

Running head: Theory of mind and attentional bias to emotional expressions

Theory of mind and attentional bias to emotional expressions: A preliminary study

Luisa A. Ribeiro

University College London

Pasco Fearon

University of Reading

Abstract

Theory of mind ability has been associated with performance in interpersonal interactions and has been found to influence aspects such as emotion recognition, social competence, and social anxiety. Being able to attribute mental states to others requires attention to subtle communication cues such as facial expressions. Decoding and interpreting emotions expressed by the face, especially those with negative valence, are essential skills to successful social interaction. The current study explored the association between theory of mind skills and attentional bias to emotional faces. According to the study hypothesis, individuals with poor theory of mind skills showed preferential attention to negative faces over both non-negative faces and neutral objects. Tentative explanations for the findings are offered emphasizing the potential adaptive role of vigilance for threat as a way of allocating a limited capacity to interpret others' mental states to obtain as much information as possible about potential danger in the social environment.

Keywords: theory of mind, attentional bias, emotional facial expressions

Theory of mind is the ability to attribute mental states to oneself and others (Premack & Woodruff, 1978). This capacity has also been referred to as *mentalizing capacity* (Morton, Frith, & Leslie, 1991), *social intelligence* (Baron-Cohen et al., 1997) and *reflective-function* (Fonagy & Target, 1997). The ability to perceive, understand, and reason about one's own and others' behaviors in terms of mental states is a crucial skill for interpersonal interactions. In fact, poor theory of mind skills have been associated with difficulties in interpersonal interaction, in areas such as emotion recognition (e.g., Ashwin, Chapman, Colle, & Baron-Cohen, 2006), social competence (Bosacki & Astington, 1999) and generalized anxiety in social and family life (e.g., Coupland, 2001).

Being able to attribute mental states to others requires awareness of minimal cues regarding non-verbal communication of emotions, such as attention to facial expressions (e.g., Baron-Cohen, Wheelwright, & Jolliffe, 1997). Much of the work on information processing of facial expressions has focused on the selective processing of positive and negative valence emotions (Beck, 1987; Mathews & MacLeod, 1986). Emotional biases for facial expressions have been most frequently demonstrated by using the Stroop colour-naming task (Power & Dalgleish, 2008), in which participants have to name the colour of a word as fast as possible while ignoring its meaning. By using a modified version of this task, Mathews and MacLeod (1985) were the first to demonstrate that, relative to normal controls, anxiety-disordered patients are slower in colour-naming threat compared to neutral words. More recently, another type of task has been used to study attentional bias: the dot-probe task. In its original version, developed by MacLeod, Mathews, and Tata (1986), pairs of words, such as threat and neutral words, are briefly showed to the participant. Immediately after, a small dot probe occurs in the location just occupied by one of the words. The rationale for the task is that individuals are faster to

respond to probes that are presented in an attended, rather than unattended, region of a visual display (Posner, Snyder, & Davidson, 1980). Typically, anxious individuals are faster to respond to probes that replace threat rather than neutral stimuli, consistent with the hypothesis of *vigilance for threat* (e.g., Broadbent & Broadbent, 1988; MacLeod et al., 1986). However, due to their lack of ecological validity, single words have been replaced with more socially relevant stimuli such as emotional facial expressions. For example, Bradley et al. (1998) used a modified version of MacLeod, Mathews and Tata's (1986) dot-probe task, including facial expressions as stimuli.

The aim of this study is to explore the relation between theory of mind skills and attentional bias to emotional faces. The goal is to detect any overall tendency to direct attention towards or away from faces (irrespective of valence), as well as to identify more specific biases in the processing of negative versus non-negative emotional expressions in individuals with good and poor theory of mind ability. It is hypothesized that individuals with poor theory of mind skills, like anxious individuals, will be more vigilant for negative emotional expressions when compared to individuals with good theory of mind skills.

Method

Participants

A total of 46 (27 female) adults aged 18 to 52 years ($M = 25.6$, $SD = 6.7$) participated in the study. Participants were paid volunteers recruited at the University College London (UCL) campus by advertisement.

