Gaze patterns in viewing static and dynamic body expressions

Correspondence

Petra M. J. Pollux (PhD)

School of Psychology

University of Lincoln

Brayford Pool

LN6 7TS

Lincoln

E-mail: ppollux@lincoln.ac.uk

Tel: +44(0)1522886360

Matthew Craddock (PhD)

School of Psychology

University of Lincoln

Brayford Pool

LN6 7TS

Lincoln

E-mail: mcraddock@lincoln.ac.uk

Tel: +44(0)152288

Kun Guo (Professor)

School of Psychology

University of Lincoln

Brayford Pool

LN6 7TS

Lincoln

E-mail: kguo@lincoln.ac.uk

Tel: +44(0)152288

Abstract

Evidence for the importance of bodily cues for emotion recognition has grown over the last two decades. Despite this growing literature, it is underspecified how observers view whole bodies for body expression recognition. Here we investigate to which extent body-viewing is face- and context-specific when participants are categorizing whole body expressions in static (Experiment 1) and dynamic displays (Experiment 2). Eye-movement recordings showed that observers viewed the face exclusively when visible in dynamic displays, whereas viewing was distributed over head, torso and arms in static displays and in dynamic displays with faces not visible. The strong face bias in dynamic face-visible expressions suggests that viewing of the body responds flexibly to the informativeness of facial cues for emotion categorisation. However, when facial expressions are static or not visible, observers adopt a viewing strategy that includes all upper body regions. This viewing strategy is further influenced by subtle viewing biases directed towards emotion-specific body postures and movements to optimise recruitment of diagnostic information for emotion categorisation.

Keywords:

Body expressions, dynamic displays, face visibility, eye-movements.

1. Introduction

The ability to recognise emotional state in others is essential for social interaction and has been shown to predict better social adjustment, mental health and workplace performance (Carton, Kessler, & Pape, 1999; Izard, Fine, Schultz, Mostow, Ackerman et al., 2001; Norwicki & Duke, 1994). Whilst facial expressions are considered to be the strongest predictors for emotional state within social interaction distance (Adolphs, 2002; Ekman, 1992), in daily life we also receive other emotional cues from the person, such as posture, gestures, vocalisations and emotion tone in speech. These cues can be aligned, thereby facilitating recognition of an emotion, or they can be in conflict and lead to confusion in emotion perception. For instance, a frightened person (expressed in bodily cues) with a happy facial expression is categorised more often as frightened compared to when both facial and bodily cues indicate a happy emotional state (Conty, Dezechache, Hugeueville, & Grèzes, 2012; Jensen & Kotz, 2011; Kokinous, Tavano, Kotz, & Schröger, 2017; Kreifelts, Ethofer, Shiozawa, Grodd, & Wildgruber, 2009; Meeren, Hadjikhani, Ahlfors, Hämäläinen, & de Gelder, 2008; Müller, Havel, Derntl, Schneider, Zilles, et al., 2011; Nelson & Mondloch, 2017; Stekelenburg & Vroomen, 2007; Yeh, Geangu, & Reid, 2016). To date, emotion research has focused predominantly on facial and vocal cues for investigating unimodal and multimodal emotion perception, whereas the role of bodily cues has received relatively less attention. Yet body cues provide critical information when the face is not clearly visible (de Gelder, 2009; de Gelder, de Borst, & Watson, 2015) or when facial expressions are ambiguous (Aviezer, Trope, & Todorov 2012) and emotional postures have been shown to modulate judgements from facial or vocal cues when perceived at the same time (Jensen & Kotz, 2011; Yeh et al., 2016). Moreover, body expressions have been found to activate action-related neural structures, suggesting that they are critical for judging action intentions and for response preparation (de Gelder, 2013; Engelen, Zhan, Sack, & de Gelder, 2018; Grezes, Pichon, & de Gelder, 2007; Meeren, Sinke, Kret, & Tamietto, 2010).

The salience of the body for human observers in social cognition is further reflected in the way viewers divide their attention over the face and the body. When viewing social scenes to identify a person, people tend to spend approximately 40% of the time looking at the body (Bindemann, Scheepers, Ferguson, & Burton, 2010). In addition, body movement (e.g. gait) has been shown to attract observers' attention when they view people for person identification (Rice, Phillips, Natu, An, & O'Toole, 2013). However, viewing of whole bodies for emotion recognition has not been investigated in great detail. A few results highlight the importance of bodily cues for perception of threat (Kret, Stekelenburg, Roelofs, & de Gelder, 2013). Using still images of people within a social context or alone, Kret et al. (2013) found that observers tend to view both the face and the body when they are visible in the images. Attention to the body increases when the emotions signalled by the body (threat) and face (happy) were incongruent, suggesting that bodily cues became more important when the emotional state is ambiguous (Kret et al., 2013). A similar effect of face cue ambiguity has been demonstrated in person identification (Rice et al., 2013), reflected in increased viewing of the body when the face is less informative for identity recognition. Several findings further suggest that the body is even more informative when movement is added in dynamic displays (Grezes et al., 2007; O'Toole, Phillips, Weimer, Roark, Ayyad, et al., 2011; Stoesz & Jakobson, 2014). For instance, compared to static images, dynamic displays of bodies with faces not visible increases person identification performance more than when motion is added to face stimuli alone (O'Toole et al., 2011). Moreover, compared to the face, the body is attended more often during free-viewing of dynamic social scenes compared to static images of these scenes (Stoesz & Jakobson, 2014). To date, it is not yet known whether a similar increase in attention to the body will emerge in viewing dynamic compared to static body expressions for emotion categorisation. To address this question, the present study will investigate gaze behaviour in viewing of static images (Experiment 1) or videos (Experiment 2) of whole body expressions at varying orientations with faces visible or not visible. Based on the findings discussed so far, it is expected that the body will attract more attention in dynamic compared to static displays and when the face is not visible compared to when it is visible.

Emotional body expressions have been described in terms of unique combinations of postures, gestures and muscle movements (Atkinson, Dittrich, Gemmel, & Young, 2004; Deal, Mortillaro, & Scherer 2012; Gunes, 2005; Gunes & Piccardi, 2006, Huis In ‘t Veld, van Boxtel, & de Gelder 2014). Similar to facial action coding system (FACS) for facial expressions, a few studies have developed coding systems for body expressions where each emotion is associated with a set of Body Action Units (BAU) (Dael et al., 2012; Gunes et al., 2006). The description of these BAUs shows considerable overlap across studies. For instance, in expressions of anger, people tend to lean forward and shake their fists or point at the cameras, whereas expressions of fear involves leaning backward and raising the arms in front of the body. For sadness, the head is often dropped and hands or arms are brought close to the body and for happiness the posture is upright, with arms raised (Atkinson et al., 2004; Atkinson, Tunstall, & Dittrich, 2007; Dael et al., 2012; Gunes et al., 2006). In addition to the postural cues, velocity is an important feature for discriminating between two emotions (Gunes et al., 2006; Atkinson et al., 2007; Gunes et al., 2015). For instance, movements for anger and happiness tend to be fast and jerky whereas movement tends to be minimal and slow for sad expressions. It is not yet known whether viewing behaviour is influenced by emotion-specific bodily cues in static and dynamic displays of whole body expressions. To investigate this question, the present study will record viewing measures separately for different body regions (head, arms, torso and legs) and for different emotions (happy, sad, fear, anger and neutral).

Emotion-specific viewing patterns have previously been observed in viewing of facial expressions (e.g. enhanced viewing of the smiley mouth for happy expressions) and are consistent with the idea that attention is drawn to facial features that are most diagnostic for different emotions (Calvo & Nummenmaa, 2008; Smith, Cottrell, Gosselin, & Schyns, 2005). A few other findings suggest however that viewing is relatively unaffected by the expressed facial expression of varying intensities (Guo, 2012). Guo argues that facial configural information (structural information about the spatial relations between local facial features) may be more informative to disambiguate subtle expressions. Under these circumstances, the use of a uniform viewing strategy that includes all facial regions may be more optimal (Guo, 2012). The importance of configural information in body expression recognition has previously been demonstrated in both static (Stekelenburg & de Gelder, 2004) and dynamic displays (Atkinsons et al., 2007). Atkinson et al. (2007) found for example that accuracy for fully lit dynamic body expressions (with faces covered) was reduced by inversion and reversion of the videos clips, suggesting that the processing of structural relations between body regions and temporal sequencing in body movements was disrupted. However, accuracy was still above chance in reversed and inversed videos suggesting that other information, such as emotion-specific body signals, remain important for body expression recognition (Atkinson et al., 2007). The results of the present study will reveal whether viewers focus more on bodily cues uniquely associated with different emotion categories, which would be reflected in emotion-specific viewing patterns, or whether viewers will adopt a more uniform viewing strategy to optimise information seeking for emotion categorisation. Body posture, such as forward lean for anger or a backward lean for fear, are expected to provide relevant diagnostic information in still images and may therefore draw attention to the body region when faces are not visible (Dael et al., 2012; Gunes et al., 2006). In contrast, attention may be drawn more to the diagnostic movements of the arms in dynamic displays given the diagnostic information they provide (i.e. pronounced movements in happy and angry expressions, minimal movement in sad and neutral displays) (Atkinson et al., 2004; Atkinson, Tunstall, & Dittrich, 2007; Dael et al., 2012; Gunes et al., 2006).

