In press, *Cognition*

Seeing the World through Others' Minds:

Inferring Social Context from Behaviour

Yvonne Teoh [a], Emma Wallis[c], Ian D. Stephen [b], Peter Mitchell [c*]

[a] School of Psychology, University of Nottingham Malaysia Campus, Semenyih, 43500, Selangor, Malaysia.

[b] Department of Psychology, Macquarie University, Sydney, Australia.

[c] School of Psychology, University of Nottingham, Nottingham, United Kingdom.

[*]Corresponding author. Email address: peter.mitchell@nottingham.ac.uk

Abstract

Past research tells us that individuals can infer information about a target's emotional state and intentions from their facial expressions (Frith & Frith, 2012), a process known as *mentalising*. This extends to inferring the events that caused the facial reaction (e.g. Pillai, Sheppard, & Mitchell, 2012; Pillai et al., 2014), an ability known as *retrodictive mindreading*. Here, we enter new territory by investigating whether or not people (perceivers) can guess a target's social context by observing their response to stimuli. In Experiment 1, perceivers viewed targets' responses and were able to determine whether these targets were alone or observed by another person. In Experiment 2, another group of perceivers, without any knowledge of the social context or what the targets were watching, judged whether targets were hiding or exaggerating their facial expressions; and their judgments discriminated between conditions in which targets were observed and alone. Experiment 3 established that another group of perceivers' judgments of social context were associated with estimations of target expressivity to some degree. In Experiments 1 and 2, the eye movements of perceivers also varied between conditions in which targets were observed and alone. Perceivers were thus able to infer a target's social context from their visible response. The results demonstrate an ability to use other minds as a window onto a social context that could not be seen directly.

Keywords: mentalising, social context, eye movements, inferences, retrodiction

1. Introduction

Over the past century researchers have struggled to understand people's ability to read others' minds. This ability has variously been called *mentalising, mindreading, mind perception, empathic accuracy, mental simulation* and *theory of mind* amongst other things. Inspired by the work of Charles Darwin (1872), researchers have investigated the ability to interpret facial expressions to infer underlying psychological states and traits (e.g. Baron-Cohen, Jolliffe, Mortimore & Robertson, 1997; Wu, Sheppard & Mitchell, 2016). Such ability has great value in professional counselling as recognized by Carl Rogers (1957), who set the goal of finding people with a talent for 'accurate empathy', in other words the ability to infer what a client is thinking and feeling. This stimulated a research tradition extending far beyond its origins in counselling psychology to determine how accurately people can read other minds, but, according to Zaki and Ochsner (2011), the early work in the field of person perception devoted little attention to the process of *how* people read minds. Subsequently, researchers working in a different tradition investigated the development of a 'theory of mind' (Wimmer & Perner, 1983), and how that development might be adversely affected by autism (Baron-Cohen, Leslie & Frith, 1985). These researchers expended much effort in trying to understand the *processes* of mentalising but did not, until recently, give much attention to how accurately mature mentalisers perform (Zaki & Ochsner, 2011).

In determining how accurately a person (the perceiver) can mentalise, it is useful to know the true mental state of the person whose mind is being read (henceforth the target). West and Kenny (2011) recognize that knowing the true state of the target's mind presents a difficult problem and they refer to the procedure devised by Ickes (e.g. 2001, 2009). In the procedure, the target is videoed in conversation with another person. The video is then played back to the target

who recalls and records what they were thinking and feeling during the conversation. Subsequently, perceivers watch the video of the target and are asked to infer what the target is thinking and feeling; they are adjudged to be correct if their responses correspond with what the target declared at any given moment.

This procedure presumes that when the target declares that they are thinking and feeling X and Y then they are really thinking and feeling these things. Another possibility is that targets do not know or at least do not recall what they were thinking and instead merely guess at these things based on visible clues in the recordings of their own observable behaviour. If so, then investigating how well the perceiver's judgment corresponds with the target's declaration is merely the same as investigating judgments made by two independent perceivers about the behaviour of a target. A procedure that overcomes this problem was developed by North, Todorov and Osherson (2010). In their task, targets were surreptitiously videoed while viewing two photographs presented one after the other. Perceivers subsequently watched the videos of the targets and were able to infer which photo the target preferred (the first or the second), presumably by recognising that the target had a more positive expression when viewing one of the photographs than the other. In this procedure, we know objectively which preference the target expressed (thus satisfying the 'truth condition' as defined by West and Kenny).

In addition to inferring what others are thinking and feeling, how well can people use others' minds as a lens onto an otherwise inaccessible view of the world? Such ability was fictionalised in Sherlock Holmes (Conan Doyle, 1902), who was able to observe and interpret fleeting clues in behaviour to infer what the person had been doing, where they had been and with whom. While ordinary people might not perform at the extraordinary level of Sherlock

Holmes, based on findings described above, perhaps they can nevertheless achieve something similar by a matter of degree.

The mind is embodied in observable behaviour, especially in the facial expressions that are made in reaction to some event. Kraut (1982) claimed, moreover, that facial expressions potentially provide information about the environmental and social contexts that caused the reaction in the target. It is for this reason, presumably, that perceivers tested by North et al. (2010) were able to infer the preferences of targets who viewed pairs of photos. Facial expressions might also reveal other information about targets' states and the aspects of the world they inhabit that caused those states. For example, Cassidy, Ropar, Mitchell, and Chapman (2013; 2015) reported that perceivers were correctly able to infer which gift had been offered to a target (chocolate, homemade novelty and monopoly money) by observing their reactions.

In another study, Pillai et al., (2012; 2014; Sheppard, Pillai, Wong, Ropar & Mitchell, 2016) examined perceivers' ability to guess what the experimenter had said to the target after viewing the target's reaction for a few seconds. Either the experimenter told a joke, gave a compliment, related her difficult day or rudely used her mobile phone to speak with a friend instead of attending to the target. As with the study by Cassidy et al. (2013), perceivers were ble to guess what the experimenter had said to the target after observing the target's reaction for a few seconds. Perceivers were thus able to infer the antecedent event based on a small sample of the target's behaviour. This ability, known as retrodictive mindreading (Gallese & Goldman, 1998), is reputed to be a common form of mentalising that allows people to determine from a facial expression (a) the proximal cause, which is the target's mental state and (b) the distal cause, which is the event in the world that gave rise to the mental state that in turn caused the facial expression. Hence, among the various practical benefits of mentalising, one of the

foremost is using another mind as a lens onto aspects of the world that are not apprehended directly. Such ability emerges very early in development in a basic form known as 'social referencing' (Sorce, Emde, Campos, & Klinnert, 1985): From about the age of 18 months it seems infants can interpret an adult's facial expression to determine their attitude on whether an aspect of the world is safe or dangerous which in turn has the effect of regulating the infant's approach and avoidance behaviour towards that particular aspect of the world.

The current study investigated a new phenomenon - people's ability to infer social context (was the target alone or accompanied by the experimenter?), along with perceivers' sensitivity to how this social context moderated target's reactions to positive or negative stimuli. The presence of the experimenter would surely have a subtle effect on the target's behaviour. A large body of research has demonstrated that individuals behave differently in different social contexts (e.g., Ekman, 1972; Fridlund, 1991; Zaalberg, Manstead, & Fischer, 2004) by inhibiting or intensifying their behaviour when in the presence of others depending on the emotions experienced (e.g., Kilbride & Yarczower, 1980; Kraut, 1982). Specifically, when individuals experience negative emotions (e.g., irritation, disgust), they tend to inhibit their behaviour (Spain, Eaton, & Funder, 2000), but when experiencing positive emotions (e.g., happiness, surprise) they tend to be more expressive (Buck, 1984; Ekman & Friesen, 1975). We already know that social context influences people's behaviour, and re-confirming such a finding was not the purpose of the current research; rather we take it as given that social context will have an effect on the target's behaviour and we move beyond this basic assumption to explore *whether or not perceivers can determine social context from the target's behaviour*. If they could do so, then it would raise the possibility that perceivers have some understanding (implicitly or explicitly) of how social presence regulates behaviour, an understanding they might use to good effect in

guessing whether the target is alone or accompanied. In so far as retrodictive mindreading is possible, it thus implicates a well-developed albeit informal understanding of social processes, such as how social presence impacts upon the way one behaves. Hence, a further aim was to shed light on the process by which perceivers made inferences of social context as elaborated below.

If perceivers understand (either implicitly or explicitly) that social presence suppresses the expression of negative emotions but intensifies the expression of positive emotions then this should be reflected not only in their judgments of whether the target is accompanied or alone; it should also be apparent in their explicit judgments of how expressive the target is. The latter was tested explicitly in Experiments 2 and 3, reported below, allowing us to investigate the possibility that perceivers infer social presence on the basis of sensing that targets were regulating their expressions of negative and positive emotions relative to whether they were alone or accompanied. In this respect, the current research forms a bridge between the tradition of investigating accuracy in person perception (which in the past has neglected the question of *process*) with the tradition of investigating process under the umbrella of research into 'theory of mind' (which in the past has neglected to consider the findings of 'accuracy research' in the area of person perception).

In addition to asking perceivers to make judgments about the targets (in Experiments 1-3), we also recorded the eye movements of the perceivers in Experiments 1 and 2, principally to rule out any low-level strategies in judging the social context of targets. In addition to ruling out the use of low level strategies, we might also find that sensitivity to social context is apparent in more subtle features of the eye movements of perceivers. Pillai et al.'s (2012) study showed that perceivers tended to focus on the mouth more than the eyes for the joke, story and compliment

scenarios, but for the waiting scenario, perceivers preferred to look at the eyes than the mouth. Meanwhile, Pillai et al. (2014), found similar patterns for the joke, story and compliment scenarios, but perceivers focused equally on the mouth and eyes in the waiting scenario. Cassidy et al.'s (2013) study suggested that perceivers' eye movements differed depending on the gift the targets received, where perceivers tended to look at the mouth more than the eyes when targets received homemade novelty and chocolate gifts, but focused more on the eyes than the mouth when targets received monopoly money. These studies collectively demonstrated that perceivers' sensitivity to what had happened to the target was apparent in their eye movements.

In Experiments 1 and 2, we showed film clips to targets with the aim of evoking a range of emotions while targets experienced two different social contexts (observed, alone). In Experiment 3, targets viewed provocative photographs. Targets' reactions to the stimuli during the different social situations were video recorded and used as stimuli. Since targets were filmed surreptitiously via the laptop's webcam, their natural and spontaneous reactions were captured unobtrusively and without awareness. This approach might have some benefits over methods in which targets were deliberately enacting emotional states (Klin, Jones, Schultz, Volkmar, & Cohen, 2002; Pillai et al., 2012).