Apparatus and Materials

Reading the Mind in the Eyes Test (Baron-Cohen, et al., 2001). This test was conceived as a measure of how well individuals can put themselves into the mind of other

people and "tune in" to their mental state. It is therefore an advanced theory of mind task for adults designed to assess individuals' ability to make inferences about other people's mental states. Participants are shown a series of 36 photographs of the eye-region of the face of different people, and are asked to choose which of four mental state terms best describes what the person in the photograph is thinking or feeling (e.g., decisive, amused, aghast, and bored). The photographs are all black and white and represent an equal number of male and female eyes, which are presented in a random order. Participants' responses are classified as correct or incorrect according to an ideal response set established on the basis of judgment consensus in a pilot study conducted by Baron-Cohen and colleagues (Baron-Cohen et al., 2001). A computer-based version of the study was used. This task also includes a glossary of all the mental state terms used, which participants are encouraged to consult before the experiment begins. If during the task participants are still unsure of a mental state term meaning, they can go over it by placing the cursor under the targeted word. The Eyes Test was found to be inversely correlated with the Autism Spectrum Quotient (AQ), a measure of autistic traits in adults of normal intelligence (Baron-Cohen et al., 2001).

Facial Emotion Dot-Probe Task

This computer-based task is a modified version of the dot-probe task used by Mansell et al. (1999) to investigate face attention in high and low socially anxious individuals. The task uses a sequence of pairs of photographs of facial emotional expressions as target stimuli and pictures of household objects as neutral stimuli. The facial-stimuli consist of 30 pictures of 6 different faces (3 male and 3 female Caucasian individuals) displaying negative (angry, fearful and sad) and non-negative (neutral and happy) emotional expressions. The facial images used were taken from the facial emotion

stimuli set developed by Matsumoto and Ekman (1988). The 30 pictures of faces were paired with photographs of 30 different household objects (e.g., table, clock, phone). All the pictures were edited to fit an upright rectangle measuring about 352 x 250 pixels and the face-object pairs were matched by eye for brightness and contrast. During the task, the pairs of pictures (face-object) are presented diagonally on the screen (for 1000ms) in one of two possible orientations: top right and bottom left, or top left and bottom right. The position of the two pictures informs participants of the two possible locations of the probe, which appears 500ms after the presentation of the pictures. The probe display consists of a letter “E” or “F” appearing in a location that corresponds to the centre of one of the pictures. Participants have to indicate which letter was displayed by pressing either the “E” key with the left forefinger or the “F” key with the right forefinger. The probe letter remains on the screen until participants’ response, after which the next trial begins. Across the 120 trials, each emotion appears in each location an equal number of times and the probe also appears in the same location as the target stimuli an equal number of times. The target stimuli are presented in a random order for each participant. The time elapsed between the presentation of the stimulus pictures and the key-response to the probe is taken as the outcome variable used to calculate *attentional bias* scores.

Procedure

The tasks were administered on desktop computers and participants were assigned to individual cubicles for the duration of the experiment. Participants were given the opportunity to read an information sheet providing a description of the computer-based tasks composing the experiment, as well as an explanation about the scientific rationale behind them. Confidentiality issues were addressed and participants were reassured that all the research data collected was to be kept separately from information concerning

personal identity. Participants were also informed that they were free to withdraw from the study at any time and were asked to provide written consent. After providing information regarding demographic characteristics (age and gender), participants began the experiment with the "Reading the Mind in the Eyes Test", which took approximately 10 minutes to complete. Instructions were received from a computer display. Having completed this task, participants were given written instructions for the "Facial Emotion Dot-Probe Task", which also took 10 minutes to complete. For both tasks, participants were encouraged to respond as quickly as possible but without sacrificing accuracy. After being paid the standard UCL rate for healthy volunteers, participants were debriefed and thanked for their participation.