In summary, it is not yet known how observers view dynamic whole body expressions for emotion categorisation responses. The aim of the present study is to fill this gap by presenting observers with still and dynamic stimuli of whole body expressions of emotions with faces either visible or not visible. Based on previous findings, it is expected that the body will attract more attention in dynamic displays (Rice et al., 2013; Stoesz & Jakobson, 2014), and that attention is drawn to emotion-specific bodily cues (Dael et al., 2012, Gunes et al., 2006).

2. EXPERIMENT 1: Static body expressions

2.1 Methods

2.1.1 Participants

Forty-one young adults (17 males, age = 20.1 ± 0.24 (Mean ± SEM) years; 24 females, age = 19.9 ± 0.22 years) were recruited via the Subject Pool of the School of Psychology at the University of Lincoln. This sample size was based on previous research in the same field and was comparable to those published reports (e.g., Guo, 2012; Shaw & Guo, 2015; Pollux, Hall, & Guo, 2014). The suitability of the sample size was confirmed by power analysis using G*power software (Faul, Erdfelder, Lang, & Buchner, 2007). In our previous study the effect size (η^2) for the effect of Emotion on categorisation accuracy was .3 and for Emotion × ROI (region of interest) in viewing time analysis was .42 (Pollux et al., 2014). A sample of 22 participants in this study would have been large enough for an effect size of .3

to be detected with a power of .95 at alpha level .05 in a repeated measures design with 5 emotions to be categorised.

Informed written consent was obtained from participants and from the actors recruited for the videos and images. All participants had normal or corrected to normal vision. Ethical approval was obtained from the Ethics Committee in the School of Psychology, University of Lincoln. All procedures complied with the British Psychological Society ‘‘Code of Ethics and Conduct’’ and with the World Medical Association Helsinki Declaration as revised in October 2008.

2.1.2. Materials and Equipment

Recording: Six Caucasian final year student actors (3 men, 3 women, age = 20 ± 0.21 years) in the Drama and Theatre Studies course at the University of Lincoln were recruited for the recording of whole body expression videos. All actors gave written consent for the use of their videos for the purpose of research. Their clothing consisted of a grey t-shirt, dark trousers and grey socks. Shoes and jewellery were removed. All women had long hair tied back in a ponytail. All men had short hair and no facial hair. The actors were positioned against a black background. Visibility of postural and movement cues generally varies from different viewpoints. For instance, a forward or backward lean is likely to be more visible from a profile viewpoint whereas sideway movements (e.g. in fear) may be more observable from a frontal viewpoint. To optimise visibility of diagnostic cues for each emotional expression, recordings were made with three cameras; one camera was facing the actors (0˚), the other two cameras were positioned at a 45˚ and 90˚ angle (to the right hand side of the actors) at a distance of 3 meters. The emotional expressions were guided by scenarios read out to the actors before each recording. In all videos, the actors started with a neutral position: Arms were held loosely next to the body and they were facing forward with neutral facial expression. Actors were encouraged to express intense emotions and not to walk more than one small step in any direction from their starting position. In order to capture natural expressions of emotions as much as possible, no other instructions were given. The emotions included for the present study were happy, sad, angry, fear and neutral. For angry expressions, all actors leaned or stepped forward and made rapid arm movements (e.g. pointing). For happy expressions, all actors raised their arms, 68% raised their arms above the head, 20% jumped up and down. For sad expressions, 68% either crossed or held their arms over their abdomen, 12% touched their lower neck, 68% put their hands together and all actors dropped their heads at some point during the video. For fearful expression, 83% either crouched slightly or leaned backward and/or sideways and 83% raised their arms up away from the body. The expressed emotions fit well with previous descriptions of body action units for each emotion category (Dael et al., 2012, Gunes et al., 2006). The recordings from the three cameras (Nikon D90, resolution 1280 × 720 pixels) were synchronised in time, grey scaled and further processed to remove any small features drawing attention in the background. A second set of the same videos was created with the faces pixelated. The duration of the final videos were either 4 seconds (neutral condition) or 5 seconds (emotion conditions). The frame rate of the videos was 25Hz.

Still Images: Selection of the frames for still images was guided by the highest intensity of both the body and the facial expression within one frame. The same frames were used for the images with pixelated faces. While actors were instructed to direct their emotional expression to the frontal view camera and their gaze was directed to the frontal camera in all images, head and/or body orientation could deviate slightly from centre in the images. The images were further processed to reduce any background noise and were re-sized to ensure that the height of the different actors was the same (13°) in the starting position.

The width of the bodies varied between 6° and 13° due to the different positions of the arms across emotion.

2.1.3. Procedure

The still images of the six actors were presented on a non-interlaced gamma-corrected colour monitor (30 cd/m^2 background luminance, 100 Hz frame rate, Mitsubishi Diamond Pro 2070SB) with a resolution of 1024 × 768 pixels. At a viewing distance of 57 cm, the monitor subtended a visual angle of 40 × 30°. During the experiment the participants sat in a chair with their head restrained by a chin-rest, and viewed the display binocularly. To calibrate eye movement signals, a small red fixation point (FP, 0.3° diameter, 15 cd/m^2 luminance) was displayed randomly at one of 9 positions (3 × 3 matrix) across the monitor. The distance between adjacent FP positions was 10°. The participant was instructed to follow the FP and maintain fixation for 1s. After the calibration procedure, the FP was presented at the centre of the screen for 500ms before it was replaced by one of the 180 images (six models × five emotions × three body orientations × two face visibility conditions). Images were presented in random order with restrictions on repetitions of actors for a specific expression. Participants were instructed to press the space bar as soon as they recognised the emotion for the purpose of recording reaction time. Once the space bar was pressed, the image disappeared and the five different emotion categories (i.e. happy, sad, angry, fear and neutral) were presented on the screen. Participant selected one of five keys to indicate their selected emotion label (See Figure 1). This procedure was chosen to reduce the influence of variability in working memory demands on response times. Once a response was selected, the instruction "on a scale of 1-9, how confident are you about your response?" appeared on the screen. Once the participant gave a verbal response, the next instruction on the screen asked for a rating of intensity "on a scale from 1-9, how intense was the emotion?". The experiment was paced by the experimenter and the responses of the participants. Pauses were included after every 30 trials.

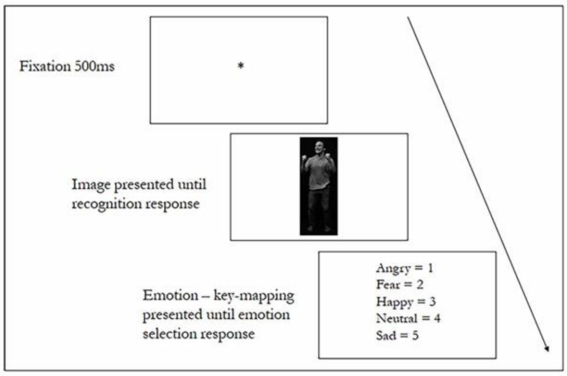

Figure 1: Illustration of trial events in Experiment 1.

During the experiment, horizontal and vertical eye positions from the self-reported dominant eye (determined through the Hole-in-Card test or the Dolman method if necessary) were measured using a Video Eyetracker Toolbox (a camera-based system tracking pupil centre and cornea reflection) with 50 Hz sampling frequency and up to 0.25° accuracy (Cambridge Research Systems, UK). The software developed in MATLAB computed horizontal and vertical eye displacement signals as a function of time to determine eye velocity and position (Guo, Mahmoodi, Robertson, & Young, 2006). The location of gaze points within each video frame was extracted from the raw data. To determine gaze allocation within body regions, each body was divided into five regions of interest (ROI) (Deal et al., 2012): face region, torso (from the base of the neck to the hips, which aligned with the bottom of the t-shirt), arms (from the point where the end of the clavica meets the top of the humerus), hands (from the wrist) and the legs (including feet) (see Figure 2). Viewing time allocated to each ROI was normalised in proportion to the total viewing time sampled in that trial (referred to as proportion viewing time).