Experiment 1 investigated whether or not perceivers could correctly identify the social context based on targets' reactions. In Experiment 2, perceivers rated the degree to which targets were hiding or exaggerating their behaviour and we investigated whether ratings varied according to the social contexts and emotions the targets experienced. Perceivers' eye movements were also recorded in both experiments. Experiment 3 investigated the association between perceivers' judgments of target social context and their rating of how expressive the targets were.

2. Stimuli Development for Experiments 1 and 2

2.1. Video stimuli presented to targets

The method was adapted from Pillai et al.'s (2012) and Cassidy et al.'s (2013) studies. Targets watched various movie clips and then rated the emotions they experienced. The content of the video clips was intended to make targets feel irritation, disgust or surprise. The main purpose of this study was not to investigate whether perceivers could identify emotions experienced by the targets – a topic that has been over-researched in the past; rather, we investigated whether or not perceivers who viewed the reactions of the targets could sense whether the targets were alone or being observed.

In the stimuli development phase, 24 university students aged between 18 and 30 (M = 20.1, SD = 2.4) watched 15 videos selected to elicit irritation, disgust or surprise (five for each emotion). Participants rated the emotion aroused by each clip on a 7-point scale where 1 = *least intense* and 7 = *most intense,* and briefly described a scene from the clip that predominantly evoked the emotion they experienced. Based on the ratings, two video clips were selected for each emotion: two clips from *YouTube* (metallicKuma, 2009; Shurken12345, 2008) to evoke irritation, clips from the movie *127 Hours* (Colson & Boyle, 2010) and *Excision* (Lewis & Bates Jr., 2012) to evoke disgust and scenes from the TV show *Britain's Got Talent* (Hurford-Jones, 2009), and an advert from Germany (mrsmithereen, 2005) to evoke surprise (links to the clips are provided in the reference section). The video clips were all freely available and varied in length between two to ten minutes.

2.2. Targets

Twenty-two male and twenty-one female students aged between 18 and 24 (M = 19.5, SD = 1.7) from the University of Nottingham Malaysia Campus were recruited. Targets were of different nationalities: 31 Malaysians, 2 Chinese, 2 Pakistani, 2 Indians, 1 Singaporean, 1 Egyptian, 1 Syrian, 1 English, 1 Sri Lankan, and 1 South Korean. Written informed consent was obtained from all participants, and the experiment was approved by the University of Nottingham Malaysia Campus Ethics Committee.

2.3. Materials and apparatus

The selected video clips were presented in full screen in random order using Psychopy (Version 1.74.01, Peirce, 2012) on a HP EliteBook 8460p laptop. The laptop's HD Webcam surreptitiously filmed targets' reactions as they viewed the videos. The targets were positioned approximately 0.6 meters from the webcam to ensure that their face, neck and shoulders were captured.

2.4. Procedure

A spacious room within the School of Psychology was used. Targets sat with their back to a white wall, while the experimenter sat across the table. Upon arrival, targets were informed that they were going to watch a few videos and rate what they felt from each clip. Each target was tested individually under two conditions presented in counterbalanced order – observed and alone. For the 'observed condition', the experimenter sat across the table from the target ostensibly to observe their behaviour as they watched the videos. The experimenter was dressed in a lab coat, scribbled notes on a clipboard and did not communicate with the target during the observation. In the 'alone condition', the experimenter made an excuse to leave the room ("I

have an important phone call to make, I'll be right outside, just come and look for me once you're done.") and targets were left alone in the room to watch the video clips.

The 14 video clips (six focal videos plus 8 fillers that were adjudged in piloting not to provoke irritation, disgust or surprise) were divided into two blocks for the two conditions (observed and alone), and targets watched seven clips (three focal videos and 4 fillers) in each block designed to evoke emotions. These clips were presented in random order via PsychoPy, and each clip was shown only once. The presentation of the two blocks of clips was counterbalanced across targets to ensure there was no confound in the content of the videos and the emotions across the conditions. Since the aim of the study was to record targets' spontaneous behaviour, the webcam was set to record as soon as targets entered the room and the light to signal recording was disabled; hence they were unaware that they were being video recorded.

As a 'decoy activity', after viewing each clip targets had to select a response out of seven choices (anger, disgust, fear, happy, neutral, sad, surprise) that best represented their emotion from viewing the clip, and the intensity of the emotion on a 7-point scale, 1 = *least intense* and 7 = *most intense*. These responses were presented on the screen in a rating scale, and targets selected their responses by using the mouse.

Once targets had completed both conditions, they were debriefed about the true nature of the study, and given the opportunity to ask any questions. Targets were asked for their consent to use their video recorded reactions for a subsequent part of the study.

2.5. Editing

The video recordings of targets' behaviour were edited to capture their reactions using video-editing software, Windows Live Movie Maker and VirtualDub (Version 1.9.11, Lee, 2010). Target reactions were selected from the most emotionally intense part of the videos they viewed, as described by participants in the stimulus development phase (see 2.1) to avoid any experimenter bias in selecting the most reactive target behaviour. The duration of the videos varied between 2 to 14 seconds. The edited clips were muted in case the audio revealed the context the targets were experiencing. The video frame was cropped to 500 pixels for both width and height, showing only the target's head to the top of their shoulders against a white wall. The rate of presentation was 29 frames per second. Only videos of Asian targets were used, out of which ten (five males, five females) were randomly selected. Thus, a total of 60 videos were used whereby for every target, a total of 6 videos were generated that consisted of two videos for each of the three emotions.

3. Experiment 1

3.1. Participants

Thirty-three perceivers (17 males, 16 females) were recruited from the University of Nottingham Malaysia Campus aged between 18 and 23 (M = 19.8, SD = 1.3). Perceivers comprised 28 Malaysians, 1 Hong Konger, 1 Indonesian, 2 Maldivians, and 1 Tanzanian. All gave written informed consent and were compensated RM5 (approx. US$2) for taking part. Three perceivers (2 males, 1 female) had less than 50 per cent of their gaze sample collected by the eye-tracker and their data were excluded from the analysis.

3.2. Materials and Apparatus

In order to demonstrate to the perceivers the two possible social contexts in which the targets were recorded, video clips of the experimenter's behaviour during the stimuli development phase were recorded using a Nikon Coolpix S5100. The experimenter sat across the table from a confederate who acted as the target, and the video camera was placed next to the confederate, aimed at the experimenter, to record how the experimenter appeared from the perspective of the target. These clips were then edited using VirtualDub (Version 1.9.11, Lee, 2010) and Windows Live Movie Maker for a duration of 5 seconds with audio muted. The clips were labelled at the top centre to indicate whether it was the alone or the observed condition (see Fig. 1). The video clips were 856 pixels in width and 480 pixels in height, with a rate of presentation at 29 frames per second.

Fig. 1. Stimuli shown to perceivers in Experiment 1 to ilustrate the two social contexts experienced by the targets (Left: alone condition, Right: observed condition).

The target videos created from the stimuli development stage were used in this study. Sixty stimuli consisting of five male and five female targets watching videos that might elicit irritation, disgust or surprise under two different social conditions (i.e. alone and observed) were used. Stimuli were presented randomly through Tobii Studio Software on a 17-inch TFT monitor

integrated with the Tobii T60 eye tracking system. Eye movements were recorded at a data sampling rate of 60Hz.

3.3. Procedure

Perceivers were tested individually in a quiet room. They sat approximately 60 cm from the monitor of a Tobii T60 eye tracker and a nine-point calibration was conducted to ensure accurate recording of the eye movements. After successfully calibrating, perceivers were shown two videos of the experimenter's behaviour in the alone and observed condition (counterbalanced across perceivers). The purpose was to give perceivers an idea of the social conditions the targets experienced. Perceivers were then instructed to rate whether the target was alone or observed in each of the 60 video clips, based on the target's behaviour. A fixation point (+) was presented for one second pseudorandomly in one of the four quadrants of the screen to prevent fixation bias before each stimulus. The video stimuli were then presented to the perceivers in random order. Each clip was viewed only once by each perceiver. After every clip, perceivers were presented with a two-alternative forced choice question to decide whether the target was alone or observed, and were required to respond using the mouse. The responses were recorded automatically by Tobii Studio. Perceivers were allowed to answer each question at their own pace.

3.4. Results and Discussion

3.4.1. Behaviour Data

To examine whether perceivers were able to identify when targets were observed by another or alone, the responses were coded based on the number of times perceivers responded 'Observed' for the different emotions and conditions.

A 3 (Emotions: irritation, disgust, surprise) × 2 (Conditions: observed, alone) repeated-measures ANOVA was conducted. Effect sizes are reported as η_p^2 for ANOVA results and are interpreted as follows; ≥ 0.01 is considered a small effect, ≥ 0.06 a medium effect, and ≥ 0.14 is a large effect (Cohen, 1992). There were large significant main effects of Emotions, $F(1.49, 43.27) = 19.07$, $p < .001$, $\eta_p^2 = .40$ and Conditions, $F(1,29) = 8.47$, $p = .007$, $\eta_p^2 = .23$ (see Fig. 2). There was no significant effect associated with the interaction term, $F(2,58) = 0.29$, $p = 0.75$, $\eta_p^2 = .10$. The main effect of Conditions resulted from perceivers judging that the target was observed significantly more often when that was true than when it was not true; the absence of an interaction effect suggests there is no reason to suppose that the ease of such discrimination varied between the three emotions. Pairwise comparisons of the main effect associated with the factor Emotions revealed that perceivers tended to judge that the target was observed significantly more often when he/she was watching the 'irritation' video than when watching the other videos (*p*s < .001).