Results

Data preparation

Participants were divided into two groups according to their performance in the "Reading the Mind in the Eyes Test". Thus, individuals were assigned to either a high or low performance group depending on whether the number of correct answers to the test was in the upper or lower median range ($n = 27$ high performance group; $n = 19$ low performance group). For the Facial Emotion Dot-Probe Task, the data analysis was based on probe detection latencies (reaction times). For each participant, outliers were removed by excluding extremely fast (less than 200 ms) and extremely slow (greater than 1000 ms) detection latencies because they typically indicate anticipatory responses and momentary inattention. Mean probe detection latencies were calculated for each type of emotional face, negative (angry, fearful, sad) and non-negative (neutral and happy), and for each probe location (probe appearing in the location of the face vs. probe appearing in the location of the object). Mean errors for each face valence and probe location were

also calculated and subjected to a separate analysis. In order to simplify the data, attentional bias scores, summarising the interaction effect of location of emotional face with probe location, were computed. Hence, for both types of emotional faces (negative and non-negative) bias scores were calculated by subtracting the mean reaction time when the emotional face and probe were in the same position, from the mean reaction time when the emotional face and the probe were in different positions. The formula by MacLeod and Mathews (1988) was used: Bias score = $0.5 \times (ELPU + EUPL - EUPU - ELPL)$ (where ELPU corresponds to the detection latency for the Emotion occurring in the Lower area and the Probe occurring in the Upper area, and so on). Positive bias scores reflect selective attention towards the emotion (*vigilance*) and negative bias scores reflect attention bias away from the emotion (*avoidance*) (Mansell et al., 1999).

Demographic variables

No correlation between age and eyes test performance was found $r = -.08, p > .05$. Also, there was no association between age and biases to negative ($r = .10, p > .05$) or non-negative ($r = .01, p > .05$) facial expressions. Moreover, men and women did not differ significantly in terms of eyes test scores, $t(44) = .63, p > .05$, bias for negative expressions, $t(44) = 1.9, p > .05$, or bias for non-negative expressions, $t(44) = .21, p > .05$.

Dot-probe task

To determine whether face valence and probe location originated different error rates in this task, a 2 x 2 (negative / non-negative emotion vs. probe appearing in the location of the face / probe appearing in the location of the object) repeated measures analysis of variance (ANOVA) of mean errors was carried out. There was no significant main effect of face type, $F(1, 45) = .002, p > .05$ and no significant main effect of probe

location, $F(1, 45) = .706, p > .05$. Moreover, no interaction "valence x probe location" was found, $F(1, 45) = 1.11, p > .05$.

Another 2 x 2 (negative / non-negative emotion vs. probe appearing in the location of the face / probe appearing in the location of the object) repeated measures analysis of variance (ANOVA) was conducted in order to test the effect of face type (valence) and probe location on probe detection latencies. The main effect of face valence was non-significant, $F(1, 45) = .172, p > .05$ and the main effect of probe location was also non-significant, $F(1, 45) = 2.07, p > .05$. A significant interaction "face valence x probe location" was found, $F(1, 45) = 4.36, p < .05$. Participants tended to be significantly faster when responding to the probe appearing in the location of the face than when responding to the probe appearing in the location of the object, but only for negative faces, $F(1, 45) = 5.06, p < .05$. For non-negative faces, no significant differences were observed in the mean probe detection latencies between the two probe locations, $F(1, 45) = .051, p > .05$

Dot-probe task and "Reading the Mind in the Eyes Test".

Error rates were examined across face valence, probe location and Eyes task performance. A 2 x 2 x 2 mixed design analysis of variance (ANOVA) of mean errors was carried out with one between-subjects factor (high vs. low performance in the Eyes task) and two within-subjects factors: face valence (negative vs. non-negative) and probe location (probe appearing in the location of the face vs. probe appearing in the location of the object). No main effects of face valence, $F(1, 44) = .012, p > .05$, or probe location, $F(1, 44) = .442, p > .05$, were found. Also, there were no two-way interactions "face valence x Eyes task performance", $F(1, 44) = .79, p > .05$, or "probe location x Eyes task performance", $F(1, 44) = .86, p > .05$. No significant interaction "face valence x probe location x Eyes task performance" was found, $F(1, 44) = 1.51, p > .05$.

