Using Smiles, Frowns, and Gaze to Attribute Conscious States to Others: Testing Part of the Attention Schema Theory

Branden J. Bio and Michael S. A. Graziano

Branden J. Bio: Department of Psychology; Princeton University, Princeton NJ, 08544; Email: bbio@princeton.edu

Michael S. A. Graziano: Department of Psychology; Princeton Neuroscience Institute; Princeton University, Princeton NJ, 08544; Email: graziano@princeton.edu

Key Words: Attention; Awareness; Gaze; Emotional expression; Social cognition; Theory of mind

Abstract: In the attention schema theory, people attribute the property of consciousness to themselves and others because it serves as a schematic model of attention. Most of the existing literature on monitoring the attention of others assumes that people primarily use the gaze direction of others. In that assumption, attention is not represented by a deeper model, but instead limited mainly to a single, externally visible parameter. Here we presented subjects with two cues about the attentional state of a face: direction of gaze and emotional expression. We tested whether people relied predominantly on one cue, the other, or both when deciding if the face was conscious of a nearby object. If the traditional view is correct, then the gaze cue should dominate. Instead, some people relied on gaze, some on expression, and some on an integration of cues, suggesting that a variety of surface strategies could inform a deeper model. We also assessed people's social cognitive ability using two, independent, standard tests. If the traditional view of attention monitoring is correct, then the degree to which people use gaze to judge attention should correlate best with their social cognitive ability. Instead, social cognitive ability correlated best with the degree to which people successfully integrated the cues together. The results strongly suggest that when people attribute a specific state of consciousness to another, rather than simply tracking gaze, they construct a model of attention, or an attention schema, that is informed by a combination of surface cues.

Introduction

In the attention schema theory, the human brain constructs a simplified, schematic model, both descriptive and predictive, to represent attention [1-3]. The model leads us to attribute consciousness to ourselves and others. The theory has two branches. In one branch, the brain constructs a model of its own attention to monitor, predict, and help control attention [3-5]. In the other branch, the brain constructs a model of the attention of other people to help in social cognition [6-11]. According to the theory, the two types of model depend on partially overlapping mechanisms in the brain [6,11-14]. The present study focuses on the second branch of the theory, modeling the attention of other agents.

Modeling the attention of others is of central importance to theory of mind [2,15,16]. The perceptual, emotional, and cognitive content of a person's mind at any one moment is almost entirely determined by the focus of that person's attention. Therefore, having a good model of what attention is, what its dynamics are, what impact it has on other cognitive processes, and how to reconstruct it from available cues, is fundamental to good social cognition. Most previous work on the tracking of someone else's attention has focused on the perception of gaze direction [16-23]. Gaze is a major external cue about a person's attention. However, attention is a deeper process than directing the eyes toward something, and is often entirely dissociated from gaze. A person can focus covert, selective attention on an object in the visual periphery, on a sound, on a thought, or on a recalled memory [24-26]. Moreover, if person A is to reconstruct the attention of person B, and person A is blind, or listening to person B over the phone, or if person B is wearing dark glasses or is engaging in social gaze aversion, then the reconstruction of attention must be more complex than registering gaze direction. According to the attention schema theory, people do not merely register gaze direction as a proxy for attention; instead, they construct a deeper, complex model of someone else's attention that is constrained by multiple surface cues.

As a useful analogy, consider the body schema [27,28]. It was once assumed that knowledge of the state of one's arm is simply a matter of registering the somatosensory signals that indicate joint angle. A century of work has shown how much deeper the process really is [27,28]. The brain constructs a rich model or simulation of the arm that has both descriptive and predictive components. The model is constrained by the convergence of many cues, including somatosensory signals, visual signals, motor feedback, and contextual knowledge. If two cues are put in opposition, the body schema works to integrate them. One of the most compelling examples is the Pinocchio illusion [29]. If you close your eyes and touch your nose with your right finger, and if someone applies a 120 Hz vibration to your right biceps muscle, two contradictory signals are generated. The sensation from the muscle tells you that your elbow is extending. The sense of touch tells you that your fingertip is still in contact with your nose. Under this circumstance, different people integrate the cues in different ways. Some people suppress the muscle signal and experience no illusion; some combine the cues and report feeling as though their finger is lengthening; and some report feeling as though their nose is lengthening. Opening the eyes immediately eliminates any illusion, because the added visual signal further constrains the body schema. This type of cue integration demonstrates the presence of an integrative model that is deeper than the tracking of any single variable.

In the case of social attention, do people monitor a single variable – gaze – or do they construct an attention schema that is similar to the body schema – a model that is constrained by multiple cues? In the present study, we presented subjects with two cues about the attentional state of a face: direction of gaze and emotional expression. The cues sometimes aligned and sometimes conflicted. We asked people to judge whether the face was aware of an object, and tested whether people relied predominantly on one cue, the other, or an integration of both.

The paradigm was based on one introduced in a brain imaging study conducted previously by our group [6] and is shown in Figure 1. Subjects viewed a face next to an object. The face could look toward or away from the object, the face could look happy or alarmed, and the object could be of positive valence (such as a birthday cake) or of negative valence (such as a bloody knife). Two cues to the state of the cartoon person's mind were therefore available: first, whether the eyes pointed toward or away from the object, and second, whether the emotional expression matched or mismatched the valence of the object. Subjects were asked to judge whether the face looked "not aware," "somewhat aware," or "very aware" of the object. After subjects completed this Attribution-of-Awareness Task, they were given two independent, standard assessments of social cognitive ability: the Reading the Mind in the Eyes Test (RMET) [30,31] and the Hinting Task [32,33].

The purpose of the study was to evaluate two contrasting hypotheses. In the traditional hypothesis, the social cognitive mechanism is tuned to rely primarily on gaze as a proxy for someone else' attention. If this hypothesis is correct, though the emotional expression cue may influence subjects' judgments, the gaze cue will dominate. Moreover, people who rely more on the gaze cue (and who therefore demonstrate a better-tuned social cognitive mechanism) should score higher on the RMET and Hinting Task, which measure general social cognitive ability. Social cognitive ability should be less well correlated with other strategies, such as reliance on the expression cue or on an integration of the two cues.

In the attention schema hypothesis, in contrast, people monitor each other's attention by constructing a deeper model that is fundamentally integrative. The gaze cue may still be of importance. Convincing evidence suggests that the brain is especially well tuned to detect the gaze of others [17-23]. But the brain is also well tuned to detect facial expression [34-41], and the processing of gaze and facial expression may interact [36,42,43]. In the attention schema hypothesis, the gaze cue should not dominate. As in the case of the Pinocchio illusion, our paradigm should reveal a variety of ways that people can use the surface cues to inform the deeper model, including a reliance on gaze, on facial expression, or on an integration of the two. Moreover, social cognitive ability should be best correlated with the integration of the two cues, and not with the use of any one cue.

Methods

Participants

All participants provided informed consent and all procedures were approved by the Princeton Institutional Review Board. We recruited 831 human volunteers through Amazon Mechanical Turk (229 females, aged 18-70, mean age = 37 years, SD = 10) and paid them $0.10 per minute. We chose online testing, a practical way to test a large sample size, because a major goal of the study was to assess the statistical range and clustering of different strategies among the population. We could not know how many different strategies might be present or what proportion of subjects might engage in each, and therefore we chose a large sample size. A disadvantage of online testing is that, outside of a controlled lab environment, people do not always pay sufficient attention or engage fully with the task. To limit this problem, we eliminated subjects whose responses indicated reduced engagement with the experiment. First, we eliminated 89 subjects because they completed fewer than 80% of the trials on the Attribution-of-Awareness Task. The Hinting Task provided a second criterion for eliminating subjects, because it required responses typed in English. We eliminated a further 154 subjects whose responses in the Hinting Task suggested an inattentive or disengaged participant (e.g. any possible word salad or nonsensical repetition of the instruction text). With these criteria, 243 of the 831 subjects were eliminated, leaving a total of 588 subjects. Eliminating a high percentage of subjects on the basis of possible inattention to the task is common for online experiments. On subsequent testing, we found that whether these subjects were eliminated or added to the data changed levels of variance but did not materially change the overall trends in the data.