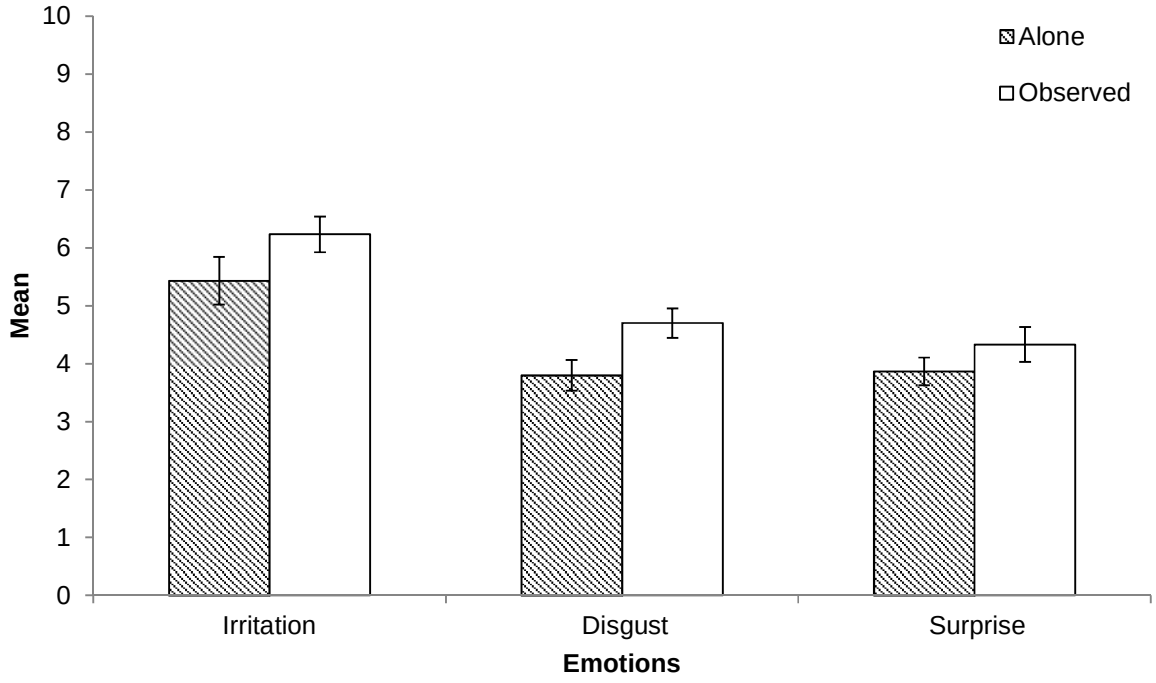

Fig. 2. Number of times perceivers responded 'observed' out of 10 for conditions in which targets were alone and observed in Experiment 1. Standard errors of the mean are represented by the error bars.

Did perceivers use a low-level strategy to determine whether targets were observed or alone? For example, did targets look away from the laptop screen (to shift their gaze towards the experimenter) more often in the observed than in the alone condition? If so, perhaps perceivers based their guess of social context on whether or not the target frequently looked away from the screen (and thus the camera). To find out, we coded the number of times targets looked away from the screen but found no evidence to suggest that this was more common in the observed than in the alone condition, $t(58) = 0.63$, $p = 0.53$. Apparently, then, this potentially low-level cue did not offer any basis for perceivers to guess whether the target was observed or alone.

3.4.2. Eye Movement Analyses

Areas of interest (AOIs) were created using Tobii Studio 3.1.6. The AOIs were drawn onto each stimulus independently to delineate the eyes, nose and mouth, following Tan, Stephen, Whitehead, and Sheppard (2012), with the addition of an AOI for the Body (see Fig. 3). The Total Fixation Duration (seconds) for each AOI was obtained from Tobii Studio. Since each stimulus varied in duration, the percentage duration of fixations on each AOI were calculated [(fixation duration of an AOI/total duration of the video) × 100].

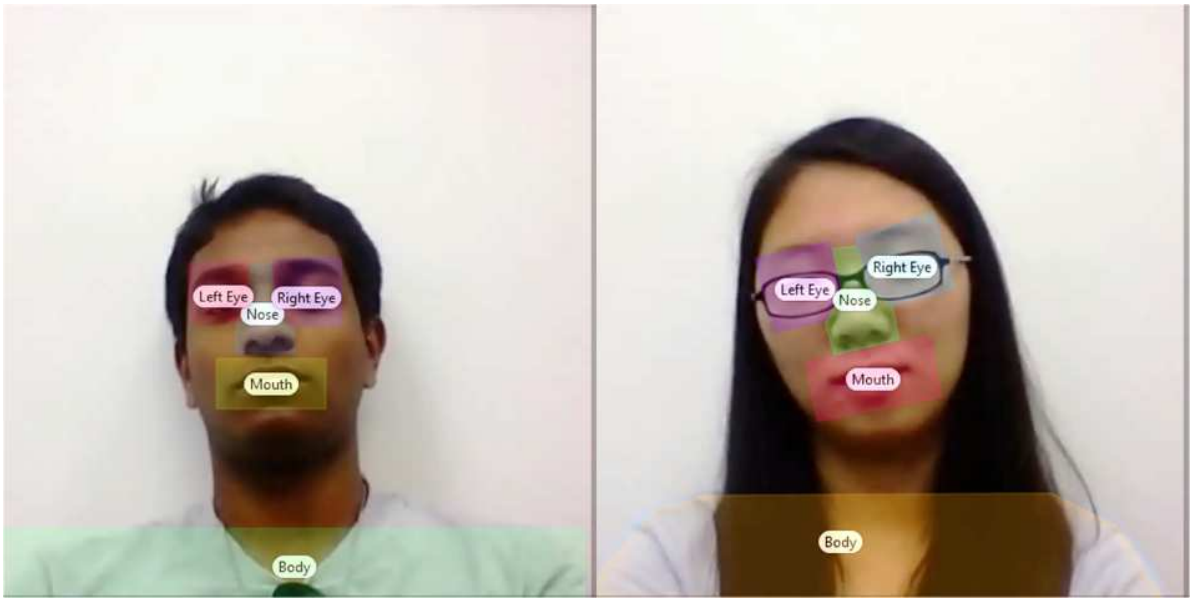

Fig. 3. The predefined areas of interest (AOIs) used to analyse eye movement behaviour.

A preliminary one-way repeated-measures ANOVA was conducted on the mean percentage gaze time of AOIs to examine which area perceivers looked at longer and a large significant main effect was found, $F(2.25, 65.36) = 31.17$, $p < .001$, $\eta_p^2 = .52$. Pairwise comparisons indicated that perceivers looked at the Eyes region (M =25.4, SD = 13.5) longer than the Mouth (M = 14.8, SD = 8.8) (p = .01), and spent more time looking at the Nose (M = 25.1, SD = 12.9) than the Mouth (p = .01). The Body (M =3.6, SD = 1.5) was the least attended region (ps < .001). The main point of interest lies in whether or not the patterns of eye movements at the different regions vary depending on the social context. Because the variance differed considerably between the different areas of interest it was appropriate to conduct separate 3 (Emotions: irritation, disgust and surprise) × 2 (Conditions: observed, alone) repeated-measures ANOVAs for the four different AOIs (Body, Mouth, Nose, Eyes: See Fig. 4).

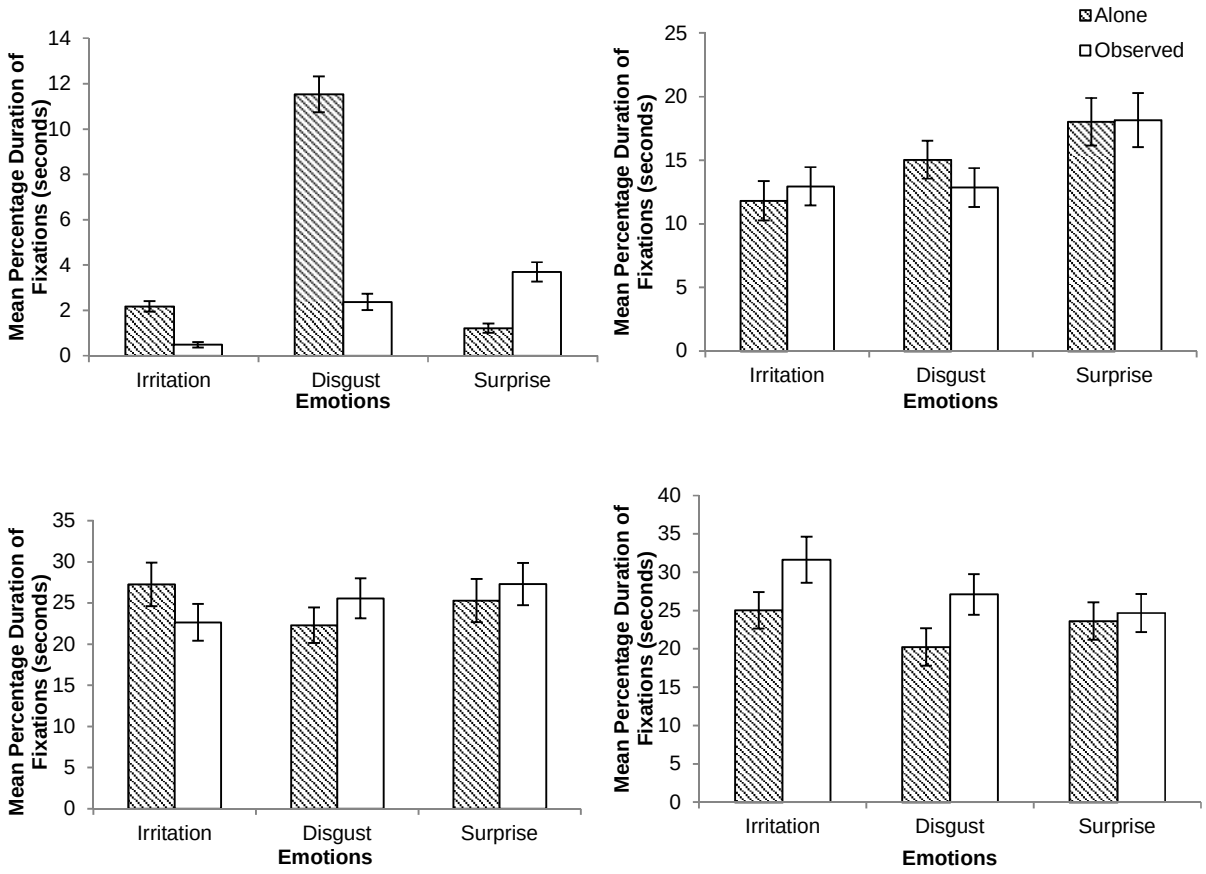

Fig. 4. Mean percentage duration of perceiver fixations (seconds) for conditions in which targets were alone and observed for the different AOIs in Experiment 1. The standard errors of the mean are represented by the error bars. (Top left: Body region, Top right: Mouth region, Bottom left: Nose region, Bottom right: Eyes region).

For the Body region, there was a large significant main effect of Emotions, $F(1.42, 41.08) = 151.09$, $p < .001$, $\eta_p^2 = .84$, a large significant main effect of Conditions, $F(1, 29) = 116.74$, $p < .001$, $\eta_p^2 = .80$, and a large significant interaction between the Emotions and Conditions, $F(1.25, 36.09) = 131.68$, $p < .001$, $\eta_p^2 = .82$. Posthoc paired-samples *t*-tests indicated

that perceivers looked longer at the Body for disgust ($t(29) = 12.60$, $p < .001$, and irritation ($t(29) = 7.31$, $p < .001$) alone than observed, and surprise observed ($t(29) = -6.50$, $p < .001$) than alone.