A 2 x 2 x 2 mixed design analysis of variance (ANOVA) of probe detection latencies data was carried out with one between-subjects factor (high vs. low performance in the Eyes task) and two within-subjects factors: face valence (negative vs. non-negative) and probe location (probe appearing in the location of the face vs. probe appearing in the location of the object). No main effect of face valence was found significant, $F(1, 44) = .28, p > .05$; there was also no main effect of probe location, $F(1, 44) = 2.21, p > .05$. The two-way interactions "face valence x eyes task performance" and "probe location x eyes task performance" were non-significant, $F(1, 44) = .52, p > .05$ and $F(1, 44) = .21, p > .05$, respectively. The two-way interaction "face valence x probe location" remained significant, $F(1, 44) = 6.16, p < .05$. Also, a three-way interaction "face valence x probe location x eyes task performance" was found, $F(1, 44) = 4.18, p < .05$. For faces with negative valence, mean probe detection latencies were significantly shorter, $F(1, 44) = 6.31, p < .05$, when the probe appeared in the location of the face than when the probe appeared in the location of the object, but only for participants scoring low in the Eyes task (see figs.1 and 2). Participants who scored high in the Eyes task did not show significantly different probe detection latencies when responding to the probe appearing in the location of the negative face or in the location of the object, $F(1, 44) = .74, p > .05$. For faces with non-negative valence, there were no significant differences in mean detection latencies between the two probe locations, for both Eyes task performance groups, $F(1, 44) = .088, p > .05$ (low performance), $F(1, 44) = .29, p > .05$ (high performance).

To simplify the previous three-way interaction, a 2 x 2 mixed design analysis of variance (ANOVA) of attentional bias scores was conducted with one between-subjects factor (high vs. low performance in the Eyes task) and one within-subjects factor

(negative faces vs. non-negative faces). There was a significant main effect of face valence, $F(1, 44) = 6.16, p < .05$, with negative faces yielding greater bias scores (*vigilance*). An interaction "face valence x performance in the Eyes task" was also found, $F(1, 44) = 4.18, p < .05$ (see figs. 3 and 4). Participants scoring low in the Eyes task showed a highly significant difference in bias scores to negative vs. non-negative faces, $F(1, 44) = 8.72, p < .01$. An attention bias towards negative faces (*vigilance*) and an attention bias away from non-negative faces (*avoidance*) were found in this group. No significant differences were found, $F(1, 44) = .12, p > .05$, between attention bias scores to negative and attentional bias scores to non-negative faces, among participants with high performance in the Eyes task.

Discussion

In the current study, attentional biases to negative and non-negative facial emotions were compared between individuals with good and poor theory of mind skills. Individuals with poor theory of mind skills showed preferential attention to negative faces over both non-negative faces and neutral objects while individuals with good theory of mind skills showed no such preference. Similarly, when attentional biases to faces with negative and non-negative valence were examined, it was found that individuals with poor theory of mind paid differential attention to negative vs. non-negative faces. They tended to look towards negative faces and away from non-negative faces. Again, individuals with good theory of mind skills showed no differential attention to negative and non-negative faces (they tend to look towards both). One possible explanation for the findings, in line with the study hypothesis, is that vigilance for negative emotions is adaptive in the presence of a poor theory of mind. In fact, given that individuals with poor theory of mind skills have difficulties in reasoning about other people's mental

states, it would be adaptive for them to selectively attend to negative faces as a way of obtaining information about potential danger in the social environment. In fact, in the absence of a good ability to differentiate subtleties in terms of facial expressions, it is probably preferable to use a limited theory of mind capacity to direct attention to negative rather than to non-negative facial stimuli, as misinterpretations of negative expressions (e.g. inability to differentiate anger from sadness) are more likely to have undesirable consequences than misinterpretations of non-negative expressions.