Stimulus Validation

Prior to using the Attribution-of-Awareness Task, we validated stimuli in an independent sample of 56 subjects recruited through Amazon Mechanical Turk. In the validation test, volunteers rated the appropriateness of the expression of a face looking at a set of 96 objects. On each trial, an object was shown with a cartoon head facing it. Objects included foods, animals, tools, landscapes, depictions of events, and so on. The cartoon either had a happy expression or an agitated expression when looking at the object. Participants were then asked to choose on a 3-point scale whether the face's expression was "inappropriate," "neither inappropriate nor appropriate," or "appropriate" for the object at which it was looking. The 30 most unambiguously positive objects (rated most highly appropriate with a happy expression and most inappropriate with an agitated expression) and 30 most unambiguously negative objects (rated most highly appropriate with an agitated expression and most inappropriate with a happy expression) were included in the final task.

Attribution-of-Awareness Task

The Attribution-of-Awareness Task was adapted from Kelly *et al.* [6]. Participants were asked to judge the extent to which an agent on screen was aware of an object (see Figure 1). Note that whereas the figure shows the relative sizes and positions of images in the display, the absolute size in degrees of visual angle could not be controlled because of the online testing procedures. On each trial, a cartoon head was displayed on the right side of the screen simultaneously with a picture of an object on the left. Participants were asked to "respond whether you think the cartoon face looks not aware, somewhat aware, or very aware of the object." Judgments were converted to a numerical scale by coding "not aware" as -1, "somewhat aware" as 0, and "very aware" as 1.

For half of the trials, the cartoon's eyes were directed toward the object, and for the other half of trials the eyes were directed away from the object. The expression of the face was happy for half of trials and agitated for half of trials. The object had a positive valence for half of trials and a negative valence for half of trials. This fully counterbalanced, 2X2X2 design was collapsed into four conditions of interest for the purpose of analysis: gaze directed away from the object paired with facial expression incongruent with the valence of the object (G-E-); gaze away, expression congruent (G-E+); gaze toward, expression incongruent (G+E-); and gaze toward, expression congruent (G+E+).

The Attribution-of-Awareness Task consisted of 240 trials (60 trials for each of the four main conditions) over 8 blocks. Subjects were allowed to rest at the end of each block and begin the next block when ready. All trial types were fully counterbalanced and their order was randomized. The task could be completed within approximately fifteen minutes. The Attribution-of-Awareness Task was always presented to subjects first, followed by the two social cognition tests.

Reading the Mind in the Eyes Task

After subjects performed the Attribution-of-Awareness Task, they were asked to complete a test of their general aptitude for social cognition. The Reading the Mind in the Eyes Task (RMET) has been widely validated as a reliable way to assess theory of mind [30,31]. The RMET consists of 36 photographs of faces restricted to the eye area. Subjects label the photos with an emotion from a small set of options given on each trial. The task measures a subject's ability to read the mental state of others given limited social cues. For each correct matching of a picture to an emotional state, one point is awarded, up to a maximum of 36. The RMET could be completed within approximately ten minutes.

Hinting Task

After completing the RMET, subjects were given a second social cognition task, the Hinting Task [32,33]. We began use of the Hinting Task part-way through the experiment, to add to the completeness of the cognitive assessment, and thus of the 588 subjects included in the analysis, 411 performed the Hinting Task. All analyses concerning the Hinting Task therefore represent that subset of the subjects.

The Hinting Task used here was an adapted version for American participants, altered from the original British version to include more familiar vocabulary and examples [33]. The task consists of 10 short written scenarios in which two individuals interact and one individual implies something to the other. For example: "Scott and Eric are coworkers who enter the elevator at the same time. Scott's arms are filled with paperwork. He says to Eric, 'Sixth floor please, I'm late for a meeting.'" Subjects are then asked a question probing what was implied by the characters, such as: "What does Scott really mean when he says this?" Each question earns 2 points for a correct answer. In previous versions, subjects were given a second chance to respond to each question and earn points, but given the online format here, we chose to limit subjects to one try at each question. Thus, the maximum number of points that could be earned was 20. The Hinting Task could be completed within approximately ten minutes and was the last of the three tests that subjects performed.

Results

K-Means Clustering for the Attribution-of-Awareness Task

In the Attribution-of-Awareness Task, for each subject, four numbers were computed, corresponding to the subject's mean responses in the four task conditions. Different subjects showed different patterns across the four means. As an initial analysis to gain a better understanding of these response patterns, we employed the commonly used, k-means clustering method [44]. K-means clustering is an unsupervised learning algorithm for finding natural categories within multi-dimensional data. In the present case, the procedure sorted participants into five clusters, as shown in Figure 2. It is important to keep in mind that the k-means clustering does not represent an absolute division of the data into cleanly separable subpopulations, but provides only an initial, roughly descriptive account of the data. A more hypothesis-driven analysis, taking into account the variance among individual subjects, is presented in subsequent sections.

Figure 3 shows the mean response pattern for each of the five k-means clusters. In the first cluster, subjects responded with a similar, elevated mean rating across all four trial conditions. These subjects tended to judge the faces to be more often "very aware" than "not aware," regardless of whether the gaze was aimed toward or away from the object or whether the facial expression matched or mismatched the valence of the object. These subjects apparently did not rely consistently on the gaze or expression cues, since our manipulation of those cues did not affect mean response. The subjects presumably noticed and may have used the two cues, but must have done so inconsistently, thus resulting in the flat average. The cluster contained 131 subjects (22.2% of total).

The second cluster of subjects in Figure 3 resembled the first. The subjects did not rely in a consistent manner on the gaze or the expression cue, since the average rating was similar regardless of cue configuration. However, the average rating was lower in this second cluster than in the first. The cluster contained 151 subjects (25.6% of total).

The third cluster of subjects in Figure 3, unlike the first two, showed evidence of relying in a more consistent manner on the available cues. These subjects appeared to rely more on the gaze of the face (whether gaze was aimed toward or away from the object) than on the expression of the face (whether the expression matched or mismatched the valence of the object). When the gaze was turned away from the object (G-E- and G-E+, first two bars shown in Figure 3, cluster 3), the subjects rated the face as unaware more often than aware of the object; and when the gaze was turned toward the object (G+E- and G+E+, second two bars shown in Figure 3, cluster 3), the subjects rated the face as aware more often than unaware of it. This cluster contained 100 subjects (17.0% of total).

The fourth cluster of subjects shown in Figure 3 appeared to rely more on the emotional expression of the face (whether the expression was congruent or incongruent with the valence of the object) than on the gaze. When the expression mismatched the valence of the object (G-E- and G+E-, first and third bars in Figure 3, cluster 4), these subjects rated the face as more unaware than aware; and when the expression matched the valence of the object (G-E+ and G+E+, second and fourth bars in Figure 3, cluster 4), the subjects rated the face as more aware than unaware. This cluster contained 115 subjects (19.5% of total).