For the Mouth region, there was a large significant effect of Emotions, $F(2, 58) = 44.01$, $p < .001$, $\eta_p^2 = .60$, and a medium significant interaction between the Emotions and Conditions, $F(1.72, 49.90) = 4.54$, $p = .02$, $\eta_p^2 = .14$. Posthoc paired-samples *t*-tests showed that perceivers looked longer at the Mouth for disgust alone than observed, $t(29) = 3.61$, $p = .001$.

Similarly, for the Nose region, there was a medium significant main effect of Emotions, $F(2,58) = 4.00$, $p = .02$, $\eta_p^2 = .12$, and a large significant interaction between Emotions and Conditions, $F(2,58) = 16.24$, $p < .001$, $\eta_p^2 = .36$. Posthoc *t*-tests showed that perceivers spent more time looking at the Nose region for irritation alone than observed, $t(29) = 4.34$, $p < .001$, disgust observed than alone, $t(29) = -3.50$, $p = .002$, and surprise observed than alone, $t(29) = -2.09$, $p = .045$.

Lastly, the Eyes region revealed a large significant main effect of Emotions, $F(2,58) = 23.27$, $p < .001$, $\eta_p^2 = .45$, Conditions, $F(1,29) = 57.52$, $p < .001$, $\eta_p^2 = .67$, and a large significant interaction between the factors, $F(2,58) = 6.03$, $p = .004$, $\eta_p^2 = .17$. Posthoc paired-samples *t*-tests revealed that perceivers attended more to the region for irritation observed than alone ($t(29) = 5.08$, $p < .001$), and disgust observed than alone ($t(29) = 5.72$, $p < .001$).

Another way to calculate the percentage of fixation duration in order to control for the varied stimuli duration is to divide the fixation duration of an AOI with the total fixation duration for each stimulus, instead of the total duration of each stimulus × 100. The findings have similar patterns to that reported above with the exception of the Nose region, where posthoc *t*-tests revealed significant differences between alone and observed for irritation ($t(29) = 5.40$, $p < .001$) and disgust ($t(29) = -2.55$, $p = .02$) only.

In summary, the eye movement data showed that the time perceivers spent looking at the different regions varied depending on whether targets were alone or observed. It was not that perceivers simply looked more at any particular region when the target was alone than observed, suggesting that perceivers were not influenced by any low-level clues in the behaviour of targets that might reveal their social context; rather, the relationship between time perceivers spent looking at a region and whether the target was alone or observed depended on the particular kind of emotion the target was experiencing.

4. Experiment 2

One purpose in Experiment 2 was to begin to illuminate the process by which perceivers had inferred the social context of the target in Experiment 1. Previous research has established that when people experience positive emotions (such as surprise) they tend to be more expressive in the company of others than when alone (Buck, 1984); conversely, when people experience negative emotions (such as irritation and disgust) they tend to hide their expressions more in the company of others than when alone (Spain et al., 2000). In Experiment 2 we asked perceivers to judge the degree to which targets were controlling their expressions (hiding or exaggerating), with the expectation that they might judge that targets are hiding their expressions when they were observed and experiencing a negative emotion (irritation, disgust) but exaggerating their expression when observed and experiencing a positive emotion (surprise). They should judge in this way if their explicit judgments of whether the target is observed or alone, as measured in Experiment 1, were based on an understanding (either implicitly or explicitly) of how social presence differentially influences the expression of positive and negative emotions. At the very least, Experiment 2 has potential to demonstrate that sensitivity to the social context of the target is also apparent in perceivers' judgments of how targets are controlling their emotions.

4.1. Participants

Twenty males and nineteen females aged between 18 and 24 (M = 19.4, SD = 1.1) were recruited from the University of Nottingham Malaysia Campus. All were Asian: 32 Malaysians, 3 Singaporeans, 2 Vietnamese, 1 Indian and 1 Bangladeshi. Written informed consent was obtained from all perceivers who were compensated RM5 (approx. US$2) for participating in the study. Less than 50% of the gaze sample was collected by the eye-tracker for six perceivers (4 males, 2 females), and data from these perceivers were duly excluded from the analysis.

4.2. Materials and Apparatus

The sixty target video stimuli used in Experiment 1 were utilised. The target stimuli were presented in random order through Tobii Studio Software. A 17-inch TFT monitor that is equipped with the Tobii T60 eye tracking system was used to record eye movements at a data sampling rate of 60Hz.

4.3. Procedure

The procedure was similar to that used in Experiment 1, except perceivers were not presented with the videos of the experimenter demonstrating the observed and alone conditions, and were not asked to judge whether the target was alone or observed. Hence, unlike in the previous experiment, perceivers were not given any information about the two possible social contexts. A nine-point calibration was performed and then, for each stimulus clip, perceivers were instructed to decide the degree to which the target was controlling their expression on a rating scale of 1-7 (1 = *hiding*… 7 = *exaggerated*). The rest of the procedure was identical to that in Experiment 1.

4.4. Results and Discussion

4.4.1. Behaviour Data

Perceivers' ratings of how the targets were controlling their expressions were submitted to a 3 (Emotions: irritation, disgust, surprise) × 2 (Conditions: observed, alone) repeated-measures ANOVA. There was a large significant main effect of Emotions, $F(2,64) = 29.36$, $p < .001$, $\eta_p^2 = .48$, which resulted from perceivers judging that targets were hiding their expressions more in the irritation condition compared with the other two emotions (ps < .001). There was no main effect associated with the factor, Condition, $F(1,32) = 1.08$, $p = 0.31$, $\eta_p^2 = .03$, but there was a medium-sized significant interaction between Emotions and Conditions, $F(2,64) = 6.63$, $p = .002$, $\eta_p^2 = .17$ (see Fig. 5). Posthoc paired-samples *t*-tests showed a significant difference between disgust alone and observed, $t(32) = 2.81$, $p = .008$, which resulted from perceivers judging that targets were hiding their expressions more in the observed than in the alone condition. There was also a significant difference between surprise alone and observed, $t(32) = 2.54$, $p = .02$, which resulted, conversely, from perceivers judging that targets were hiding their expressions less in the observed than in the alone condition. Hence, perceivers were effectively able to discriminate when targets were alone and observed by judging the extent to which targets were controlling their expressions. However, it was not the case that perceivers consistently judged that targets were hiding their expressions more when observed than when alone; rather, the direction of the effect also depended on the particular emotion the target was experiencing.

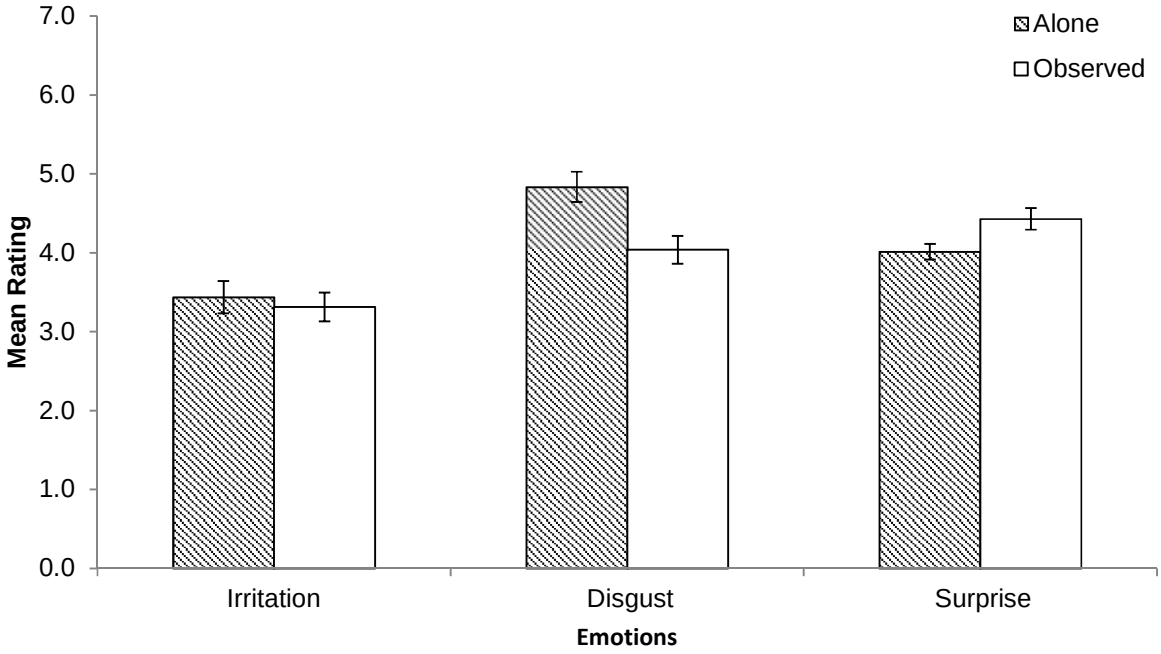

Fig. 5. Mean ratings made by perceivers of how targets were controlling their expressions (1= *hiding*… 7= e*xaggerated*) for the alone and observed condition in Experiment 2. The standard errors of the mean are represented by the error bars.

The findings were consistent with expectations. Generally, perceivers tended to judge that targets were exaggerating when the target was experiencing a positive emotion (surprise) and was observed rather than alone. Conversely, they tended to judge that targets were hiding their expressions when the target was experiencing a negative emotion (irritation or disgust) and was observed rather than alone. Such behaviour in the targets is consistent with what we might have expected from the literature (Buck, 1984; Spain et al., 2000) and the novel finding presented here suggests that perceivers were attuned to this phenomenon, for their judgments of how the targets were controlling their expressions effectively discriminated between the two social conditions experienced by the targets.

4.4.2. Eye Movement Analyses

Areas of interest (AOIs) for the stimuli were created using Tobii Studio 3.1.6 to calculate perceivers' total fixation duration (seconds) on each target AOI (Eyes, Nose, Mouth, Body). In order to control for any differences in the stimulus duration, the percentage of fixation duration was calculated [(fixation duration of an AOI/total duration of the video) × 100].

To investigate which AOI perceivers focused on more, a one-way repeated-measures ANOVA was conducted on the mean percentage gaze time of each AOI, revealing a large significant main effect between the regions, $F(1.89, 60.44) = 17.21$, $p < .001$, $\eta_p^2 = .35$. Pairwise comparisons revealed that the Body ($M = 4.0$, $SD = 1.8$) was attended to less (ps < .001) than the Nose (M =25.7, $SD = 12.2$), Eyes ($M = 21.5$, $SD = 17.0$), and Mouth ($M = 21.4$, $SD = 14.0$).