An alternative explanation can be offered for the fact that only individuals with poor theory of mind skills showed differential attentional bias to negative and non-negative faces. Attentional bias towards negative faces (and away from non-negative faces) might have interfered with theory of mind task performance. In fact, it is possible that some individuals in this group obtained a low performance in the Eyes task due to their preferential attention to negative rather than non-negative faces, which might have contributed to increase the number of errors in this task. In fact, it would have been interesting to analyse attentional bias in relation to error rates in the Eyes task in order to investigate, for instance, whether individuals with differential vigilance for negative versus non-negative faces are the ones showing greater error rates in certain states of mind included in the Eyes task. This would have made possible to assess whether the Eyes task worked as a "pure" theory of mind test or if its results were confounded by an attentional bias towards negative faces (and away from non-negative faces).

Attentional bias observed in the dot-probe task suggest a general tendency for participants to attend more to facial emotion expressions than to neutral objects, although this tendency was only significant for faces with negative valence (angry, fearful, and sad). Faces with non-negative valence (neutral and happy) did not seem to be

significantly more attended to than neutral objects. Hence, it appears that individuals participating in this study showed a greater vigilance for negative facial expressions, tendency also previously found in individuals with high levels of trait anxiety (e.g., Mansell et al., 1999; Bradley et al., 1998). Hence, the attention "preference" for negative over non-negative faces found in the current study may raise the issue as to whether the present sample is composed of individuals with higher levels of trait anxiety than those expected in the general population.

In fact, one limitation of the study was the failure to include an anxiety measure which could help to disentangle the effect of theory of mind ability from that of trait anxiety on emotional biases to emotional expressions. It is possible that individuals with low theory of mind skills are also the ones exhibiting higher levels of anxiety. This association would account for the common pattern in terms of attentional bias shared by individuals with high levels of trait anxiety and individuals with poor theory of mind skills. In fact, anxiety-related difficulties have been found to be a prevalent feature of disorders characterized by poor theory of mind skills, such as the autism spectrum disorders (e.g., Chalfant, Rapee, & Carroll, 2007). Hence, the difficulty in interpreting other people's thoughts and feelings would lead to a sense of unpredictability in social relationships and contribute to increase anxiety levels; on the other hand, high anxiety levels would interfere with social interactions and make it more difficult for the individual to make sense of other people's behaviours. Nevertheless, only the assessment of participants' anxiety levels would permit to clarify this issue and offer a valid interpretation of the observed relation between poor theory of mind skills and attentional bias towards negative and away from non-negative faces. It is hoped that future research can clarify the issues raised by this preliminary study, contributing to better understand

the relation between theory of mind skills and attentional bias to facial emotions over and above the effect of anxiety.

References

- Baron-Cohen, S., Jolliffe, T., Mortimore, C., & Robertson, M. (1997). 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.
- Baron-Cohen, S., Wheelwright, S., Hill, J., Raste, Y., & Plumb, I. (2001). 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.
- Beck, A.T. (1987). Cognitive models of depression. *Journal of Cognitive Psychotherapy, An International Quarterly*, 1, 5-37.
- Bosack, S. & Astington, J.W. (1999). Theory of mind in preadolescence: Relations between social understanding and social competence. *Social Development*, 8, 237-255.
- Bradley, B.P., Mogg, K., Falla, S.J., & Hamilton, L.R. (1998). Attentional bias for threatening facial expressions in anxiety: Manipulation of stimulus duration. *Cognition and Emotion*, 12, 737-753.
- Broadbent, D., & Broadbent, M. (1988). Anxiety and attentional bias: State and trait. *Cognition and Emotion*, 11, 25-42.
- Chalfant, A.M., Rapee, R., & Carroll, L. (2007). Treating anxiety disorders in children with high functioning autism spectrum disorders: A controlled trial. *Journal of Autism and Developmental Disorders*, 37, 1842-1857.
- Coupland, N.J. (2001). Social phobia: Etiology, neurobiology, and treatment. *Journal of Clinical Psychiatry*, 62(Suppl.1), 25-35.
- Fonagy, P. & Target, M. (1997). Attachment and reflective function: Their role in self-organization. *Development and Psychopathology*, 9, 679-700.