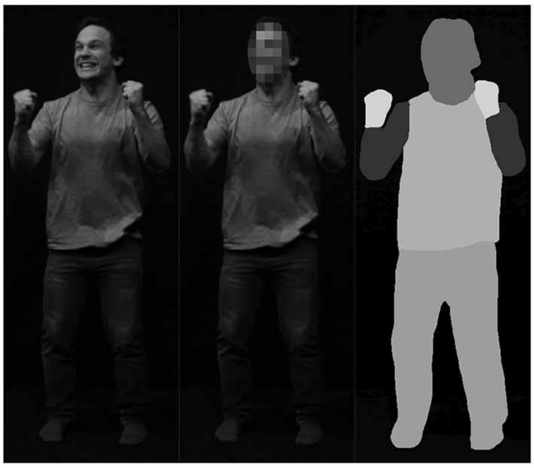

Figure 2: Example of still images and regions of interest (ROI)

All measures were averaged for the three body orientations to ensure that the orientations with optimal visibility of emotion-specific postural cues were included in the average. Viewpoint was not analysed further given the variability in deviation of head orientation from straight ahead in the still images. Behavioural measures (proportion correct responses, confidence ratings and intensity ratings, response times (the response to indicate recognition of the emotion) and total viewing times were analysed using ANOVA with the factors Face visibility (face visible or not visible) and Emotion (happy, angry, sad, fear, neutral). The intensity ratings were not analysed for the trials that participants perceived as neutral. Total viewing times (Experiment 1) refers to the sum of fixation durations on all ROIs (total viewing duration of the person). Proportions of viewing time for each ROI were calculated with reference to the total viewing times. ROI was added for the analysis of proportion viewing times (head, torso, arms, hands, legs). Greenhouse-Geisser was used for sphericity corrections. Generalised eta squared (η^2) was used to indicate effect sizes and Tukey's HSD test was used for pair-wise post-hoc comparisons.

3. Results

3.1 Behavioural measures:

3.1.1. Accuracy (percentage correct responses) was high and varied with Emotion [F(2,58) = 13.4; $p < 0.001$, $\eta^2 = 0.17$], Face visibility [F(1,40) = 25.01; $p < 0.001$, $\eta^2 = 0.02$] and Emotion × Face visibility [F(4,138) = 11.21; $p < 0.001$, $\eta^2 = 0.04$] (See Figure 3). When the face was not visible, accuracy reduced significantly for fearful, happy and sad expressions compared to when faces were visible *(all p's ≤ 0.05).* When the face was visible, accuracy was significantly higher for happy, angry and neutral compared to fearful and sad expressions *(p's ≤ 0.001)*. For images with faces not visible, accuracy was highest for angry expressions (angry>happy/neutral>sad/fear: *p's ≤ 0.05*).

3.1.2. Confidence ratings also varied as a function of Emotion [F(3,116) = 11.2; $p < 0.001$, $\eta^2 = 0.07$], Face visibility [F(1,40) = 33.3; $p < 0.001$, $\eta^2 = 0.04$] and Emotion × Face visibility [F(3,99) = 5.7; $p = 0.002$, $\eta^2 = 0.01$]. Confidence ratings were significantly lower for all expressions when faces were not visible (p's ≤ 0.05). When faces were visible, the highest confidence ratings were associated with happy and neutral expressions (happy > sad/fear/angry: *p's ≤ 0.004*; neutral > sad/fear: *p's < 0.001*). For images with faces not visible, happy, angry and neutral expressions were associated with higher confidence ratings compared to fearful and sad expressions (happy/angry/neutral > sad/fear: *p's ≤ 0.05)*.

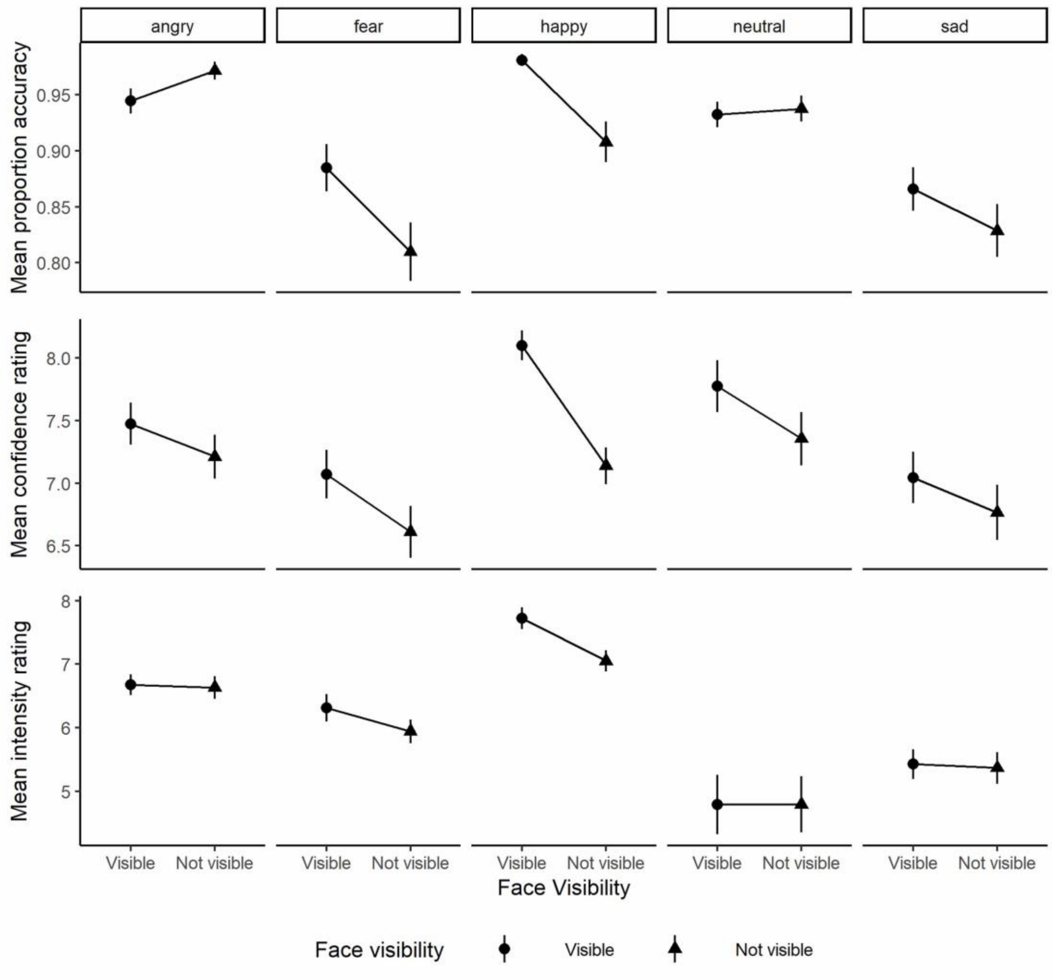

Figure 3: Experiment 1: Behavioural measures as a function of Emotion and Face Visibility.

3.1.3. Intensity ratings varied as a function of Emotion [$F(3,115) = 29.2$; $p < 0.001$, $\eta^2 = 0.22$], Face visibility [$F(1,40) = 21.5$; $p < 0.001$, $\eta^2 = 0.004$] and Emotion × Face visibility [$F(3,95) = 7.3$; $p < 0.0005$, $\eta^2 = 0.006$]. Pixilation of faces reduced intensity ratings significantly for happy and fearful expressions (p's < 0.001). When faces were visible, happy expression attracted the highest intensity rating followed by angry and fearful, and then by sad expressions (happy>angry/fear> sad, $p's \leq 0.05$). When faces were not visible, intensity ratings were highest for angry and happy expressions, followed by fear, and ratings were lowest for sad expressions (happy/angry>fear>sad: $p's \leq 0.05$). Average intensity rating across the six actors varied between 5.8±0.19 and 6.3±0.21 and the range of the ratings (maximum-minimum rating) varied between 4.1 and 5.03 for the different actors.