Separate 3 (Emotions: irritation, disgust, surprise) × 2 (Conditions: observed, alone) repeated-measures ANOVAs were conducted for the four different AOIs (Body, Mouth, Nose, Eyes (see Fig. 6). For the Body region, the ANOVA revealed a large significant main effect of Emotions, $F(2,64) = 89.66$, $p < .001$, $\eta_p^2 = .74$, a large significant main effect of Conditions, $F(1,32) = 42.07$, $p < .001$, $\eta_p^2 = .57$, and a large significant interaction between Emotions and Conditions, $F(1.45,46.38) = 151.52$, $p < .001$, $\eta_p^2 = .83$. Posthoc paired-samples t-tests showed that perceivers fixated longer for irritation ($t(32) = 8.47$, $p < .001$) and disgust ($t(32) = 11.96$, $p < .001$) when targets were alone than observed, and surprise ($t(32) = 9.55$, $p < .001$) when targets were observed than alone.

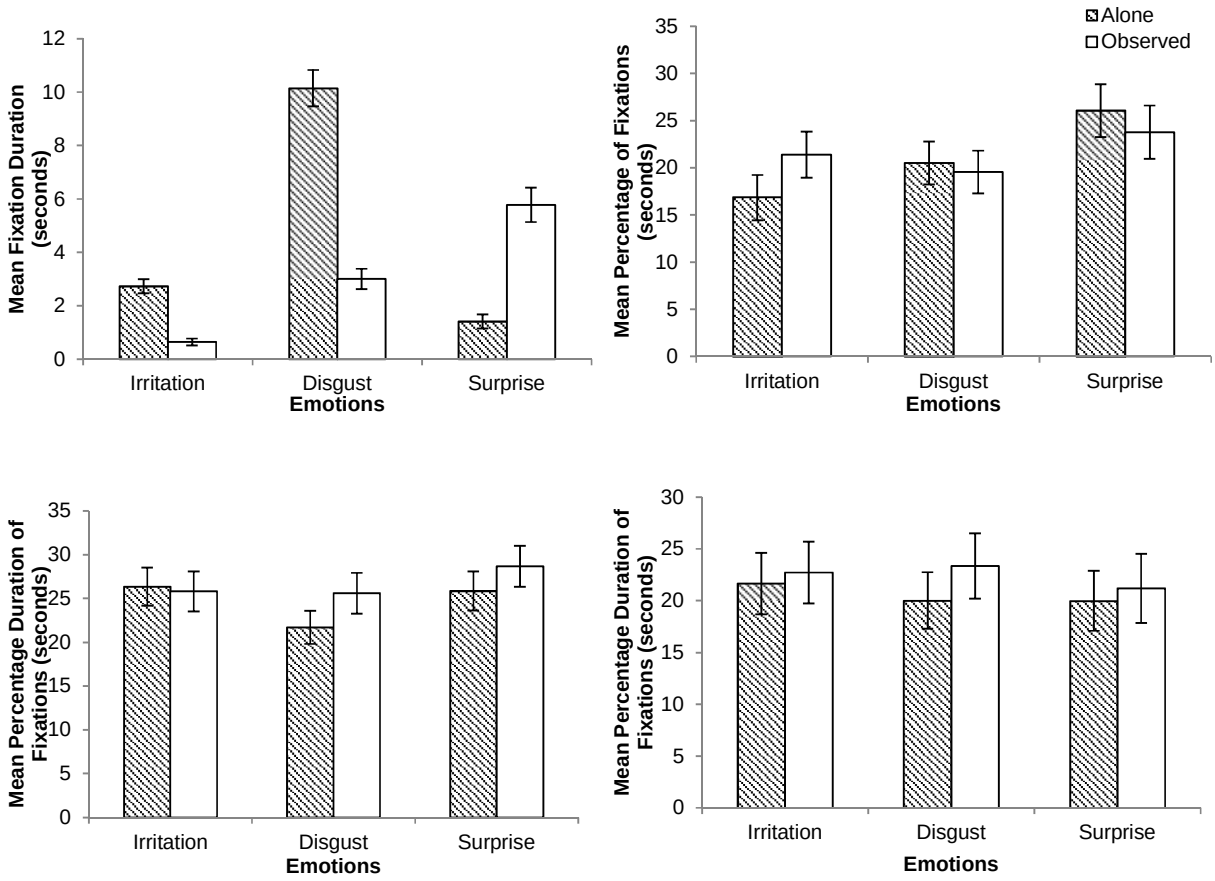

Fig. 6. Mean percentage duration of perceiver fixations (seconds) for conditions in which targets were alone and observed for the different AOIs in Experiment 2. The standard errors of the mean are represented by the error bars. (Top left: Body region, Top right: Mouth region, Bottom left: Nose region, Bottom right: Eyes region).

For the Mouth region there was a large significant main effect of Emotions, $F(2, 64) = 43.70$, $p < .001$, $\eta_p^2 = .58$, and a large significant interaction between Emotions and Conditions, $F(2, 64) = 20.72$, $p < .001$, $\eta_p^2 = .39$. Posthoc paired-samples *t*-tests showed that perceivers focused less on the Mouth when targets felt irritated when alone than observed, $t(32) = -5.91$, $p < .001$, and looked longer at the Mouth for surprise alone than observed, $t(32) = 2.91$, $p = .007$.

For the Nose region there was a large significant main effect of Emotions, $F(1.65, 52.78) = 12.34$, $p < .001$, $\eta_p^2 = .28$, a large significant main effect of Conditions, $F(1,32) = 18.64$, $p < .001$, $\eta_p^2 = .37$, and a medium significant interaction between Emotions and Conditions, $F(2,64) = 5.09$, $p = .009$, $\eta_p^2 = .14$. Posthoc *t*-tests revealed that perceivers spent less time gazing at the Nose region of the target for disgust alone than observed, $t(32) = -3.47$, $p = .001$, and surprise alone than observed, $t(32) = -2.76$, $p = .01$. For the Eyes region there was a large significant main effect of Conditions, $F(1,32) = 13.70$, $p = .001$, $\eta_p^2 = .30$, which resulted from perceivers attending to the region more when targets were observed than alone ($p = .001$).

As mentioned previously, the percentage fixation duration can be calculated by dividing the fixation duration of an AOI with the total fixation duration for each stimulus × 100 to control for the different durations of the stimuli. The results showed similar patterns to those reported above except for the one-way ANOVA that compared between the means of gaze time for the different areas of interest where a large significant main effect was found, $F(1.66, 52.97) = 21.03$, $p < .001$, $\eta_p^2 = .40$. For the Mouth region, there was a medium significant main effect of Conditions, $F(1,32) = 4.23$, $p = .048$, $\eta_p^2 = .12$, and posthoc paired-samples *t*-tests showed significant differences between alone and observed for irritation ($t(32) = -3.33$, $p = .002$, and surprise ($t(32) = 6.25$, $p < .001$, but not for disgust. For the Nose region, posthoc *t*-tests revealed significant differences between alone and observed for disgust ($t(32) = -3.67$, $p = .001$) only. Lastly, for the Eyes region, the ANOVA showed a large significant main effect of Emotions, $F(2,64) = 18.71$, $p < .001$, $\eta_p^2 = .37$, a large significant main effect of Conditions, $F(1,32) = 6.53$, $p = .02$, $\eta_p^2 = .17$, and a large significant interaction between Emotions and Conditions, $F(2,64) = 9.31$, $p < .001$, $\eta_p^2 = .23$. Posthoc paired-samples *t*-tests revealed significant differences between alone and observed for disgust, $t(32) = -5.19$, $p < .001$.

In summary, the eye-tracking results revealed that perceivers' gaze time differed between the different regions when targets were observed and alone for the various emotions. As with the first experiment, it was not the case that perceivers simply looked more at any particular region when the target was alone, suggesting that perceivers were not influenced by any low-level clues in the behaviour of targets that might reveal their social context; rather, the relationship between time perceivers spent looking at a region and whether the target was alone or observed depended on the particular kind of emotion that the target was experiencing.

5. Experiment 3

Building upon the first two experiments, the main purpose of Experiment 3 was to investigate the link between perceivers' estimation of how expressive the target was and their judgments that the target was alone or accompanied. After viewing each target video, perceivers were asked to make both kinds of judgment, creating the opportunity to determine if judgments of social context are largely or even completely explained by estimations of target expressiveness. If so, then the process by which perceivers judge social context will be apparent – judging social context would be little more than a proxy of judging target expressiveness.

Experiment 3 also provides an opportunity to investigate the robustness of perceivers' ability to infer social context. Experiments 1 and 2 used a common set of target stimuli, in which targets watched movie clips. In Experiment 3, a new set of stimuli were created with European (or European descent) targets reacting not to videos but to photographs from the International Affective Picture System (IAPS, Lang, Bradley & Cuthbert, 2008). An advantage of using IAPS is that presentation time can be brief and constant across all stimuli; also, the stimuli are pre-validated with values indicating how emotionally positive or negative each picture is.

A further potentially important modification was made to the social context experienced by the targets. In Experiments 1 and 2, the Experimenter, when present in the room, actively observed the target as he or she viewed the stimuli. In Experiment 3, in contrast, the Experimenter did not openly observe the target but quietly sat beside him or her while they looked together at the screen showing IAPS. In order for the experimenter's presence to cause an impact on target reactions that is perceptible to perceivers, perhaps it is necessary for the experimenter to be openly and actively observing the target. In that case, the results of Experiments 1 and 2 will not be replicated in Experiment 3. Alternatively, if the experimenter's mere presence is sufficient to moderate the target's reaction in a way that is perceptible to perceivers, then perceivers will be able to discriminate between the two kinds of social context in Experiment 3, just as they did in Experiments 1 and 2. One further modification was that perceivers used rating scales not only in estimating how expressive targets were but also in rating the likelihood that the target was accompanied.

5.1 Stimuli Development for Experiment 3

Targets viewed 20 IAPS images on a laptop screen while the laptop's webcam surreptitiously recorded them. The pictures included 2 positive, 2 negative and 16 neutral images and were selected based on their valence ratings (Lang et al., 2008). The positive and negative images were the focal images and target reactions to these were pre-selected for presenting to perceivers. These images were numbered 1722, 2080, 3170, 9410 in the IAPS data base. The two positive images were selected based on high valence ratings (7.04, 8.09 on a 10-point scale, where a hypothetical score of 10 is most positive), while the two negative images were selected to have low valence ratings (1.46, 1.51). The neutral images had valence ratings in the range 4-6.

5.1.2. Targets

Twenty target participants (10 males, 10 females) were selected under a convenience sampling method. They were mostly students from the University of Nottingham, aged 18-28 years (*M*= 22.8 years, *SD*= 2.42 years). Targets were predominantly British (16) but included other nationalities: 1 German, 1 Polish, 1 Italian and 1 American. Informed consent was obtained from all before the experiment began, which was approved by the University of Nottingham Ethics Committee. Targets received an inconvenience allowance of £2 for their participation.