- MacLeod, C., Mathews, A., Tata, P. (1986). Attentional bias in emotional disorders. *Journal of Abnormal Psychology, 95*, 15-20.
- MacLeod, C., & Mathews, A. (1988). Anxiety and the allocation of attention to threat. *The Quarterly Journal of Experimental Psychology A: Human Experimental Psychology, 40*, 653-670.
- Mansell, W., Clark, D.M., Ehlers, A., & Chen, Y.P. (1999). Social anxiety and attention away from emotional faces. *Cognition and Emotion, 13*, 673-690.
- Mathews, A., MacLeod, C. (1985). Selective processing of threat cues in anxiety states. *Behaviour Research and Therapy, 23*, 563-569.
- Mathews, A., & MacLeod, C. (1986). Discrimination of threat cues without awareness in anxiety states. *Journal of Abnormal Psychology, 95*, 131-138.
- Matsumoto, D., & Ekman, P. (1988). *The Japanese and Caucasian Facial Expressions of Emotion (JACFEE) and Neutrals (JACNeuF)*. [Slides]: Intercultural and Emotion Research Laboratory, Department of Psychology, San Francisco State University, 1600 Holloway Avenue, San Francisco, CA 94132.
- Morton, J., Frith, U., & Leslie, A. (1991). The cognitive basis of a biological disorder: Autism. *Trends in Neurosciences, 14*, 434-438.
- Posner, M.I., Snyder, C.R., & Davidson, B.J. (1980). Attention and the detection of signals. *Journal of Experimental Psychology: General, 109*, 160-174.
- Power, M., & Dalgleish, T. (2008). *Cognition and emotion: From order to disorder*. New York: Psychology Press.
- Premack, D., & Woodruff, G. (1978). Does the chimpanzee have a "theory of mind"? *Behaviour and Brain Sciences, 4*, 515-526.

Fig.1: Probe detection latencies for negative faces as a function of probe location and Eyes task performance (N = 46)

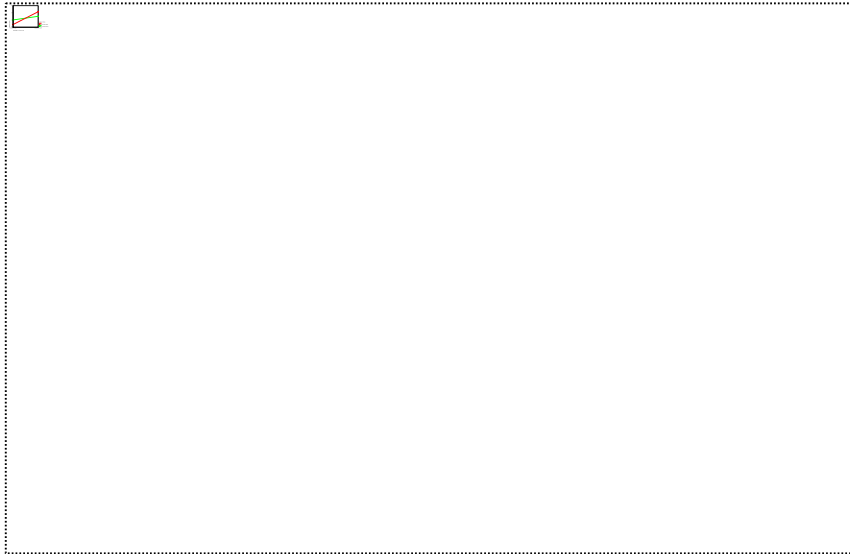


Fig.2: Mean probe detection latencies for negative faces (N = 46)



Fig.3: Bias scores as a function of face valence and Eyes task performance (N = 46)



Fig 4: Mean bias scores as a function of face valence and Eyes task performance (N = 46)