3.1.4. Response times: Significant effects were found for Emotion [F(4, 143) = 11.8; $p < 0.001$, η^2= 0.13; mean RTs were 1575±0.11ms, 1729±0.11ms, 1970±0.13ms, 1840±0.14ms and 1518±0.09ms for happy, sad, fear, angry and neutral expressions, respectively], Face visibility [F(1,40) = 16.38; $p < 0.001$, η^2=0.03; mean RTs were 1650±0.10ms and 1803±0.12ms when faces were visible and not visible, respectively] and Emotion × Face visibility [F(4,129) = 2.45; p = 0.04, η^2 = 0.01]. Face visibility reduced response times for happy, fear and neutral expressions (*p's* ≤ 0.03). When faces were visible, RT was shortest for neutral and happy expressions (anger/fear/sad>happy/neutral: *p's* ≤ 0.026). When faces were not visible, RT was longest to fearful expressions (fear>angry/happy/sad>neutral; p's ≤ 0.02).

3.2. Viewing times:

3.2.1. Total viewing times: Across all the trials, on average each participant spent at least 93% of image presentation time to explore the displayed actor or actress. Significant effects were found for Emotion [(4,128) = 16.3, $p < 0.001$, η^2 = 0.04], Face visibility [F(1,40) = 13.7, $p < 0.001$, η^2 = 0.006] and Emotion × Face visibility [F(4,140) = 7.3, $p < 0.001$, η^2 = 0.007; Fig. 4]. The effects for viewing time were the same as for RTs: When faces were not visible, total viewing times were longer compared to when they were visible for happy, fearful and neutral expressions (*p's* ≤ 0.01). When faces were visible, angry, fearful and sad expressions attracted longer viewing time than happy and neutral expressions (angry/fear/sad>happy/neutral: *p's* ≤ 0.001). When faces were not visible, fearful expressions attracted the longest viewing time, followed by angry, happy and sad expressions, and then by neutral expressions (fear>angry/happy/sad>neutral: *p's* ≤ 0.04).

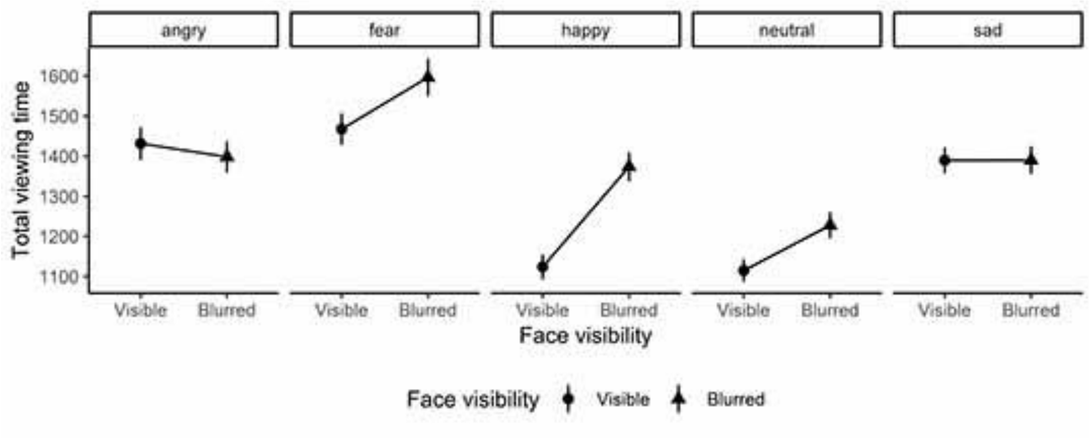

Figure 4: Experiment 1: Total viewing times (in milliseconds) as a function of Emotion and Face visibility.

3.2.2. Proportion viewing times: Viewing time allocated at different local body regions was significantly influenced by Emotion and Face visibility, as reflected in significant effects of Emotion × ROI [F(5,199) = 19.94, $p < 0.001$, $\eta^2 = 0.04$], Face visibility × ROI [F(4,140) = 71.9, $p < 0.001$, $\eta^2 = 0.06$] and Emotion × Face visibility × ROI [F(7,283) = 4.9, $p < 0.001$, $\eta^2 = 0.005$]. Figure 5 shows that viewing patterns were qualitatively similar for all expressions. Proportion viewing times of the head and body were higher compared to viewing of arms, legs and hands *(all p's ≤ 0.017).* The head was also viewed more than the body for all expressions when faces were visible *(p's ≤ 0.002)* but this difference was only significant for happy expressions when faces were not visible *(p = 0.03)*. Emotion-specific effects were revealed in comparisons of proportion viewing times separately for each ROI. The arms were viewed more for happy and fearful expressions (Face visible: happy/fear>angry/sad/neutral: *p's ≤ 0.035*, Face not visible: happy/fear> sad/neutral: *p's ≤ 0.018*), the body was viewed more for sad and fearful expressions (for both face visibility conditions sad/fear>happy/angry: *p's ≤ 0.04*), whereas the head was viewed more for happy, angry and neutral expressions (Face visible: happy/angry/neutral>sad/fear: *p's* ≤ *0.001*; Face not visible : happy>angry: $p = 0.02$ and happy/angry/neutral>sad/fear: *p's* ≤ *0.001*). Hands and legs received little attention, although hands were viewed more in sad and fearful expressions, particularly when faces were not visible (sad/fear>happy/angry/neutral: *p's* ≤ *0.01*).

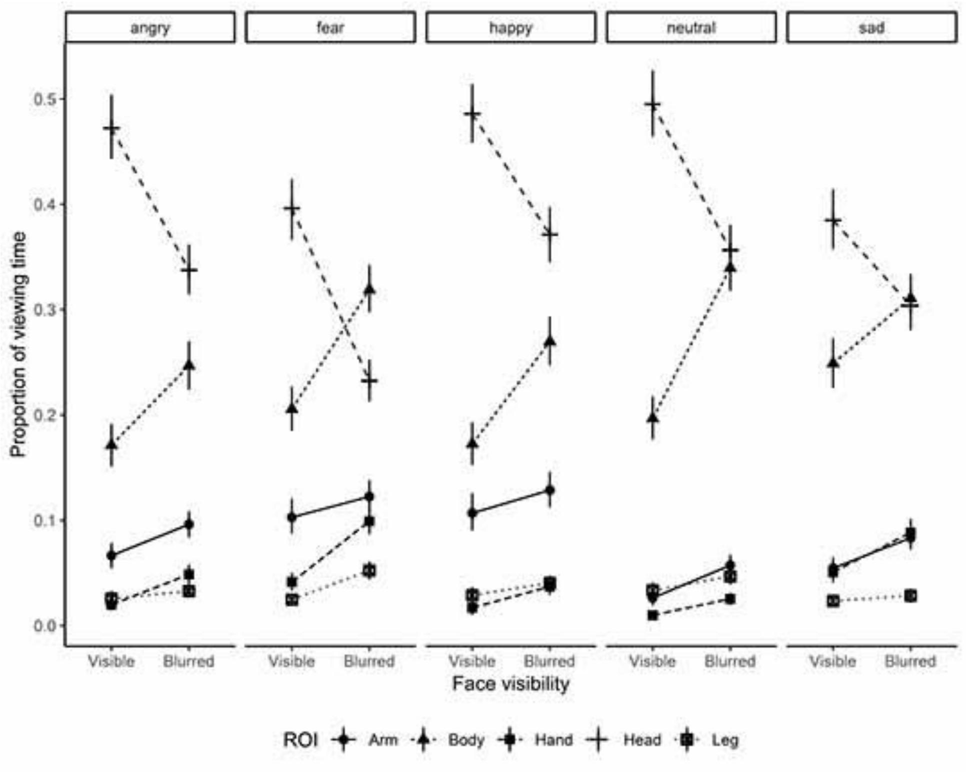

Figure 5: Experiment 1: Proportion viewing times as a function of Emotion, Face Visibility, and Region of Interest (ROI).

4. Summary of results: Analysis of behavioural measures showed that accuracy and confidence ratings were high for all expressions and that these measures reduced significantly for most expressions when faces were not visible. Overall, accuracy and confidence tended to be the highest for happy, angry and neutral expressions. Intensity ratings were the highest for happy, angry and fearful expressions and were overall affected less by face visibility than

accuracy and confidence ratings (only intensity ratings of happy and fearful expressions were reduced). While the gaze distribution over the different body regions was similar for the five expressions, several emotion-specific patterns were observable. For instance, the arms were viewed more for happy and fearful expressions whereas the body was viewed more for sad and fearful expressions. The results of Experiment 1 will be discussed after Experiment 2 where viewing of dynamic body expressions will be investigated.