5.1.3. Materials and Apparatus

The IAPS images were presented to targets on an Asus X53z laptop using Psychopy (Version 1.82.01, Peirce, 2015). The images were displayed in full screen (15.6 inches) and the order was randomised. The internal HD webcam recorded targets' reactions discretely while the webcam light was disabled to conceal the recording. Targets sat approximately 60-70cm from the screen to ensure the camera captured their face and bodies. Windows Movie Maker was used to record from the webcam.

5.1.4. Procedure

Targets were seated in a small Lab with no windows in the School of Psychology, University of Nottingham. The webcam was set up prior to their arrival such that targets remained unaware of the camera recording as they viewed the photographs. Targets were told that they were about to participate in a memory task and would be presented with 20 images, lasting 5 seconds each, and that they would subsequently perform a cued recall task. They were also informed that after viewing 10 images there would be a two-minute break.

Half of the targets were accompanied for the first half of the task and were alone for the second half. The rest of the targets experienced the two social contexts in the opposite order (alone followed by accompanied). The experimenter's presence or absence was managed in the same way as was the case in Experiments 1 and 2. In the accompanied condition, the experimenter, who was wearing casual clothes (not a lab coat, unlike in Experiments 1 & 2) sat beside the target and quietly watched the screen with them.

The 20 IAPS images were presented randomly (5 seconds for each image) using Psychopy (but with a two-minute break in the middle during which the experimenter either left the room or returned to the room). Once targets had completed the task, they were informed of the true purpose of the experiment and were made aware that they had been filmed. All consented to the use of their videos in subsequent experiments.

5.1.5. Editing

Targets' expressions in reaction to the 4 focal IAPS images were captured by the laptop's internal webcam using Windows Movie Maker. VLC Media Player was used to edit each clip, and each was then individually saved as a Windows Media Audio/Video file. Each of the 80 edited clips (20 targets each viewing 4 IAPS) prepared for the second phase of the experiment ranged between 4 and 6 seconds in duration.

5.2. Perceiver Phase Method

Thirty perceiver participants (15 males, 15 females), aged 18-80 years (M = 29.80 years, SD = 14.48 years) were selected under a convenience sampling method. Participants' nationalities included 20 British, 3 Pakistani, 3 Ukrainian, 1 Greek, 1 German, 1 Chinese and 1

Norwegian. Informed consent was obtained from perceivers before the experiment began and all received an inconvenience allowance of £2.

The perceivers viewed 80 silent video clips of targets who were either accompanied or alone as they watched positive and negative IAPS. Perceivers had no information about the stimuli being viewed by targets – they did not know that targets were viewing photographs and they were not told that some of these were positive and some negative. The video clips were 12x16 cm in size and presented in the centre of the screen in a randomised order using Psychopy (Version 1.82.01, Peirce, 2015) on an Asus X53z laptop (15.6 inch screen).

After viewing each target video clip, perceivers used a mouse to rate how expressive targets were on a scale ranging from 1-6 (1 = Non-expressive … 6 = Very expressive); they also rated the likelihood that the target was accompanied on a similar scale (1 = Not likely … 6 = Very likely). Perceivers were not given any additional information about the social contexts the targets experienced – unlike in Experiments 1 and 2, they were not shown photographs of the experimenter wearing a white lab coat. After each video clip ended, the rating scales appeared side by side in counterbalanced position on an otherwise empty screen. There was no time limit and as soon as perceivers had completed the two ratings the next video appeared. Ratings were automatically recorded in a data output file in Microsoft Excel.

5.3. Results and Discussion

Figure 7 shows perceivers' ratings that the target was likely accompanied as the target viewed positive or negative photographs. A 2 (accompanied or alone) × 2 (positive or negative photographs) repeated measures ANOVA was conducted on perceivers' ratings that the target was likely to be accompanied. A large significant main effect indicated that perceivers judged it

was more likely that the target was accompanied when indeed that was truly the case, $F(1, 29) = 14.32$, $p < .001$, $\eta_p^2 = .33$. There was no main effect associated with the valence of the stimuli, $F(1, 29) = 1.55$, $p = .22$, $\eta_p^2 = .05$) and no significant interaction between these factors, $F < 1$.

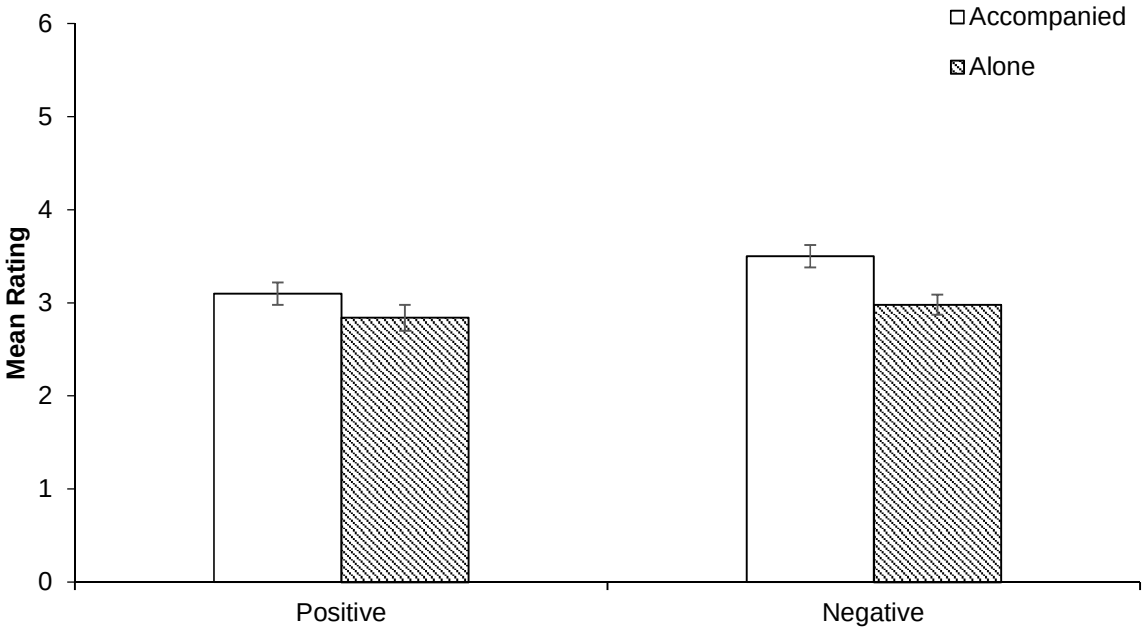

Fig. 7. The mean ratings (out of 6) by perceivers that the target was likely to be accompanied when targets were either accompanied or alone and when they were viewing either positive or negative photographs. The standard errors of the mean are represented by the error bars.

Figure 8 shows perceivers' ratings of how expressive the targets were. A 2 (accompanied or alone) × 2 (positive or negative photographs) repeated measures ANOVA revealed a large significant main effect indicating that perceivers judged accompanied targets to be more expressive than targets who were alone, $F(1, 29) = 19.64$, $p < .001$, $\eta_p^2 = .40$. Another large main effect indicated perceivers judged that targets who were viewing negative pictures were more

expressive than targets who viewed positive pictures, $F(1, 29) = 82.44$, $p < .001$, $\eta_p^2 = .74$. There was also a large significant interaction between these factors, $F(1, 29) = 43.73$, $p < .001$, $\eta_p^2 = .60$. Post-hoc *t*-tests showed that perceivers judged targets who were viewing positive images to be more expressive when accompanied than when alone, $t(29) = 8.16$, $p < .001$; conversely, targets viewing negative images were adjudged to be more expressive when alone than when accompanied, $t(29) = 2.34$, $p = .03$. When targets were alone, perceivers judged them to be more expressive as they viewed negative images than as they viewed positive images, $t(29) = 10.22$, $p < .001$; no such difference was apparent when targets were accompanied.

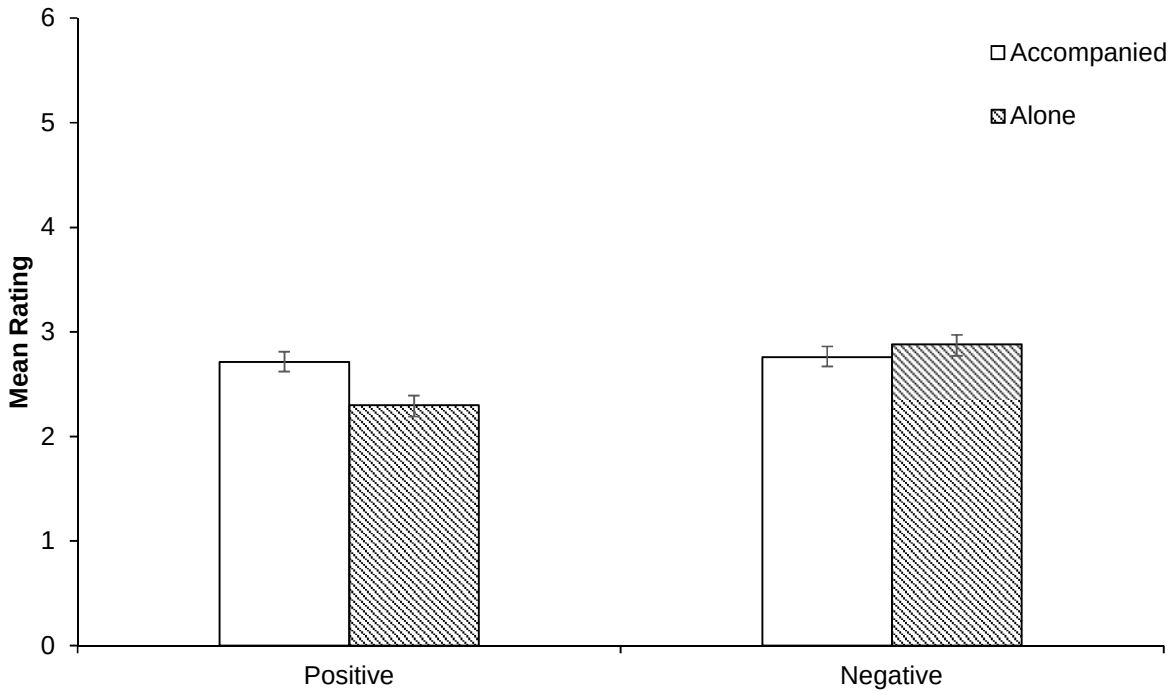

Fig. 8. The mean ratings (out of 6) by perceivers that targets were expressive when targets were either accompanied or alone and when they were viewing either positive or negative photographs. The standard errors of the mean are represented by the error bars.