Finally, the fifth cluster of subjects followed a pattern consistent with relying roughly equally on both the gaze and expression cue, integrating the two to inform judgments. These subjects rated the face as mostly unaware of the objects only when both gaze and expression cues mismatched the object (G-E-, first bar in Figure 3, cluster 5); they rated the face as mostly aware of the object only when both gaze and expression cues matched the object (G+E+, fourth bar in Figure 3, cluster 5); and when the two cues conflicted (G-E+ and G+E-, second and third bars in Figure 3, cluster 5), the subjects judged the face at an intermediate level of awareness, on average closer to unaware than aware. Cluster five contained 91 subjects (15.4% of total).

In summary, the gaze cue did not dominate. The results are not consistent with gaze as the proxy signal that drives social attention. Instead, the results are consistent with subjects constructing a deeper model of the face's awareness of the object, and informing that deeper model through surface cues. Just as in the case of the Pinocchio illusion, where conflicting cues could be used in a variety of ways to inform the body schema [29], here different subjects used different cue strategies, and among those subjects who relied on the cues, the three main strategies were roughly equally represented.

RMET And Hinting Task Performance

The purpose of including the RMET and Hinting Task in the present study was to provide independent assessments of social cognitive ability and to determine whether that ability was in any way correlated with the cue-utilization strategies found in the Attribution-of-Awareness Task. Before describing how these many tasks were correlated, we first describe the results of the RMET and Hinting Task individually. The average score on the RMET was 21.1, with a standard deviation of 8.36 and a range from 5-36. Performance on the RMET in this study was roughly similar to, though lower than, performance in past studies (Fertuck *et al.* [45] mean=25.00, standard deviation=3.63; Jankowiak-Siuda *et al.* [46] mean=24.98, standard deviation=4.53; Kynast *et al.* [47] mean=23.40, standard deviation =10.53).

The average score on the Hinting Task was 10.31, with a standard deviation of 8.18 and a range from 0-20. The Hinting Task scores collected here are not comparable to scores from previous studies due to simplifications that we introduced to the scoring of this dataset (see **Methods**).

Although the RMET and Hinting Task are independent measures of social cognition and assess participants in very different ways, one focusing on the visual inspection of eyes and emotional expression, and the other focusing on hidden intentions and written stories, scores for the two assessments were highly significantly correlated (r^2=0.66, F=807.8, p=7.14e-99).

RMET and Hinting Task S*cores for Each K-Means Cluster*

As described above, subjects were sorted into five clusters based on their performance on the Attribution-of-Awareness Task. Figure 4 shows how each of these five clusters scored on the two social cognition tests (see also Table 1). The RMET scores are shown in Figure 4A and the Hinting Task scores are shown in Figure 4B.

Note that the five clusters form two larger, overarching groups. The first group could be called the "non-strategy" subjects (clusters 1 and 2), or subjects who did not use the available visual cues in a consistent manner. Subjects in this non-strategy group performed comparatively poorly on both of the social cognition tasks. It is possible that these subjects represent the low end of the social cognition spectrum; but it is also possible that many of these subjects were unengaged with the task. In the Discussion, we address this possibility further.

The second large group could be called the "strategy" subjects (clusters 3, 4, and 5), or subjects who used the available cues in a consistent, systematic manner. The strategy group performed significantly better than the non-strategy group on both of the social cognition tests (two tailed Welch's t test: for RMET, t = 29.7, p = 7.23e-119; for Hinting Task, t = 22.12, p = 1.29e-69).

Before describing the next, more detailed analysis, here we note some interpretational limits to the initial analysis involving clustering. The clusters of subjects described here cannot be considered to cleanly represent distinct response strategies. Though each cluster may, on average, emphasize one response pattern more than another, the subjects appear to utilize the cues in a complex and mixed manner. Dividing the subjects into a small number of clusters and averaging within each cluster, though providing a useful initial picture, is not sufficient for understanding the data. The following section describes a more detailed analysis.

Does Better Social Cognitive Ability Predict a Specific Cue-Utilization Strategy?

The traditional view of how people monitor each other's attention implies that use of the gaze strategy should correlate best with general social cognitive ability. In contrast, the attention schema hypothesis suggests that an integrator strategy should correlate best with social cognitive ability. To test these possibilities, we first constructed three metrics to quantify how much each individual subject relied on a gaze strategy, an expression strategy, or a cue-integration strategy. Figure 5 shows the three hypothetical, idealized response strategies. Consider Figure 5A first. Here the hypothetical strategy relies on the gaze cue only. When gaze is away from the object (first two bars, Figure 5A), the face is rated as totally unaware of the object, with a rating of -1. When gaze is toward the object (second two bars, Figure 5A), the face is rated as very aware of the object, with a rating of +1. A simple way to quantify how much each real subject relies on a pure gaze strategy is to compute a difference score between the subject's response and this idealized response. For each of the four conditions shown in Figure 5A, the subject's own mean score was subtracted from the idealized score. Each of these four numbers was then squared and added together. The result was a sum-of-squares deviation score, a single score for each subject that varied between 0 and 16, for which 0 indicates that the subject perfectly matched the idealized response pattern of using the gaze cue only, 8 indicates that the subject's strategy was orthogonal to or totally unrelated to the gaze strategy, and 16 indicates that the subject was as far as mathematically possible from that response pattern [gaze difference score = (-1 - subject's mean score for G-E- condition)^2 + (-1 - subject's mean score for G-E+ condition)^2 + (1 - subject's mean score for G+E- condition)^2 + (1 - subject's mean score for G+E+ condition)^2].

Similarly, Figure 5B shows an idealized response pattern that relies only on facial expression to solve the task. Using the same method as for the gaze difference score, an expression difference score was computed for each subject. Once again, the lower the score, the better the subject matched the idealized, facial expression strategy [expression difference score = (-1 - subject's mean score for G-E- condition)^2 + (1 - subject's mean score for G-E+ condition)^2 + (-1 - subject's mean score for G+E- condition)^2 + (1 - subject's mean score for G+E+ condition)^2].

Finally, Figure 5C shows an idealized response pattern that relies on an integration of the two cues to solve the task, weighting the two cues equally. Using the same method as for the gaze difference score, an integrator difference score was computed for each subject. The lower the score, the better each subject matched this integrator response strategy [integrator difference score = (-1 - subject's mean score for G-E- condition)^2 + (0 - subject's mean score for G-E+ condition)^2 + (0 - subject's mean score for G+E- condition)^2 + (1 - subject's mean score for G+E+ condition)^2].

With these three scores computed for each subject (gaze difference score, expression difference score, and integrator difference score), we could perform correlation analyses, asking whether people who adhered relatively more closely to a particular cue-utilization strategy would also tend to score higher on the two social cognition tests. The results are shown in Figure 6 (see also Tables 2 and 3).

We first briefly summarize the overall pattern of results and then describe the details. Overall, subjects' social cognitive ability, as measured by performance on the two social cognition tests, was significantly correlated with the use of all three strategies in the Attribution-of-Awareness Task. However, it was much better correlated to the strategy of integrating the two cues together than it was to the alternative strategies of relying on any one cue by itself. The correlation between the RMET score and the integrator strategy accounted for 44% of the intersubject variance (r^2=0.44, see Table 2) as compared to 13% for the gaze strategy and 11% for the expression strategy. Similarly, the correlation between the Hinting Task score and the integrator strategy accounted for 43% of the variance in the data as compared to 14% for the gaze strategy and 9% for the expression strategy (see Table 3). The traditional hypothesis of a dominance of the gaze strategy was not confirmed; instead, the attention schema hypothesis was supported.