5. EXPERIMENT 2: Dynamic body expressions

5.1. Methods

5.1.1. Participants

Twenty four participants (5 men (22.1 ± 0.17 years) and 19 women (21.6 ± 1.15 years)) were recruited via the Subject Pool of the School of Psychology at the University of Lincoln. This sample size was determined by checking the effect size reported in previous comparable studies. For instance, Nelson and Mondloch (2017) report an effect size of .4 in proportion viewing time (dynamic displays) for the interaction between Emotion and ROI. With a more conservative effect size of .3, a sample size of 22 would be large enough for this effect to be detected with a power of .95 at alpha level .05. Informed written consent was obtained prior to the testing and all participants had normal or corrected to normal vision.

5.1.2. Procedure

The experimental procedure and the participants' task were similar to those used in Experiment 1. In Experiment 2, videos of 4 actors (two men and two women) were selected from six actors described in Experiment 1. The videos (14 × 15°) were presented on a black window (23 × 17°). The actor was in the centre of the window with 2° distance from the head to the top of the window and from the feet to the bottom of the window in the starting position. Eye-movements were measured using the Video Eyetracker Toolbox (Cambridge Research Systems) via GazeTracker software (http://www.eyetellect.com/gazetracker-software-interface-analysis/). After initial calibration, each trial started with a FP for 2 s, followed by the presentation of a body expression video. Participants were instructed to watch the video until the end before identifying verbally which emotion was expressed. Following their response, participant were required to rate confidence and intensity verbally. All responses were entered by the experimenter. To reduce potential memory effects, videos with faces visible or not visible were presented in separate blocks with a break in between. The order of face visible and face not visible blocks was counterbalanced whereas the order of the videos within each block was pseudorandomised. In total 120 videos were presented (4 actors × 5 emotions × 3 body orientations × 2 face visibility conditions). The looking zones (ROIs) created for Experiment 2 used the same regions as in Experiment 1. Based on the observation that little attention was allocated to the hands in Experiment 1, arms and hands were now combined in one ROI. The size and location of each ROI was adjusted on a frame-to-frame basis using the GazeTracker software. Viewing times were extracted from the number of gaze points at each ROI per video frame and were then normalised with reference to the total tracked viewing time for each video (referred to as proportion viewing times).

6. Results

6.1. Behavioural measures:

6.1.1. Accuracy analysis revealed significant effects of Emotion [F(3,62) = 8.1, $p < 0.001$, $\eta^2 = 0.18$], Face visibility [F(1,22) = 13.6, $p < 0.001$, $\eta^2 = 0.03$] and Emotion × Face visibility [F(2,43) = 14.1, $p < 0.001$, $\eta^2 = 0.15$] (see Fig. 6). Face pixilation only reduced accuracy for sad expressions ($p < 0.001$). When faces were visible, accuracy was the lowest for fear (happy/sad/anger/neutral>fear: *p's* $\leq$ *0.02*), whereas when faces were not visible accuracy was the lowest for sad expressions (angry/happy/fear/neutral>sad: *p's* $\leq$ *0.019*).

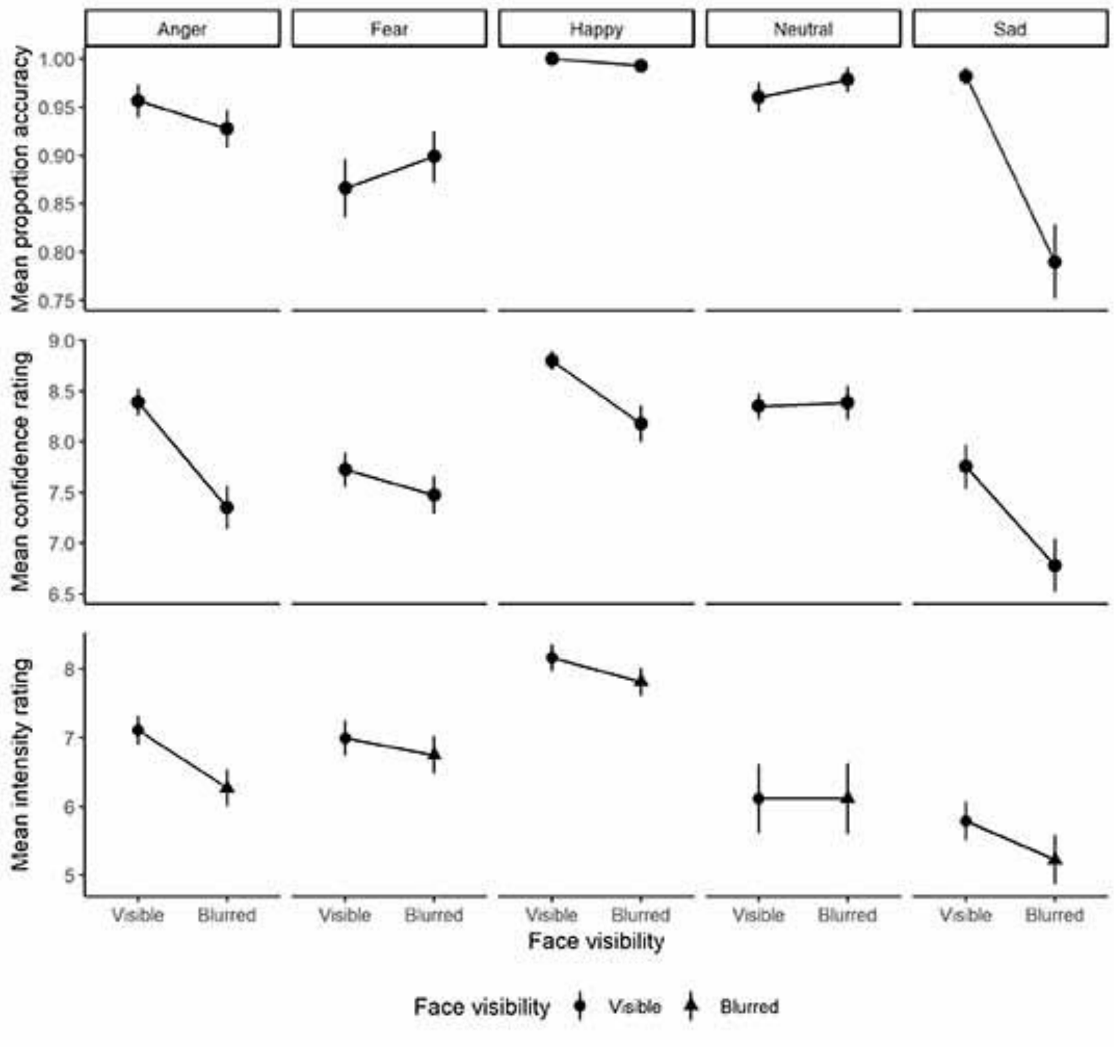

Figure 6: Experiment 2: Behavioural measures as a function of Emotion and Face Visibility.

6.1.2. Confidence ratings: Significant effects were found for Emotion [F(3,48) = 24.11, $p < 0.001$, $\eta^2 = 0.23$], Face visibility [F(1,22) = 29.8; $p < 0.001$, $\eta^2 = 0.1$] and Emotion × Face visibility [F(3,69) = 10.5, $p < 0.001$, $\eta^2 = 0.06$]. Face pixilation reduced confidence ratings for angry, happy and sad expressions *(p's < 0.001)*. Confidence was the highest for happy, angry and neutral expressions and the lowest for sad expressions (Face visible:

happy/angry/neutral>fear/sad: *p's ≤ 0.00*5; Face not visible: happy/neutral>angry/fear>sad: *p's ≤ 0.01*).

6.1.3. Intensity ratings: The analysis revealed significant effects for Emotion [F(2,30) = 19.9, $p < 0.001$, $\eta^2 = 0.23$], Face visibility [F(1,22) = 20.7, $p = 0.002$, $\eta^2 = 0.02$] and Emotion × Face visibility [F(4,80) = 7.9, $p < 0.001$, $\eta^2 = 0.009$]. Face pixilation reduced intensity ratings significantly for angry, happy and sad expressions *(p's < 0.001)*. Happy and sad expressions were associated with the highest and lowest intensity ratings, respectively (Face visible: Happy>fear/angry>sad: *p's ≤ 0.04;* Face not visible: Happy>fear/angry>sad: *p's ≤ 0.03).* Average intensity ratings across the actors varied between 5.9±0.20 and 6.4±0.25 and the range of the ratings per actor (max-min rating) varied between 4.2 and 5.2.