Each perceiver made two ratings in relation to each of the 80 target video clips. In order to determine if perceivers' judgments that the target was accompanied were based on judgments that the target was expressive, we performed correlation analyses on each perceiver's set of ratings. The resulting 30 correlation values (one value for each perceiver) ranged from -.66 to .91, with a mean of r = .32 (SD = .35). Twenty-five of these 30 correlation values were significantly positive (p <.05). On average, then, perceivers who judged that the target was expressive also tended to judge that the target was likely to be accompanied. However, the mean correlation value was not high and there was a wide distribution that extended well into the negative range, suggesting that some perceivers who judged that targets were expressive also judged that targets were alone; and they judged that targets who were inexpressive were accompanied!

In view of the finding that perceivers who judged targets to be expressive also somewhat tended to judge targets were accompanied, does it also follow that perceivers who were successful in identifying the target's social context based their judgments in a simple way on their estimation of target expressivity? In other words, was it that case that success in identifying social context depends on judging that targets who are expressive are accompanied and judging that targets who are inexpressive are alone? To address this question, it was first necessary to create a value for each perceiver that represented success in identifying social context, and *d*-prime was used for this purpose. To calculate *d*-prime, we coded likelihood ratings in the range of 4-6 as the perceiver judging that the target was accompanied; we coded ratings in the range of 1-3 as the perceiver judging that the target was alone. We then compared these categorical codings against the true social context experienced by the target, thus allowing us to calculate hits, false alarms, correct rejections and misses. This provided sufficient information to calculate

a *d*-prime value for each perceiver, a value that represented how successful each was in identifying the target's social context. Across the 30 perceivers, the mean *d*-prime value was .24 (*SD* = .34, range = -.38 to .98). If perceivers as a group were unable to identify social context then the mean *d*-prime value would not be above zero. As expected, though, the observed value was significantly greater than zero, $t(29) = 3.88$, $p < .001$, which is consistent with the finding reported above, showing that perceivers were more likely to rate the target as being accompanied when that was truly the case than when it was not the case.

Based on *d*-prime values, it seems that some perceivers achieved a relatively high score, indicating success in identifying target social context, while some gained a relatively low score, indicating lack of success. We compared each perceiver's *d*-prime score with their tendency to link judgments that the target was likely accompanied with estimations that the target was expressive. Such a tendency to link these two ratings was indexed by each perceiver's correlation value based on estimations of social context and expressivity for 80 video clips of targets. However, there was no evidence to suggest that perceivers' success in identifying social context was related with an index of perceivers' tendency to link judgments of social context with judgments of target expressivity, $r(29) = -.14$, $p = .45$.

One reason for conducting Experiment 3 was to investigate whether the relationship between perceivers' judgments that the target was accompanied and that the target was expressive varied depending on whether the target was viewing a positive or negative photograph. Specifically, we might expect that when targets were viewing a positive photograph then perceivers who judged the target to be accompanied (with a high likelihood rating) should also judge the target to be expressive (with a high expressiveness rating); in contrast, perceivers who judged the target to be alone should judge the target to be inexpressive. In this case, then,

when targets were viewing a positive photograph, we might expect on average to find a positive correlation between perceivers' ratings that the target was accompanied and their ratings that the target was expressive and such an expectation was supported by the data (mean r = .28, SD = .37); this mean r-value was significantly positive, $t(29) = 4.21$, $p < .001$.

When targets were viewing negative photographs we might expect a different kind of relationship between perceivers' ratings of target social context and ratings of target expressiveness. In this case, we might expect perceivers who judged the target to be alone (as indicated by a low likelihood rating) to judge the target to be expressive (as indicated by a high expressiveness rating); in contrast perceivers who judged the target to be accompanied (with a high likelihood rating) should judge the target to be inexpressive. We should thus find a negative correlation between perceivers' ratings that the target was accompanied and their ratings that the target was expressive when the target was viewing negative photographs. Ironically, though, the mean correlation value across the 30 perceivers' not only was significantly positive (mean r=.36, SD = .36), $t(29) = 5.51$, $p < .001$, but was significantly more positive than the mean correlation value reported above, associated with the positive photo condition, $t(29) = 2.17$, $p = .04$.

In summary, the results of Experiment 3 show that as a group, perceivers were able on average to determine when targets were accompanied and when they were alone. Perceiver judgments of how expressive the target was also discriminated between the two target social contexts. Moreover, when perceivers judged that the target was expressive they were also inclined to judge that the target was accompanied. However, linking judgments of social context with judgments of expressivity did not explain or account for perceivers' success in identifying target social context. Is it possible, then, that a low-level cue allowed perceivers to guess the social context of the target? Specifically, did targets look away from the screen (and towards the

experimenter) more often when the target was accompanied than when alone? Each stimulus presented to targets appeared only for a few seconds, giving targets little opportunity to look away from the screen. Out of 80 video clips, targets looked away from the screen on a total of 11 occasions, six when the target was accompanied and five when the target was alone. Hence, there is no basis for suggesting that perceivers determined target social context from the low-level cue of targets looking away from the screen. Rather, it seems that perceivers interpreted other clues (not related with target expressivity) in successfully inferring the target's social context.

6. General Discussion

The purpose of this research was to explore the scope and process of human ability in knowing the world through other minds, a process known as 'retrodictive' mindreading (Gallese & Goldman, 1998). Previous research (e.g., Pillai et al., 2012, 2014; Sheppard et al., 2016) suggests that people are capable of making a sequence of two 'backwards' inferential steps from the behaviour of a target to their mental state (the proximal cause) and then from their mental state (as embodied in their behaviour) to the antecedent cause of that mental state (the distal cause). In that research, perceivers could make somewhat accurate inferences about the content and manner of the experimenter's communication with the target. The results presented here take us into new territory: From the manner of the target's emotional reaction to a video or photo, perceivers were able to infer whether the target was alone or accompanied. People can thus interpret another person's expression of a mental state to determine their social context.

The results, particularly those of Experiments 2 and 3, speak to the process by which perceivers inferred the social context of targets from an embodiment of their mental states. When targets were viewing positive stimuli, perceivers judged that targets were more expressive (or

'exaggerating' their expressions) when it so happened the target was accompanied than when alone. When targets were viewing negative stimuli, in contrast, perceivers judged that targets were less expressive (or 'hiding' their expressions) when targets were accompanied than when alone. These results thus show that perceivers were sensitive to the social context of the targets in their judgments of how targets controlled their expressions. Furthermore, the results also provide information about the basis on which perceivers explicitly judged that targets were accompanied or alone in Experiments 1 and 3, though the relationship between perceiver judgments of target expressivity and judgments of target social context requires careful consideration, as explained below.

The results of Experiment 3 reveal a weak but reliable relationship between perceivers judging that the target was expressive and judging that the target was accompanied. Is it simply the case, then, that perceivers were able to determine target social context in so far as they judged that targets were expressive? Further analysis of data reported in Experiment 3 revealed no basis for supposing that perceivers who linked judgments between target expressiveness and judgments or target social context were successful in determining target social context. Consequently, it appears that perceivers' judgments of target social context were affected by two factors: (1) a factor, whatever that might be, that allowed them to make correct judgments about social context that was independent of judgments that the target was expressive; (2) bias in perceivers to judge that expressive targets (rightly or wrongly) were accompanied. West and Kenny (2011) argued that variance in judgments about others' mental states can be partitioned into at least two components: (1) variance associated with correct performance; (2) variance associated with bias. West and Kenny's model is useful in interpreting our data in the sense that 'bias', if not identified for what it is, could be wrongly interpreted as informing us about the

process involved in correct performance. In the light of West and Kenny's model, it seems that perceivers' performance in determining target social context is not explained by supposing that target's linked judgments of social context with judgments that the target was accompanied – such a tendency merely reflects bias.

Nevertheless, it could have been that perceivers linked target expressivity with target social context in a more sophisticated way. We know from classic research in social psychology that people are more expressive when accompanied than when alone if they are experiencing positive emotions (e.g. Buck, 1984); conversely, it seems people are less expressive when accompanied than when alone if experiencing negative emotions (Spain et al, 2000). Perceivers in our studies might have had sufficient insight into this social process to be able to determine target social context from their perception of how the target's emotional reaction seems to be moderated (or controlled) depending on whether the target seemed to be having a positive or a negative emotion. However, the results of Experiment 3 offered no evidence to suggest in particular that perceivers had any such insight: The relationship between perceivers' judgments about target social context and target expressivity did not vary in the expected way, depending on whether the target was having a positive or negative emotion.

Having ruled out two possible explanations for the basis on which perceivers inferred target social context, it remains a matter for speculation on how perceivers actually did infer target social context. Apparently, perceivers interpret another aspect of target behaviour, not how expressive the target is, in determining whether the target is accompanied or alone. Perhaps people who are accompanied are more aroused and visibly more alert than people who are alone. In other words, perhaps perceivers are effectively basing their judgments of target social context

on how aroused the target appears, irrespective of whether the target is having a positive or negative emotion (Kraut, 1982).

Notwithstanding, we should consider whether perceivers were using a low-level strategy to determine whether the target was observed or alone. For example, did targets look away from the screen (and look instead at the experimenter) more often when the experimenter was present than absent? If so, perhaps perceivers were judging social context not based on the manner of the target's emotional reaction but on whether the target seemed to be distracted by the presence of another person in the room. There was no evidence to suggest that targets looked away from the screen more often in the accompanied than in the alone condition, meaning that perceivers could not use the apparent distractibility of the target as a clue to social context. Besides, in Experiment 2 perceivers were told nothing of the social context that sometimes the target was alone and sometimes observed; indeed, they were not even told anything about the videos being watched by the targets in any of the experiments. In Experiment 2 they merely had to judge whether the target was hiding or exaggerating their expression and they did so in a way that effectively showed they were sensitive to the two social contexts of the target. In Experiment 3, when perceivers judged how expressive the targets were, a very similar pattern emerged.

Data from Experiments 1 and 2 are consistent in showing that eye movements of the perceivers were different for several different areas of interest depending on the social context. It was not that the perceiver simply looked more at certain areas when the target was observed than alone (or the opposite); rather the patterns of eye movements revealed a complicated relationship between the social context of the target and the particular emotion that he or she was experiencing. This suggests once again that the social context mediated the target's response to the video; and this in turn impacted upon the visual attention of the perceiver in a way that

demonstrated sensitivity in the perceiver to the context (and the emotions experienced by the target) at the level of eye movements.

In Experiments 1 and 3, perceivers were explicitly asked about the target's social context. One might think that this instruction could have influenced the perceivers' eye movements. For example, they could have looked for clues as to whether the target was looking away from the screen and (by implication) at the observer. If so, one might expect the eye movement patterns to be quite different in Experiments 1 and 2 in that perceivers were not given any information about the targets' social context in the latter. While the patterns of data are not identical in the two experiments, nevertheless perceivers' eye movements suggested they were sensitive to the target's social context in both: This effect was not confined to Experiment 1. Notably, perceivers were demonstrably sensitive to conditions in which targets were alone and observed even in a circumstance where it is implausible to suppose that the perceiver was trying to detect if the target was looking away from the screen (Experiment 2). Moreover perceivers were not simply looking more at a particular region (such as the eye region) in the observed than in the alone condition; how much they looked at any particular region also depended upon the particular emotion the target was experiencing.