We examined these relationships in greater detail, to better understand the difference between the gaze strategy, expression strategy, and integrator strategy. First consider Figure 6A, which shows the relationship between the gaze difference score and the RMET score. Each data point represents a single subject. Overall, a clear relationship can be seen. As subjects scored better on the RMET (higher scores on the X axis), they also showed greater use of the strategy of relying on the gaze cue when performing the Attribution-of-Awareness Task (lower scores on the Y axis). This correlation is highly significant (r^2=0.13, F=89.78, p=6.50e-20). Note that the r^2 value is relatively low (accounting for 13% of the variance) but the relationship is reliable, resulting in a small computed p value. However, even a quick inspection of the data in Figure 6A shows more complexity than a simple linear relationship. The first half of the data along the X axis is obviously different from the second half. To help quantify that pattern, we divided the subjects into two groups: those that scored at or below the median on the RMET (lower half of the range on the X axis, median = 20), and those that scored above the median (upper half of the range on the X axis). The two groups show different behavior. Within the low-social-scoring group, a clear correlation was present: subjects who performed better in the RMET tended to rely more on the gaze cue in the Attribution-of-Awareness task. This correlation was statistically significant (r^2=0.045, F=14.46, p=0.00017). In contrast, within the high-social-scoring group (subjects who scored above the median on the RMET), no significant correlation was found anymore between the RMET score and the gaze difference score (r^2 =0.0005, F=1.52, p=0.218).

Figure 6B shows a similar result for the relationship between the expression difference score and the RMET score. The subset of subjects who were relatively poor at social cognition (at or below the median on the X axis) showed a significant correlation. Among those subjects, better performance on the RMET predicted greater reliance on the expression cue in the Attribution-of-Awareness task (r^2=0.046, F=14.73, p=0.00015). The subset of subjects who were relatively better at social cognition (above the median on the X axis) did not show a significant correlation. Among these subjects, better performance on the RMET did not predict more reliance on the expression cue (r^2=0.001, F=0.25, p=0.614).

Figure 6C shows the result for the integrator difference score. Here the pattern was different. The data no longer showed a clear division between those subjects who were below the median and those who were above. The same trend applied to both halves of the data. The subset of subjects who were relatively poor at social cognition (at or below the median on the RMET) showed a significant correlation, in which better performance on the RMET was associated with greater reliance on the strategy of integrating the two cues together in the Attribution-of-Awareness task (r^2=0.10, F=33.83, p=1.52e-08). The subset of subjects who were relatively better at social cognition (above the median on the RMET) also showed a significant correlation in which better performance on the RMET was associated with greater reliance on integrating the two cues together (r^2=0.034, F=9.79, p=0.0019).

The results shown in Figure 6D, E, and F show a similar pattern with respect to the Hinting Task (see also Table 3).

In summary, scores on the two social cognition tests were by far better correlated with a reliance on an integrator strategy than with any other strategy. The more subjects relied on the integrator strategy, the better their social cognition scores tended to be. This overall pattern appeared to be partly related to a split in the data between low-social-scoring subjects and high-social-scoring subjects. Among the low-social-scoring subjects, social cognition scores were significantly correlated with all three cue-utilization strategies, though the correlation was still two to three times greater for the integrator strategy. Among the high-social-scoring subjects, social cognition scores were significantly correlated only with the integrator strategy.

Discussion

Monitoring the attention of others is fundamental to all social cognition. Most of the literature on social attention assumes that the gaze direction of others serves as a proxy for the attention of others, and that the social cognitive machinery is tuned specifically to rely on gaze [15-23]. A weakness of this approach is that it trivializes the representation of other people's attention, reducing it to a single parameter that is visually tracked. In contrast, in the attention schema theory, people construct a deeper model of the attention of others, and the model can be informed by more than one cue. Here we tested whether the more standard assumption, or the attention schema hypothesis, is correct. We tested how people combine two cues, a gaze cue and an emotional expression cue, to judge the awareness of a cartoon face. Sometimes the cues were aligned and sometimes they were in conflict with each other. In direct contradiction to the gaze-dominance hypothesis, we found a variety of cue-utilization strategies. Some subjects did not rely on the cues in a consistent manner and simply found the face to be generally aware of the nearby object. Some subjects relied mainly on the gaze cue, much as in the traditional view of attention monitoring. Some subjects relied on the emotional expression of the face and ignored the gaze cue. Some relied more heavily on a strategy of integrating the two cues together, weighing them against each other. The range of strategies suggests a deeper model that can be constrained in various ways by surface cues.

We also tested subjects on two independent assessments of social cognitive ability: the RMET and the Hinting Task. In both cases, we found the same pattern of results. Social cognitive ability, as measured by performance on the two social cognition tests, was better correlated to the strategy of integrating the two cues together than it was to the alternative strategies of relying on any one cue by itself. Neither test was best correlated to the use of the gaze cue.

To gain greater insight into the data, we separately examined subjects who performed poorly and who performed well on the two social cognition tests. Among those subjects who performed relatively poorly on the social cognition tests, a better social cognition score was significantly correlated with a greater use of any of the cue strategies (though the correlation was still two to three times stronger for the integrator strategy). By implication, if you are bad at social cognition, then an ability to use any cue and any strategy will help. But among those subjects who were above the median in their social cognitive ability, only one of the cue-utilization strategies was significantly associated with better social cognition scores – the strategy of integrating the two cues together. Looking at Figure 6A, B, and C, and focusing on the upper end of the X axis range, one sees the result especially starkly. Good social cognition is not especially associated with reliance on the gaze of others (Figure 6A) as the dominant cue; nor is it associated with reliance on the emotional expression of others (Figure 6B); instead, good social cognition is associated with reliance on a strategy of integrating cues together (Figure 6C).

These results directly contradict the most common assumptions about how humans monitor the attention of other humans. The direction of someone's gaze is apparently not automatically used as the dominant, proxy signal for that person's attention. A person's attention is a complex, hidden property, not equatable to any simple external feature. Instead, the results point to a deeper process. The social mechanism evidently constructs an integrative model, and that model can be informed by more than one cue.

The gaze cue is clearly an important one, with special representation in the brain [15-23]. The emotional expression of faces is also obviously an important feature with special representation in the brain [34-41]. We do not argue that any one cue is unimportant or does not have special neuronal representation. Our argument here is that the representation of other people's attention is not the same as a representation of other people's gaze. It lies at a deeper level. It can be informed by both gaze and expression cues, and perhaps by other cues. The point of the present experiment is also not to search out all cues or define all attributes of the model, but to establish the basic principle that social attention is deeper than the gaze cue, and that the process involves integration of cues to inform a model.

Other studies have looked at interactions between gaze and facial expression, but have done so in a different manner [42,43]. In those previous studies, expression was not set up to provide specific information about attentional state. Instead, gaze was the primary cue to attention and facial expression was a secondary, modulating factor. In the present study, the expression of the face could match or mismatch an object, just as the gaze could match or mismatch the object. In this way, facial expression and gaze were potentially equally informative about whether the face was attending to the object. We suggest that in real-life situations, though gaze may indeed often be an important cue to someone else's attention, sometimes other cues are also informative. A person's speech and body language, tone of voice and facial expression, and the general context itself, can provide strong cues to attention. Most importantly, reconstructing attention is not about tracking a simple parameter, but, we suggest, about constructing a deeper model and attributing a mind state to others.