6.1.4. Comparison of behavioural measures of Experiment 1 and 2:

To analyse the effect of body movement (still image *vs* dynamic video display), an additional ANOVA was conducted with the factors of Experiment and Face visibility. In addition to the expected Face visibility effects for all three behavioural measures (F(1,62) > 36.9, *p's < 0.001*, $\eta^2 > 0.08$), the effect of Experiment was significant for accuracy [F(1,62) = 6.18, $p = 0.02$, $\eta^2 = 0.08$] and confidence ratings [F(1,62) = 9.18, $p = 0.004$, $\eta^2 = 0.12$] with higher measures for dynamic video displays. Intensity ratings were unaffected by movement and no significant interaction effects were found.

6.2. Proportion viewing times:

Across all trials, on average each participant spent at least 92% of video presentation time to view the displayed actor or actress. Viewing time allocated at local body regions were influenced by Emotion and Face visibility. Significant effects were found for Emotion × ROI [F(5,110) = 16.5, $p < 0.001$, $\eta^2 = 0.11$], Face visibility × ROI [F(2,29) = 457.7, $p < 0.001$, η^2

= 0.84] and Emotion × Face visibility × ROI [F(6,120) = 8.8, $p < 0.001$, $\eta^2 = 0.06$]. Figure 7 shows that whilst viewing patterns were fairly similar across emotions, they were different when the face was visible or not visible. Viewers looked mostly at the face when it was visible whereas other regions received little attention (head>body/arms/legs: *p's* < *0.001* for all expressions). When the face was not visible, the head was viewed more than the other regions for all expressions except fear (anger/happy/sad: head>body/arms>legs: *p's* < *0.001*; neutral: head>body>arms>legs: *p's* < *0.001)*. For fear, the body was viewed more than the other regions (body>head>arms>legs; *p's* ≤ *0.01)*. Additional emotion-specific effects were revealed in the comparisons between expressions for each ROI. When faces were visible, arms were viewed more for happy expressions (happy>angry/neutral/sad: *p's* ≤ *0.04*) and the head was viewed more for angry expressions (angry>fear/happy/sad: *p's* < *0.001*). When faces were not visible, arms were viewed more for angry expressions (angry>fear/happy/sad/neutral: *p's* ≤ *0.04*; fear/happy/sad>neutral: *p's* < *0.001*), whereas the body was viewed more for fearful expressions (fear> angry/happy/sad/neural: *p's* < *0.001*; sad>angry: $p = 0.03$) and the head was viewed less for fearful expressions (angry/happy/neutral/sad>fear; *p's*< *0.001*).

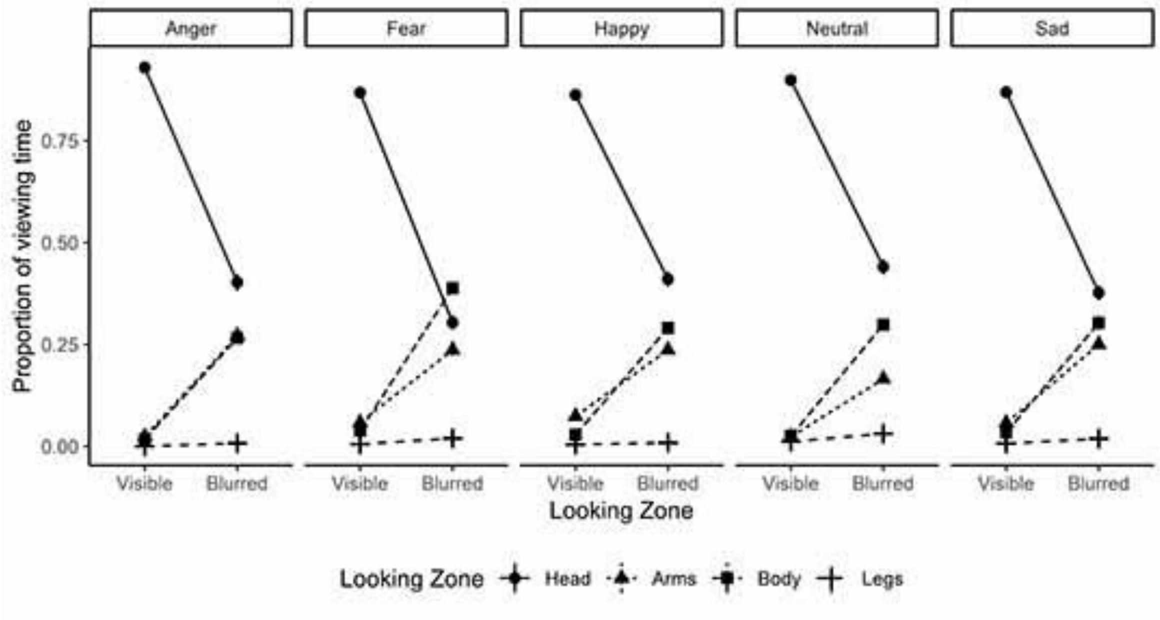

Figure 7: Experiment 2: Proportion viewing times as a function of Emotion, Face Visibility, and Region of Interest (ROI).

7. General discussion

The aim of the present study was to investigate gaze strategies in viewing of static and dynamic body expressions with faces either visible or not visible for expression categorisation. Categorisation accuracy and confidence ratings were higher for dynamic compared to still displays, whereas intensity ratings were not significantly affected by body movement. Face pixilation reduced accuracy more in still displays (affecting all expressions) than in dynamic displays (affecting only sad expressions) suggesting that bodily cues were more informative for emotion recognition in the dynamic videos. The eye movement data revealed that in viewing of still images, observers attended the head, the torso and the arms (whether the face was visible or not visible) and the body was attended more when the face was not visible. A similar viewing pattern was associated with viewing of dynamic body expressions with faces not visible, suggesting that viewers adopted a relatively uniform gaze strategy to optimise information seeking in these three conditions. Conversely, attention was almost exclusively allocated to the face when they were visible in dynamic displays. In addition, small variations in viewing patterns were associated with different expressions in all conditions, suggesting that allocation of attention was influenced by emotion-specific postures and movements in the stimuli.

The strong face bias observed in viewing of dynamic body expressions with visible faces is surprising given that the body is generally viewed more in dynamic compared to static displays in different task demands (Rice et al., 2013, Stoesz et al., 2015). A plausible explanation is that the dynamic facial expressions were more informative for emotion categorisation than the static facial expressions. Consistent with this idea, a few studies have shown that facial expression recognition is enhanced for dynamic compared to static displays (Harwood, Hall, & Shinkfield, 1999), particularly when expressions are subtle (Ambadar, Schooler, & Cohn, 2005). Moreover, in viewing of static images viewers attended faces more when they were visible compared to when they were not, yet in both conditions attention was still allocated to all upper body regions to recruit more diagnostic information. Overall, the pattern in proportionate viewing of face and body in both experiments suggests that viewing for emotion categorisation responds flexibly to the informativeness of facial cues in whole body expression recognition. This finding aligns with those eye-tracking studies investigating person identification, where viewing of the body was found to increase when it was a better indicator for identity than the face (Rice et al., 2013). Interestingly, the head region was still viewed most when the face was not visible in dynamic displays and received as much attention as the body in viewing of still images, suggesting that the position and movement of the head provided useful cues for emotion recognition (Atkinson et al., 2007; Dael et al., 2012; Gunes et al., 2006).