Having discussed how perceivers infer target social context by interpreting an embodiment of the target's mental state, we now turn to *why* they make such inferences. Assuming that any given ability develops because it is functional and adaptive (in most cases), it seems fair to consider the value of the particular ability that allows us to infer social context by interpreting clues in the target's behaviour (Tinbergen, 1963). Presumably, the ability in question broadly enables perceivers to know things about the world inhabited by the target; making inferences about social context is but a specific implementation of this broader ability. Other

examples demonstrated in past research are guessing what another person said to the target (Pillai et al., 2012, 2014; Sheppard et al., 2016), guessing what gift had been offered to the target (Cassidy et al., 2013, 2015) and guessing which stimulus a target selected as their preferred choice (North et al, 2010). The full scope of the ability to infer events in the world from a target's embodied mental state awaits further exploration: Perceivers might be able to guess the gender of a person interacting with the target, they might be able to guess the expression (positive or negative) of a person the target is interacting with and they might be able to guess which cue word a target is looking at (pride, embarrassment, happiness, shame) in a context where targets are instructed to think of a time they experienced the emotion indicated by the cue word.

In essence, the results presented here might testify to the broader ability of people knowing the world through the lens of another mind. According to Mitchell (1996), this basic process serves as the foundation for a more advanced ability to learn from other people. In its most highly developed form this could include the kind of explicit instruction from others that is formalised in education. Here, students learn things about the world that extends well beyond what they could reasonable expect to know from their own personal discovery. Accordingly, the expansive knowledge and wisdom enjoyed by humans depends on an underlyinbg ability to apprehend aspects of the world through the lens of other minds (cf Vygotsky, 1978).

Such ability seemingly has its roots in early development. From the age of about 18 months, infants begin to show signs of 'social referencing' (Sorce et al., 1985), meaning that they look to their caregiver and are influenced by the mental state or attitude that is embodied in his or her facial expression. If the caregiver's mental state is positive then the infant is likely to be more adventurous in an unfamiliar environment; if the mental state is negative they will be

less adventurous. Apparently, then, the infant seems to learn something about the status of the environment (whether or not it is safe) by interpreting the caregiver's attitude towards that particular aspect of the environment. Perhaps the basic capacity to learn about the world from other minds is refined during development, ultimately to a point where we are able to determine such things as whether a target is alone or accompanied, what a third party said to a target or what a third party gave to a target.

In summary, the three experiments reported here yielded a variety of data which converge in showing that perceivers are sensitive to the social context of the target. This is apparent in their explicit judgments on whether the target is observed or alone and it is apparent in their judgments of how the target was controlling their expressions and in their eye movements. The explicit judgments might be made on the basis of knowledge (either explicit or implicit) that people display their emotions differently in the presence or absence of others depending on the particular emotion in question. Such an intuition, if it deserves to be called that, accords with research into the effect an audience has on the intensity of emotional displays (e.g. Ansfield, 2007; Spain et al., 2000). In consequence, the evidence testifies to the ability of our perceivers as natural psychologists.

References

Ansfield, M. E. (2007). Smiling when distressed: when a smile is a frown turned upside down. *Personality & social psychology bulletin*, *33*(6), 763–75. doi:10.1177/0146167206297398.

Baron-Cohen, S., Leslie, A.M. & Frith, U. (1985). Does the autistic child have a "theory of mind"? *Cognition*, *21*, 37-46.

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 and Allied Disciplines*, *38*, 813-822.

Buck, R. (1984). *The communication of emotion*. New York: Guilford Press.

Cassidy, S., Ropar, D., Mitchell, P., & Chapman, P. (2013). Can Adults With Autism Spectrum Disorders Infer What Happened to Someone From Their Emotional Response?. *Autism Research*, *7*(1), 112-123. doi:10.1002/aur.1351.

Cassidy, S., Ropar, D., Mitchell, P., & Chapman, P. (2015). Processing of spontaneous emotional responses in adolescents and adults with Autism Spectrum Disorders: effect of stimulus type. *Autism Research*, *8*, 534-544.

Cohen, J. (1992). A power primer. *Psychological Bulletin*, *112*(1), 155.

Boyle, D., 127 hours, 2010, Twentieth Century Fox Film Corporation; United States of America.

Conan Doyle, A. (1902). *The Hound of the Baskervilles*. London: George Newnes.

Darwin, C. (1872). *The origin of species by means of natural selection* (6th ed.). London: John Murray.

Ekman, P. (1972). Universal and cultural differences in facial expressions of emotion. In J. Cole (Ed.), *Nebraska symposium on motivation, 1971* (Vol. 19, pp. 207–282). Lincoln: University of Nebraska Press.

Ekman, P., & Friesen, W. V. (1975). *Unmasking the face*. Englewood Cliffs, N.J.: Prentice-Hall.

Fridlund, A. J. (1991). Evolution and facial action in reflex, social motive, and paralanguage. *Biological Psychology, 32*(1), 3-100.

Frith, C. D., & Frith, U. (2012). Mechanisms of social cognition. *Annual Review of Psychology, 63*, 287–313.

Gallese V., & Goldman A. (1998) Mirror neurons and the simulation theory of mind-reading. *Trends in cognitive sciences, 2*(12), 493-501.

Hurford-Jones, G., Britain's got talent, broadcast April 11, 2009, ITV; Great Britain.

Ickes, W. (2001). Measuring empathic accuracy. *Interpersonal sensitivity: Theory and measurement*, 219-241.

Ickes, W. (2009). Empathic accuracy: its links to clinical, cognitive, developmental, social, and physiological psychology. *The social neuroscience of empathy*, 57-70.

Kilbride, J. E., & Yarczower, M. (1980). Recognition and imitation of facial expressions: A cross-cultural comparison between Zambia and the United States. *Journal of Cross-Cultural Psychology, 11*(3), 281–296.

Klin A., Jones W., Schultz R., Volkmar F., & Cohen D. (2002). Visual fixation patterns during viewing of naturalistic social situations as predictors of social competence in individuals with autism. *Archives of General Psychiatry 59*(9), 809-816.

Kraut, R. E. (1982). Social presence, facial feedback, and emotion. *Journal of Personality and Social Psychology, 42*(5), 853–863. doi:10.1037//0022-3514.42.5.853.

Lang, P. J., Bradley, M. M., & Cuthbert, B. N. (2008). International affective picture system (IAPS): Affective ratings of pictures and instruction manual. *Technical report A-8.*

Lee, A. (2010). VirtualDub (Version 1.9.11) [Software]. Available from http://sourceforge.net/projects/virtualdub/files/virtualdub-win/1.9.11.32842/.

Bates R., Jr. & Lewis, D. H., Excision, 2012, Anchor Bay Entertainment; United States of America.

metallicKuma. (2009, August 7). *The most irritating 9:36.47 minutes of your life* [Video file]. Retrieved from http://www.youtube.com/watch?v=z-TykljBz-E.

Mitchell, P. (1996). *Acquiring a Conception of Mind: A review of Psychological Research and Theory*. Hove: Psychology Press.

mrsmithereen. (2005, July 30). *Ghost Car* [Video file]. Retrieved from http://www.youtube.com/watch?v=Wz1W_omigwg.

North, M. S., Todorov, A., & Osherson, D. N. (2010). Inferring the preferences of others from spontaneous, low-emotional facial expressions. *Journal of Experimental Social Psychology*, *46*(6), 1109-1113.

Peirce, J. W. (2012). PsychoPy (Version 1.74.01) [Software]. Available from http://sourceforge.net/projects/psychpy/.

Peirce, J. W. (2015) PsychoPy (Version 1.82.01) [Software]. Available from https://github.com/psychopy/psychopy/releases

Pillai, D., Sheppard, E., & Mitchell, P. (2012). Can people guess what happened to others from their reactions? *PloS one*, *7*(11), e49859. doi:10.1371/journal.pone.0049859.

Pillai, D., Sheppard, E., Ropar, D., Marsh, L., Pearson, A., & Mitchell, P. (2014). Using Other Minds as a Window onto the World: Guessing What Happened from Clues in Behaviour. *Journal of Autism and Developmental Disorders, 44*(10), 2430-2439.

Rogers, C. R. (1957). The necessary and sufficient conditions of therapeutic personality change. *Journal of consulting psychology*, *21*(2), 95.

Sheppard, E., Pillai, D. Wong, G. T. L., Ropar, D., & Mitchell, P. (2016). How easy is it to read the minds of people with Autism Spectrum Disorder? *Journal of Autism and Developmental Disorders*, *46*, 1247-1254.

Shurken12345. (2008, December 23). *I will respect anyone who can watch the whole thing!!!* [Video file]. Retrieved from http://www.youtube.com/watch?v=vxOuECN03RQ.

Sorce, J.F., Emde, R.N., Campos, J., & Klinnert, M.D. (1985). Maternal emotional signaling: its effects on the visual cliff behavior of 1-year-olds. *Developmental Psychology*, *21*, 195–200.

Spain, J. S., Eaton, L. G., & Funder, D. C. (2000). Perspectives on personality: the relative accuracy of self versus others for the prediction of emotion and behaviour. *Journal of Personality*, *68*(5), 837–67.

Tan, C. B. Y., Stephen, I. D., Whitehead, R., & Sheppard, E. (2012). You look familiar: how Malaysian Chinese recognize faces. *PloS One*, *7*(1), e29714. doi:10.1371/journal.pone.0029714.

Tinbergen, N. (1963). On aims and methods of ethology. Zeitschri & fur Tierpsychologie, 20, 410-433.

Vygotsky, L.S. (1978). Mind in Society. Cambridge, MS: Harvard University Press.

West, T. V., & Kenny, D. A. (2011). The truth and bias model of judgment. *Psychological review*, *118*(2), 357.

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.

Wu, W., Sheppard. E. & Mitchell, P. (2016). Being Sherlock Holmes: Can we sense empathy from a brief sample of behavior? *British Journal of Psychology, 107*, 1-22.

Zaalberg, R., Manstead, A., & Fischer, A. (2004). Relations between emotions, display rules, social motives, and facial behaviour. *Cognition & Emotion*, *18*(2), 183–207. doi:10.1080/02699930341000040.

Zaki, J., & Ochsner, K. (2011). Reintegrating the study of accuracy into social cognition research. *Psychological Inquiry*, *22*(3), 159-182.