An alternative interpretation of at least some of the present results is worth considering. In both the k-means clustering analysis and the regression analysis, a difference was found between subjects who performed relatively poorly across all tasks and those who performed relatively better. One possibility is that some subjects genuinely have poor social cognition, and consequently are not as consistent or strategic in their use of the two cues. However, it is also possible that, given the nature of online testing, some subjects were inattentive or unmotivated and therefore performed poorly across tasks. It is extremely difficult to disentangle these possibilities. It is likely that at least some subjects were disengaged, contaminating the data set. However, we do not believe that inattentive subjects and random responses represent a major interpretational problem to the data. We offer the following reasons.

First, to address exactly this concern, we eliminated a high percentage of subjects who demonstrated a possible lack of engagement with the tasks (see **Methods**).

Second, even if all poorly-performing subjects are eliminated entirely from the analysis, the conclusions of the study remain. When we considered only the high-performing subjects – those who performed above the median on the social cognition tests – this select sample was presumably the least likely to include unmotivated, inattentive subjects. It was among this group that the clearest result emerged: social cognitive ability was significantly correlated with a cue-integration strategy and was not significantly correlated with the strategy of relying separately on the gaze cue or on the expression cue.

Acknowledgements

Supported by the Princeton Neuroscience Institute Innovation Fund. Thanks to Elyssa Barrick for her suggestions and support with analysis and data visualizations. We also thank Jemma Bio-Barrick for her encouragement.

Data Availability:

The data in this study are available at: http://arks.princeton.edu/ark:/88435/dsp01zw12z8390

References

1. M. S. A. Graziano, S. Kastner, Human consciousness and its relationship to social neuroscience: A novel hypothesis. Cognitive Neuroscience 2, 98-113 (2011).

2. M. S. A. Graziano, *Consciousness and the Social Brain.* New York: Oxford University Press (2013).

3. T. W. Webb, MS. Graziano, 2015. The attention schema theory: a mechanistic account of subjective awareness. Frontiers in Psychology 6, 500 (2015).

4. T. W. Webb, H. H. Kean, M. S. A. Graziano, Effects of awareness on the control of attention. Journal of Cognitive Neuroscience 28, 842-851 (2016).

5. A. I. Wilterson, C. M. Kemper, N. Kim, T. W. Webb, A. M. W. Reblando, M. S. A. Graziano, sAttention control and the attention schema theory of consciousness. Progress in Neurobiology 195, doi: 10.1016/j.pneurobio.2020.101844 (2020).

6. Y. T. Kelly, T. W. Webb, J. D. Meier, M. J. Arcaro, M. S. A. Graziano, Attributing awareness to oneself and to others. Proceedings of the National Academy of Sciences USA 111, 5012-5017 (2014).

7. A. Pesquita, C. S. Chapman, J. T. Enns, Humans are sensitive to attention control when predicting others' actions. Proceedings of the National Academy of Sciences USA 113, 8669-8674 (2016).

8. A. Guterstam, H. H. Kean, T. W. Webb, F. S. Kean, M. S. A. Graziano, An implicit model of other people's visual attention as an invisible, force-carrying beam projecting from the eyes. Proceedings of the National Academy of Sciences USA 116, 328-333 (2018).

9. M. S. A. Graziano, We are machines that claim to be conscious. Journal of Consciousness Studies 26, 95-104 (2019).

10. A. Guterstam, M. S. A. Graziano, Visual motion assists in social cognition. Proceedings of the National Academy of Sciences USA, published online, DOI 10.1073/pnas.2021325117 (2020).

11. A. Guterstam, B. J. Bio, A. I. Wilterson, M. S. A. Graziano, Temporo-parietal cortex involved in modeling one's own and others' attention. Elife 10: e63551, doi: 10.7554/eLife.63551 (2021).

12. T. W. Webb, K. Igelström, A. Schurger, M. S. A. Graziano, Cortical networks involved in visual awareness independently of visual attention. Proceedings of the National Academy of Sciences USA 113, 13923-13928 (2016).

13. A. I. Wilterson, S. A. Nastase, B. J. Bio, A. Guterstam, M. S. A. Graziano. Attention, awareness, and the right temporoparietal junction. Proceedings of the National Academy of Sciences USA 118, e2026099118. doi: 10.1073/pnas.2026099118. PMID: 34161276 (2021).

14. B. J. Bio, T. W. Webb, M. S. A. Graziano. Projecting one's own spatial bias onto others during a theory-of-mind task. Proceedings of the National Academy of Sciences USA 115, E1684-E1689 (2018).

15. S. Baron-Cohen, Mindblindness: An Essay on Autism and Theory of Mind. Cambridge, MA: MIT Press (1997).

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

17. D. I. Perrett, P. A. J. Smith, D. D. Potter, A. J. Mistlin, A. S. Head, A. D. Milner, M. A. Jeeves, Visual cells in the temporal cortex sensitive to face view and gaze direction. Proceedings of the Royal Society of London, B: Biological Sciences 223, 293-317 (1985).

18. H. Kobayashi, S. Kohshima, Unique morphology of the human eye. Nature 387, 767-768 (1997).

19. C. K. Friesen, A. Kingstone, The eyes have it! Reflexive orienting is triggered by nonpredictive gaze. Psychonomic Bulletin and Review 5, 490-495 (1998).

20. A. Puce, T. Allison, S. Bentin, J. C. Gore, G. McCarthy, Temporal Cortex Activation in Humans Viewing Eye and Mouth Movements. Journal of Neuroscience 18, 2188-2199 (1998).

21. 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).

22. E. A. Hoffman, J. V. Haxby, Distinct representations of eye gaze and identity in the distributed human neural system for face perception. Nature Neuroscience 3, 80 (2000).

23. A. Frischen, A. P. Bayliss, S. P. Tipper, Gaze cueing of attention. Psychological Bulletin 133, 694-724 (2007).

24. D. M. Beck, S. Kastner, Top-down and bottom-up mechanisms in biasing competition in the human brain. Vision Research 49, 1154-1165 (2009).

25. M. M.Chun, J. D. Golomb, N. B. Turk-Browne, A taxonomy of external and internal attention. Annual Review of Psychology 62, 73-101 (2011).

26. T. Moore, M. Zirnsak, Neural Mechanisms of Selective Visual Attention. Annual Review of Psychology 68, 47-72 (2017).

27. M. S. A. Graziano, M. M. Botvinick, How the brain represents the body: insights from neurophysiology and psychology. In: *Common Mechanisms in Perception and Action: Attention and Performance XIX.* Eds. W. Prinz, B. Hommel. Oxford, UK: Oxford University Press, 136-157 (2002).

28. N. P. Holmes, C. Spence, The body schema and the multisensory representation(s) of peripersonal space. Cognitive Processing 5, 94-105 (2004).

29. J. R. Lackner, Some proprioceptive influences on the perceptual representation of body shape and orientation. Brain 111, 281-297 (1988).

30. S. Baron-Cohen, T. Jolliffe, C. Mortimore, M. Robertson, Another advanced test of theory of mind: evidence from very high functioning adults with autism or Asperger Syndrome. Journal of Child Psychology and Psychiatry 38, 813-822 (1997).

31. S. Baron-Cohen, S. Wheelwright, J. Hill, Y. Raste, I. Plumb, The "Reading the Mind in the Eyes" Test Revised Version: A Study with Normal Adults, and Adults with Asperger Syndrome or High-functioning Autism. Journal of Child Psychology and Psychiatry 42, 241-251 (2001).

32. R. Corcoran, G. Mercer, C. D. Frith, Schizophrenia, symptomatology and social inference: Investigating "theory of mind" in people with schizophrenia. Schizophrenia Research 17, 5–13 (1995).