When presented with static images and videos with faces pixilated, observers viewed the head, the torso and the arms whereas the legs were almost entirely ignored in all conditions. This finding is consistent with the observation that leg movements are not often identified as diagnostic for body expressions (Dael et al., 2012) and have previously been shown to receive less attention compared to the upper body when the body is viewed for emotion recognition (Kret, 2013). The upper body bias may have been partly influenced by a general, task independent, centre of gravity effect, previously demonstrated as the allocation of attention to the centre of a scene (Bindeman et al., 2010) or to the centre of faces (Bindeman et al., 2009) in the first fixation. However, consistent with the literature highlighting emotion-specific gestures in different body expressions (Dael et al., 2012; Gunes et al., 2006; Huis in‘t Veld et al., 2014), viewing was also characterised by subtle emotion-specific gaze patterns in the present study, likely resulting from a combination of factors, such as the presence, visibility and the duration of emotion-specific facial or bodily cues. The most noticeable emotion-specific effect was associated with fearful expressions in dynamic displays when faces were not visible, where the body was viewed more than the head whereas the head region was viewed most for all other expressions. Changes in body postures evolved relatively slowly in dynamic displays of fearful expressions, which may have contributed to prolonged viewing of the body in these displays. Moreover, in the comparisons of proportion viewing times between different expressions for each body region separately, the arms were found to be viewed more for happy and fearful expressions in all still displays and for happy and angry expressions dynamic displays when faces were visible and not visible, respectively. The finding that viewers allocated their attention to all upper body regions in static images and in dynamic displays with faces not visible suggests however that viewers did not rely on the presence of just one diagnostic gesture (e.g. pointing for angry expression) and attended all body regions to optimise information seeking. Interestingly, emotion categorisation accuracy was still high when faces were not visible in both experiments, which is consistent with previous studies showing that information from the body can be sufficient for above chance expression categorisation (Atkinson et al., 2004, 2007; de Gelder, 2009). Face visibility did increase behavioural measures, but this effect was not consistent for all expressions and conditions.

One limitation of the present study is that a face-alone (without body) condition was not included in the design, thereby prohibiting a direct comparison of performance based on only body or only facial cues. The inclusion of a face-alone conditions would also be relevant to further investigate the assumption that proportionate viewing of the body is dependent on the informativeness of facial emotion cues in categorisation of whole body expressions. A second limitation is that the frames for static images in Experiment 1 were selected from the videos, based on their proximity to the apex of the body expressions and the clarity of the facial expression. Although these images have high ecological validity and ensure a meaningful direct comparison with video stimuli, creating the still images in this way may have inevitably resulted in some variability in luminance or image quality and in variability in facial expression intensity. The results of the present study show however that average intensity ratings were high for most whole body expressions in both experiments (particularly for happy, angry and fearful expressions), suggesting that the actors expressed high intensity emotional states in the images. In addition, both categorisation accuracy and viewing of facial expressions has recently been shown to be quite resilient to reduced image quality (Guo, Soornack, & Settle, 2018). Guo et al. (2018) showed for example that accuracy in expression categorisation and viewing measures were unaffected when pixel resolution was reduced to as low as 48 × 64 pixels. Moreover, previous eye-tracking results have shown that gaze allocation in viewing of faces and social scenes is more guided by top-down (task-driven) information than low-level features and visual saliency (Birmingham, Bischof, & Kingstone, 2009; Malcolm, Lanyon, Fugard, & Barton, 2008). It is therefore unlikely that the quality of the images could have been responsible for differences in behavioural measures and viewing patterns between still and dynamic displays in the present study. Nonetheless, further studies will be necessary to specify the relationship between the intensity of facial expressions, image quality and the amount of viewing that the body receives for whole body expression categorisation.

A second limitation could be raised based on the effect sizes for statistical effects in the analysis of viewing measures for static and dynamic displays. The results show however that smaller effects sizes were predominantly associated with Experiment 1 whereas effect sizes were moderate or large in Experiment 2, suggesting increased consistency in viewing of dynamic stimuli. This finding is in line with previous studies showing that movement in a scene can reduce variability in eye-movements when observers are engaged in free viewing of dynamic compared with static displays (Carmi & Itti, 2006; Dorr, Martinetz, Gegenfurtner & Barth, 2010). In the present study facial and body movements constitute important diagnostic emotion cues for the task at hand and attention is therefore strongly guided by top-down processes (Birmingham, Bischof, & Kingstone, 2009; Malcolm, Lanyon, Fugard, & Barton, 2008; Aviezer et al, 2012; Kret et al, 2013). Future studies are however needed to systematically investigate the influence of low level saliency effects associated with movement in viewing of the dynamic body expressions.

Comparison of behavioural measures for static and dynamic displays showed that accuracy and confidence ratings were higher for dynamic compared to static displays whether faces were visible or not. Whilst this result should be considered with some caution given that a smaller number of stimuli were used in Experiment 2, the trend in this analysis suggests that the addition of movement in face and body improved whole body emotion categorisation. An outstanding question is which information in dynamic displays is responsible for this behavioural improvement. Previous studies have shown that inversion of dynamic facial expressions (Calder, Young, Keane, & Dean, 2000; Calder & Jansen, 2005; Ambadar et al., 2005,) and dynamic body expressions (Atkinson et al., 2007) reduces emotion categorisation judgments, suggesting that motion benefits processing of configural information for both. Ambadar et al. (2005) showed however that the effect of inversion was the same for static and dynamic facial expressions, suggesting that the motion benefit could not be entirely attributed to additional configural information. Instead, they argued that the change in the composition of features from neutral to an emotional state provided the most crucial additional information in dynamic displays. Consistent with this idea, Ambadar et al. (2005) found a comparable benefit of videos and the sequential presentation of the first and last frame of the video compared to static expressions, suggesting that the temporal sequential presentation of the changes was less relevant than the perceived change from neutral itself. For body expressions, Atkinson et al. (2007) showed that performance was still above chance for videos in the inverse condition and argued that emotion-specific gestures may have been sufficiently recognisable for above chance expression recognition. Interestingly, the videos in the present study (and in Atkinson et al., 2007) showed the actors in the neutral position at the start of the video. As such, it could be that the perceived change from neutral may have provided crucial information in face covered dynamic body expressions in the present experiment. Future studies could further specify the relative importance of additional configural information and the perceived change from neutral in dynamic displays of whole body expressions using similar methodologies.

Conclusion: The aim of the present study was to investigate gaze strategies in viewing of dynamic and static whole body expressions of emotions. The findings revealed a stronger face-bias in dynamic compared to static displays when faces were visible. When faces were not visible or when observers viewed static images, viewing was distributed over the head, torso and the arms. Viewing was also subtly influenced by emotion-specific gestures. Together, these findings suggest that viewers adopt a uniform gaze strategy for optimal recruitment of diagnostic information in viewing of whole body expressions for emotion categorisation.

8. Acknowledgements:

We would like to express our thanks to Ferenc Igali and Georgia Sapsford for their contribution to the recording of the videos and the processing of the stimuli. We would also like to thank Emma Grange and Grace Severn for their contribution to data collection.

9. References

Adolphs, R. (2002). Neural systems for recognizing emotions. Current opinion in neurobiology, 12, 169-177.

Ambadar, Z., Schooler, J. W., & Cohn, J. (2004). Deciphering the enigmatic face. The importance of facial dynamics in interpreting subtle facial expression. Psychological Science, 16, 403-410.

Atkinson A. P., Dittrich, W., Gemmell, A. J., & Young, A. W. (2004). Emotion perception from dynamic and static body expressions in point-light and full-light displays. Perception 33, 717-746.

Atkinson, A. P., Tunstall, M. L., & Dittrich, W. H. (2007). Evidence for distinct contributions of form and motion information to the recognition of emotions from body gestures. Cognition 104, 59-72.

Aviezer, J., Trope, Y., & Todorov, A. (2012). Bodily cues, not facial expressions, discriminate between intense positive and negative emotions. Science, 338, 1225-1229.

Bindemann, M., Scheepers, C., & Burton, A. M. (2009). Viewpoint and center of gravity affect eye movements to human faces. Journal of Vision, 9, 7.

Bindemann, M., Scheepers, C., Ferguson, H. J., & Burton, A. M. (2010). Face, body, and center of gravity mediate person detection in natural scenes. Journal of Experimental Psychology: Human Perception and Performance, 36, 1477–1485.

Birmingham, E., Bischof, W. F., & Kingstone, A. (2009). Saliency does not account for fixations to eyes within social scenes. Vision Research, 49, 2992-3000.

Calder, A. J. & Jansen, J. (2005). Configural coding of facial expressions: The impact of inversion and photographic negative. Visual Cognition, 12, 495–518.

Calder, A. J., Young, A. W., Keane, J., & Dean, M. (2000). Configural information in facial expression perception. Journal of Experimental Psychology: Human Perception and Performance, 26, 527–551.

Calvo, M. G. & Nummenmaa, L. (2008). Detection of emotional faces: Salient physical features guide effective visual search. Journal of Experimental Psychology: General, 137, 471-494.

Carmi, R. & Itti, L. (2006). Visual causes versus correlates of attentional selection in dynamic scenes. Vision Research, 46, 4333-4354.

Carton, J. S., Kessler, E. A., & Pape, C. L. (1999). Nonverbal decoding skills and relationship well-being in adults. Journal of Nonverbal Behaviour, 23, 91–100.

