanding of deception. Cognition 13, 103–128.

Winawer, J., Huk, A.C., Boroditsky, L., 2008. A motion aftereffect from still photographs depicting motion. Psychol. Sci. 19, 276–283.

Winer, G.A., Cottrell, J.E., Gregg, V., Fournier, J.S., Bica, L.A., 2002. Fundamentally misunderstanding visual perception: adults' belief in visual emissions. Am. Psychol. 57, 417.

9