33. T. C. Greig, G. J. Bryson, M. D. Bell, Theory of Mind Performance in Schizophrenia: Diagnostic, Symptom, and Neuropsychological Correlates. The Journal of Nervous and Mental Disease 192, 12-18 (2004).

34. P. Ekman, Facial expression and emotion. American Psychologist 48, 384-392 (1993).

35. T. Ganel, K. F. Valyear, Y. Goshen-Gottstein, M. A. Goodale, The involvement of the 'fusiform face area' in processing facial expression. Neuropsychologia 43, 1645-1654 (2005).

36. A. D. Engell, J. V. Haxby, Facial expression and gaze-direction in human superior temporal sulcus. Neuropsychologia 45, 3234-3241 (2007).

37. C. J. Fox, S. Y. Moon, G. Iaria, J. J. Barton, The correlates of subjective perception of identity and expression in the face network: an fMRI adaptation study. Neuroimage 44, 569-580 (2009).

38. X. Xu, I. Biederman, Loci of the release from fMRI adaptation for changes in facial expression, identity, and viewpoint. Journal of Vision 10, 36 (2010).

39. A. Achaibou, E. Loth, S. J. Bishop, Distinct frontal and amygdala correlates of change detection for facial identity and expression. Social Cognitive and Affective Neuroscience 11, 225-233 (2016).

40. R. B. Adams Jn., D. N. Albohn, K. Kveraga, Social Vision: Applying a Social-Functional Approach to Face and Expression Perception. Current Directions in Psychological Science 26, 243-248 (2017).

41. Y. Li, R. M. Richardson, A. S. Ghuman, Posterior Fusiform and Midfusiform Contribute to Distinct Stages of Facial Expression Processing. Cerebral Cortex 29, 3209-3219 (2019).

42. J. K. Hietanen, J. M. Leppänen, Does Facial Expression Affect Attention Orienting by Gaze Direction Cues? Journal of Experimental Psychology: Human Perception and Performance 29, 1228-1243 (2003).

43. E. Hori, T. Tazumi, K. Umeno, M. Kamachi, T. Kobayashi, T. Ono, H. Nishijo, Effects of facial expression on shared attention mechanisms. Physiology and Behavior 84, 397-405 (2005).

44. J. MacQueen, Some methods for classification and analysis of multivariate observations. In: L. M. Le Cam, J. Neyman, Eds, *Berkeley Symposium on Mathematical Statistics and Probability*, Vol 5.1, pp. 281-297 (1967).

45. E. A. Fertuck, A. Jekal, I. Song, B. Wyman, M. C. Morris, S. T. Wilson, B. S. Brodsky, B. Stanley, Enhanced 'Reading the Mind in the Eyes' in borderline personality disorder compared to healthy controls. Psychological Medicine 39, 1979-1988 (2009).

46. K. Jankowiak-Siuda, S. Baron-Cohen, W. Bialaszek, A. Dopierala, A. Kozlowska, K. Rymarczyk, Psychometric evaluation of the "Reading the Mind in the Eyes" test with samples of different ages from a Polish population. Studia Psychologica 58, 18-31 (2016).

47. J. Kynast, E. M. Quinque, M. Polyakova, T. Luck, S. G. Riedel-Heller, S. Baron-Cohen, A. Hinz, A. V. Witte, J. Sacher, A. Villringer, M. L. Schroeter, Mindreading from the eyes declines with aging–evidence from 1,603 subjects. Frontiers in Aging Neuroscience 12:550416. doi: 10.3389/fnagi.2020.550416 (2020).

	RMET			Hinting		
	N	Mean	SE	N	Mean	SE
Cluster 1	134	13.00	0.45	89	2.88	0.63
Cluster 2	148	14.32	0.36	109	5.03	0.59
Cluster 3	100	25.72	0.62	72	15.61	0.59
Cluster 4	115	26.72	0.54	78	16.10	0.60
Cluster 5	91	27.38	0.44	63	16.70	0.47

Table 1: Performance of five clusters of subjects on two social cognition tests. For definition of clusters, see Figure 3. N = number of subjects tested, SE = standard error, RMET = Reading the Mind in the Eyes Test, Hinting = Hinting Task.

	All			< M			> M		
	R^2	F	P	R^2	F	P	R^2	F	P
Gaze	0.13	89.78	6.50e-20	0.045	14.46	0.00017	0.005	1.524	0.218
Expression	0.11	72.30	1.54e-16	0.046	14.73	0.00015	0.001	0.2543	0.614
Integrator	0.44	456.10	2.80e-75	0.100	33.83	1.52e-08	0.034	9.787	0.0019

Table 2: Relationship between the RMET score and the three cue-utilization strategies. Gaze = the strategy of relying on the gaze cue, as quantified by the gaze difference score. Expression = the strategy of relying on the expression cue, as quantified by the expression difference score. Integrator = the strategy of relying on an integration of both cues, as quantified by the integrator difference score. All = all subjects, ≤M = the subset of subjects who scored at or below the median on the RMET, >M = the subset of subjects who scored above the median on the RMET.

	All			< M			> M		
	R^2	F	P	R^2	F	P	R^2	F	P
Gaze	0.14	64.59	1.00e-14	0.16	44.08	2.27e-10	0.003	0.5844	0.446
Expression	0.09	39.94	6.84e-10	0.11	28.39	2.38e-07	0.009	1.682	0.196
Integrator	0.43	303.30	3.19e-51	0.32	104.80	1.73e-20	0.073	14.060	0.00024

Table 3: Relationship between the Hinting Task score and the three cue-utilization strategies. Gaze = the strategy of relying on the gaze cue, as quantified by the gaze difference score. Expression = the strategy of relying on the expression cue, as quantified by the expression difference score. Integrator = the strategy of relying on an integration of both cues, as quantified by the integrator difference score. All = all subjects, ≤M = the subset of subjects who scored at or below the median on the Hinting Task, >M = the subset of subjects who scored above the median on the Hinting Task.