Conty, L. Dezechache, G., Hugeueville, L, & Grèzes, J. (2012). Early binding of gaze, gesture and emotion: neural time course and correlates. The Journal of Neuroscience, 32, 4531-4539.

Dael, N., Mortillaro, M., & Scherer, K. R. (2012). The body action and posture coding system (BAP). Development and reliability. Journal of Nonverbal Behaviour, 36, 97-121.

Dael, N., Mortillaro, M., & Scherer, K. R. (2012). Emotion expression in body action and posture. Emotion, 12, 1085–1101.

de Gelder, B. (2009). Why bodies? Twelve reasons for including bodily expressions in affective neuroscience. Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences, 364, 3475-3484..

de Gelder, B. (2013). From body perception to action preparation: a distributed neural system for viewing bodily expressions of emotion. In M. Shiffrar and K. Johnson (Eds.). People Watching: Social, Perceptual, and Neurophysiological Studies of Body Perception (pp. 350-368). New York, NY: Oxford University Press.

de Gelder, B., de Borst, A. W., & Watson, R. (2015). The perception of emotion in body expressions. WIREs Cognitive Science, 6, 149-158.

Eisenbarth, H. & Alpers, G. W. (2011). Happy mouth and sad eyes: scanning emotional facial expressions. Emotion, 11, 860–865.

Dorr, M., Martinetz, T., Gegenfurtner, K.R., & Barth, E. (2010). Variability of eye movements when viewing dynamic natural scenes. Journal of Vision, 10, 1-17.

Ekman, P. (1992). An argument for basic emotions. Cognition and Emotions, 6, 169-200.

Engelen, T., Zhan, M., Sack, A. T., & de Gelder, B. (2018). Dynamic interactions between emotion perception, action perception and preparation areas for reacting to social threat: A combined cTBS-fMRI study. eNeuro, ENEUR).0408-17.2018.

Faul, F., Erdfelder, E., Lang, A., & Buchner, A. (2009). G power 3.1. Dusseldorf, Germany: Institute for Experimental Psychology, Herinrich Heine University of Dusseldorf. Retrieved from http://www.psycho.uni-duesseldorf.de/abteilungen/aap/gpower3/

Grezes, J., Pichon, S., & de Gelder, B. (2007). Perceiving fear in dynamic body expressions. NeuroImage, 35, 959-967.

Gunes H., & Piccardi, M. (2006). A bimodal face and body gesture database for automatic analysis of human nonverbal affective behaviour. 18th International Conference on Pattern Recognition, Hong Kong, 1148-1153.

Gunes, H., Shan, C., Chen, S., & Tian, Y. (2015). Bodily expression for automatic affect recognition. In A Konar, A Chakraborty (Eds.): Emotion recognition: A pattern analysis approach (pp. 343-377). John Wiley and Sons.

Guo, K. (2012). Holistic gaze strategy to categorize facial expression of varying intensities. PLOS One, 7, e42585.

Guo, K., Mahmoodi, S., Robertson, R., &Young, M. (2006). Longer fixation duration while viewing face images. Experimental Brain Research, 171, 91-98.

Guo, K. & Shaw, H. (2015). Face in profile view reduces perceived facial expression intensity: an eye-tracker study. Acta Psychologica, 155, 19-28.

Guo K, Soornack, Y., & Settle. R. (2018). Expression-dependent susceptibility to face distortions in processing of facial expressions of emotion. Vision Research (in press).

Harwood, N. K., Hall, L. J., & Shinkfield, A. J. (1999). Recognition of facial emotional expressions from moving and static displays by individuals with mental retardation. American Journal on Mental Retardation, 104, 270–278.

Huis In ‘t Veld, E. M. J., van Boxtel, G. J. M., & de Gelder, B. (2014). The Body Action Coding System II: Muscle activations during the perception and expressions of emotion. Frontiers in behaviour Neuroscience, 8, 1-13

Huis in ‘t Veld, E. M. J., Van Boxtel, G. J. M., & de Gelder, B. (2014). The body action coding system I: muscle activations during the perception and expression of emotion. Social Neuroscience. 9, 249–264.

Izard, C. E., Fine, S., Schultz, D., Mostow, A., Ackerman, B., et al. (2001). Emotion knowledge as a predictor of social behavior and academic competence in children at risk. Psychological Science, 12, 18–24.

Jensen S. & Kotz, S. A. (2011). The temporal dynamics of processing emotions from vocal, facial and bodily expressions. NeuroImage, 58, 665-674.

Kokinous J. Tavano, A, Kotz, S. A., & Schröger, E. (2017). Perceptual integration of faces and voices depends on the interaction of emotional content and spatial frequency. Biological Psychology, 123, 155-165.

Kreifelts, B., Ethofer, T., Shiozawa, T., Grodd, W., & Wildgruber, D. (2009). Cerebral representation of non-verbal perception: fMRI reveals audiovisual integration area between voice- and face-sensitive regions in the superior temporal sulcus. Neuropsychologia, 47, 3059-3066.

Kret, M. E., Stekelenburg, J., Roelofs, K., & de Gelder, B. (2013). Perception of face and body expressions using electromyography, pupillometry and gaze measures. Frontiers in Psychology, 4-12.

Leppänen, J. M., Tenhunen, M., & Hietanen, J. K. (2003). Faster choice-reaction times to positive than to negative facial expressions: The role of cognitive and motor processes. Journal of Psychophysiology, 17, 113-123.

Malcolm, G. L., Lanyon, L. J., Fugard, A. J. B., & Barton, J. J. S. (2008). Scan patterns during the processing of facial expression versus identity: An exploration of task-driven and stimulus-driven effects. Journal of Vision, 8, 1-9.

Meeren, H. K. M, Hadjikhani, N., Ahlfors, S. P., Hämäläinen, M. S., & de Gelder, B. (2008). Early category-specific cortical activation revealed by visual stimulus inversion. PLoS ONE, 3:e3503.

Meeren, H. K., Sinke, C. B., Kret, M. E., & Tamietto, M. (2010). Standing up for the body. Recent progress in uncovering the networks involved in the perception of bodies and bodily expressions. Neuroscience Biobehavioral Reviews, 34, 513–527.

Müller, V. I., Habel, U., Derntl, B., Schneider, F., Zilles, K., Turetsky, B. I., & Eickhoff, S. B. (2011). Incongruence effects in cross-modal emotional integration. NeuroImage, 54, 2257-2266.

Nelson, N. L. & Mondloch, C. J. (2017). Adults' and children's perception of facial expressions is influenced by body postures even for dynamic stimuli. Visual Cognition, 25, 563-574.

Nowicki, S. Jr. & Duke, M. P. (1994). Individual differences in the nonverbal communication of affect: The diagnostic analysis of nonverbal accuracy scale. Journal of Nonverbal Behaviour, 19, 9–35.

O'Toole, A. J., Phillips, P. J., Weimer, S., Roark, D.A., Ayyad, J., Barwick, R., & Dunlop, J. (2011). Recognizing people from dynamic and static faces and bodies: Dissecting identity with a fusion approach. Vision Research, 51, 74-83.

Pollux P. M. J, Hall, S., & Guo, K. (2014). Facial expression training optimises viewing strategy in children and adults. PLoS ONE 9(8): e105418.

Rice, A., Phillips, P. J., Natu, V., An, X., & O'Toole, A. J. (2013). Unaware body recognition from the body when face identification fails. Psychological Science, 24, 2235 – 2243.

Sato, W., Yoshikawa, S., Kochiyama, T., & Matsumura, M. (2004). The amygdala processes the emotional significance of facial expressions: an fMRI investigation using the interaction between expression and face direction. NeuroImage, 22, 1006-1013.

Smith, M. L., Cottrell, G. W., Gosselin, F., & Schyns, P. G. (2005). Transmitting and decoding facial expressions. Psychological Science, 16, 184–189.

Stekelenburg J. & Vroomen, J. (2007). Neural correlates of multisensory integration of ecologically valid audiovisual event. Journal of Cognitive Neuroscience, 19, 1964-1973.

Stekelenburg, J. & de Gelder, B. (2004). The neural correlates of perceiving human bodies: and ERP study on the body-inversion effect. NeuroReport, 15, 777-780.

Stoesz, B, M. & Jakobson, L. S (2014). Developmental changes in attention to faces and bodies static and dynamic scenes. Frontiers in psychology, 5, article 193.

Yeh, P, Geangu, E., & Reid, V. (2016). Coherent emotional perception from body expressions and the voice. Neuropsychologia, 91, 99-108.