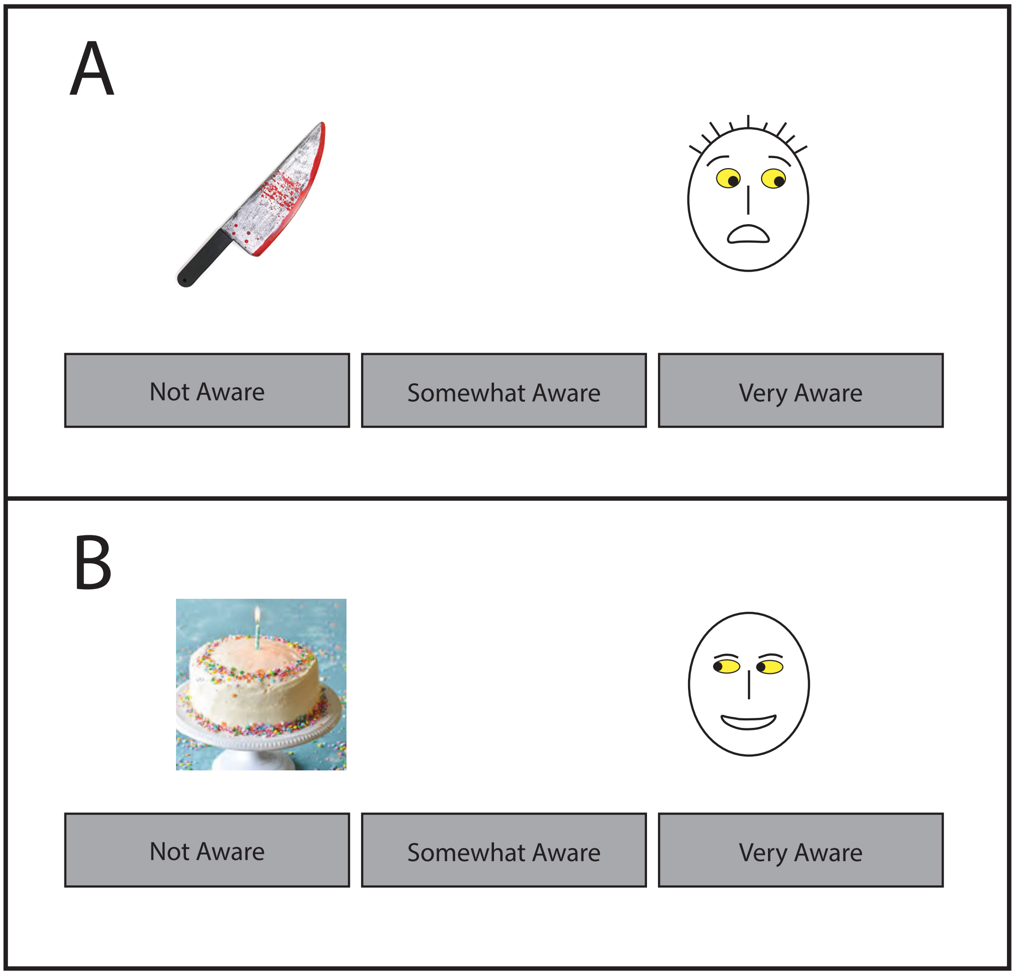

Figure 1. Paradigm for the Attribution-of-Awareness Task. On each trial subjects saw a cartoon face next to an object. The face could look toward or away from the object, the object could have positive or negative valence, and the face could have a happy or alarmed expression. The result was a 2X2 design: gaze toward or away and expression congruent or incongruent. Subjects rated whether the face seemed not aware, somewhat aware, or very aware of the object. Here two examples of the four possible categories are shown. Top: gaze away, facial expression congruent. Bottom: gaze toward, facial expression congruent.

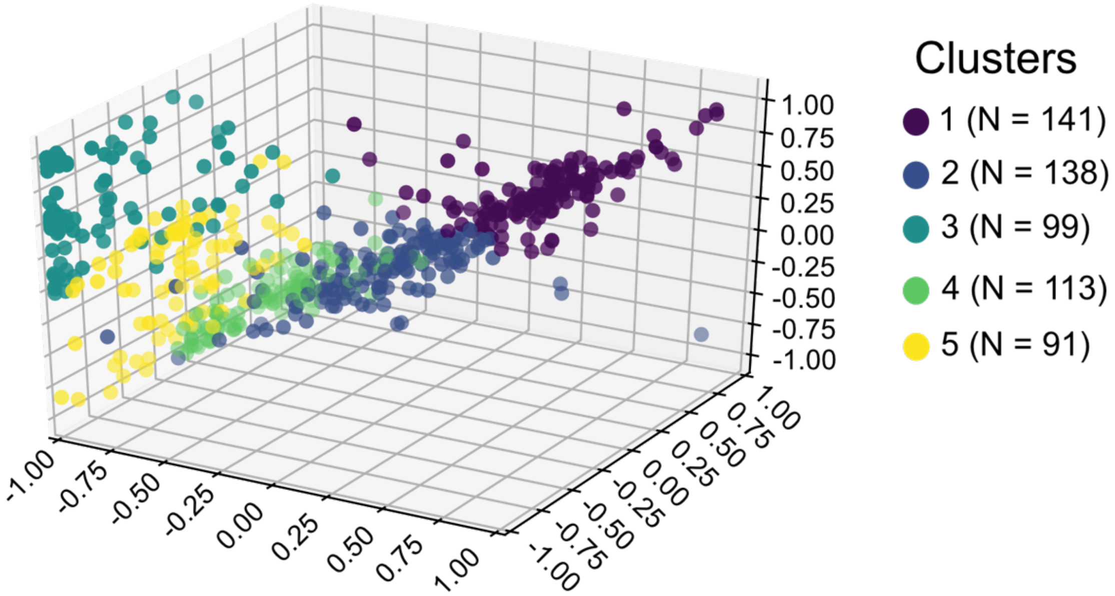

Figure 2. K-means clustering results for the Attribution-of-Awareness task. Each data point represents a single subject. While only three axes are visually displayed here, the data are actually distributed in four dimensions, corresponding to the four means obtained for each subject. The clusters are therefore less overlapping and more separable than they appear here. The five colors represent the five k-means clusters.

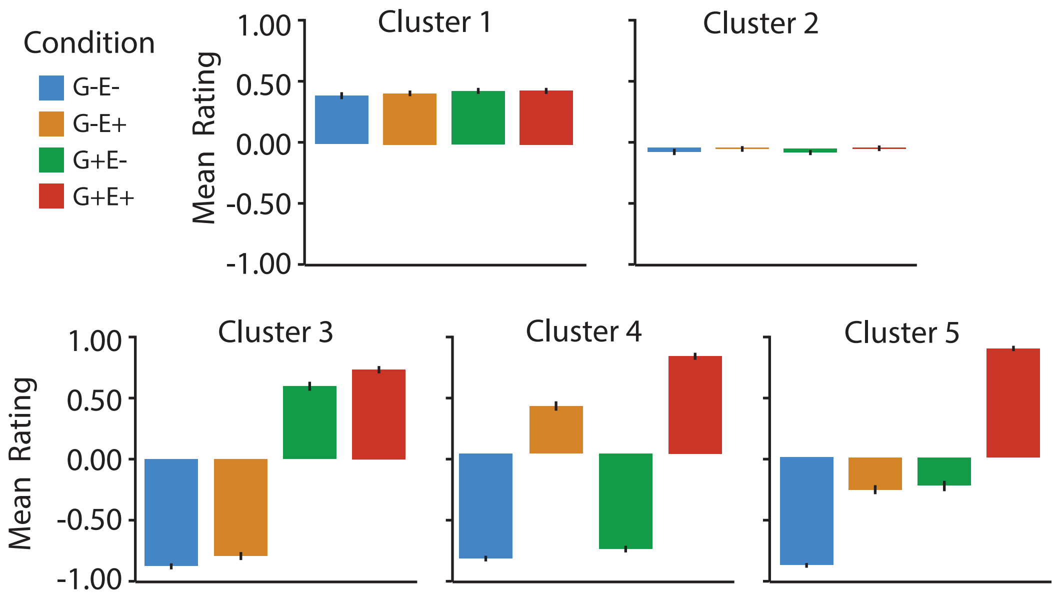

Figure 3. Five clusters of subjects. Cluster 1 subjects were characterized by similar, elevated ratings across all conditions. Cluster 2 subjects were characterized by similar, intermediate ratings across all conditions. Cluster 3 subjects were characterized by ratings mainly dependent on whether the gaze matched or mismatched the object. Cluster 4 subjects were characterized by ratings mainly dependent on whether the expression matched or mismatched. Cluster 5 subjects were characterized by ratings dependent on an integration of gaze and expression. Bars show mean ratings among subjects in the four task conditions. Error bars = standard error. G-E- = task condition in which the gaze is away from the object and the expression mismatches the object. G-E+ = task condition in which the gaze is away and the expression matches the object. G+E- = task condition in which the gaze is toward and the expression mismatches. G+E+ = task condition in which the gaze is toward and the expression matches.

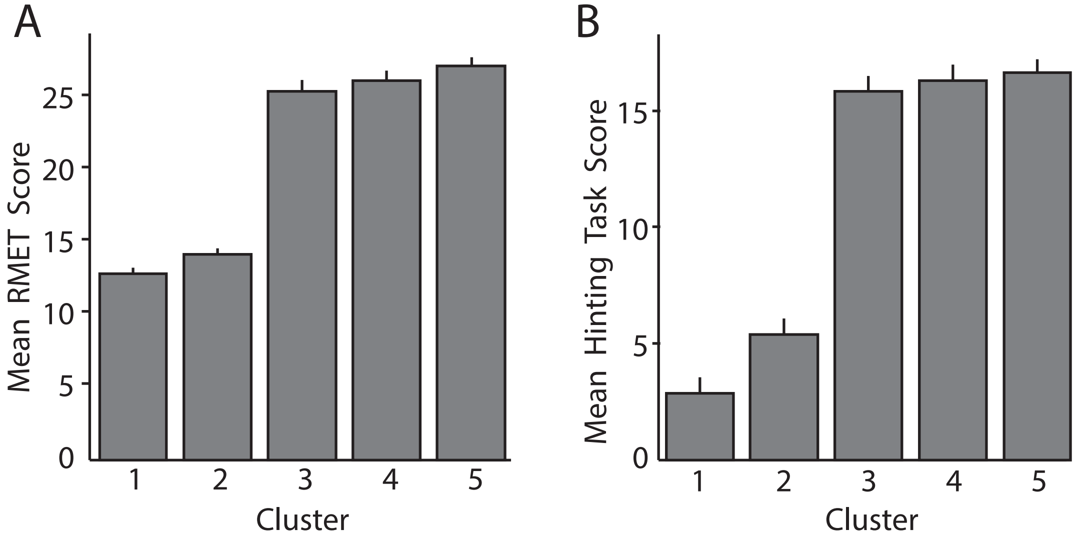

Figure 4. Social cognition scores by cluster. A. Average score on the RMET for each of the five clusters of subjects. Error bars = standard error. B. Average score on the Hinting Task for each of the five clusters of subjects. Error bars = standard error.

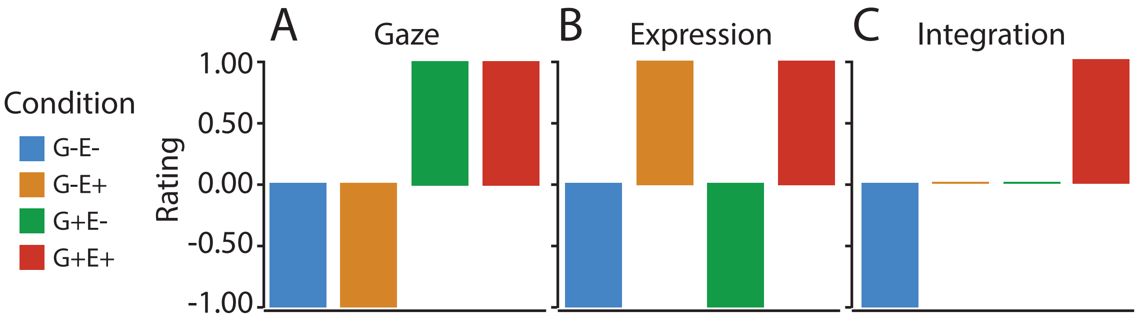

Figure 5. Three idealized strategies. A. The idealized strategy for a hypothetical subject who relies entirely on the gaze cue and ignores the expression cue. B. The idealized strategy for a hypothetical subject who relies entirely on the expression cue and ignores the gaze cue. C. The idealized strategy for a hypothetical subject who relies on integration and equal weighting of both the gaze and expression cues.

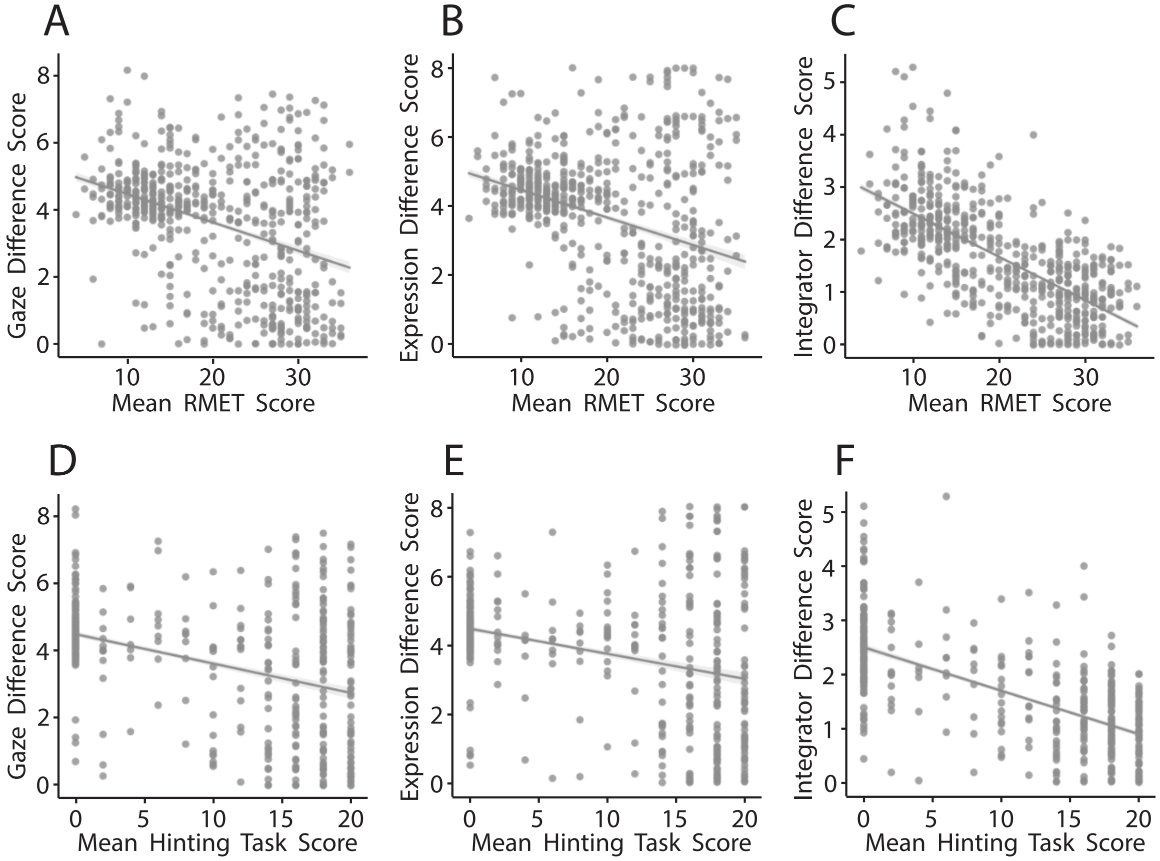

Figure 6. Relationship between social cognitive ability and cue-utilization strategy. A. For each subject, the score on the RMET (a measure of social cognitive ability) is plotted on the X axis and the gaze difference score (lower scores indicate more reliance on the gaze strategy) is plotted on the Y axis. Shading around regression line shows standard error. B. RMET score versus expression difference score. C. RMET score versus integrator difference score. D. Score on the Hinting Task (an independent measure of social cognitive ability) versus the gaze difference score. E. Hinting Task score versus expression difference score. F. Hinting Task score versus integrator difference score.