Running Head: DYNAMIC EMOTION PERCEPTION

Title: Dynamic emotion perception and prior expectancy

Authors: Ilvana Dzafic[a,b], Andrew K. Martin[ac], Julia Hocking[d], Bryan Mowry[a,e] &

Hana Burianová[b]

[a]Queensland Brain Institute, University of Queensland, Brisbane, Australia

[b]Centre for Advanced Imaging, University of Queensland, Brisbane, Australia

[c]University of Queensland Centre for Clinical Research, University of Queensland, Brisbane, Australia

[d]School of Psychology and Counselling, Queensland University of Technology, Brisbane, Australia

[e]Queensland Centre for Mental Health Research, Brisbane, Australia

Correspondence should be addressed to:
Ilvana Dzafic
Queensland Brain Institute,
University of Queensland,
Brisbane, QLD, 4072, AUSTRALIA
E-mail: i.dzafic@uq.edu.au
Telephone: +61 7 3346 6300
Fax: +61 7 3346 6301

Conflict of Interest: The authors declare no competing financial interests.

Abstract

Social interactions require the ability to rapidly perceive emotion from various incoming dynamic, multisensory cues. Prior expectations reduce incoming emotional information and direct attention to cues that are aligned with what is expected. Studies to date have investigated the prior expectancy effect using static emotional images, despite the fact that dynamic stimuli would represent greater ecological validity. The objective of the study was to create a novel functional magnetic resonance imaging (fMRI) paradigm to examine the influence of prior expectations on naturalistic emotion perception. For this purpose, we developed a dynamic emotion perception task, which consisted of audio-visual videos that carry emotional information congruent or incongruent with prior expectations. The results show that emotional congruency was associated with activity in prefrontal regions, amygdala, and putamen, whereas emotional incongruency was associated with activity in temporoparietal junction and mid-cingulate gyrus. Supported by the behavioural results, our findings suggest that prior expectations are reinforced after repeated experience and learning, whereas unexpected emotions may rely on fast change detection processes. The results from the current study are compatible with the notion that the ability to automatically detect unexpected changes in complex dynamic environments allows for adaptive behaviours in potentially advantageous or threatening situations.

1. Introduction

The cognitive processes involved in emotion perception are of significant interest as they provide insights into how humans make sense of complex social interactions. Social situations can rapidly change, requiring a fluid and adaptable system able to recognize and predict another's emotions (Palermo and Rhodes, 2007). Neuroimaging studies have found that recognition of emotion is influenced by 'prior expectations' about which emotions are likely to arise in certain contexts (Barbalat *et al.*, 2013; Dieguez-Risco *et al.*, 2015); however, the behavioural and neural influences of prior expectations on perception have been examined using static representations of emotion. Static images only represent a single facial pattern within a single modality and fail to capture the dynamic complexity of emotion perception relevant in everyday life. In addition, it is the dynamics of emotion that preferentially incite predictive mechanisms, leading to a greater reliance on prior expectations (Palumbo and Jellema, 2013; Kaufman and Johnston, 2014). Thus, it is of paramount importance that functional Magnetic Resonance Imaging (fMRI) studies, which investigate the influence of prior expectations on emotion perception, incorporate tasks that rely on multimodal sensory information and dynamic emotional presentations. Therefore, the first aim of the study was to develop and validate a novel task that is suitable for use in neuroimaging experiments and that incorporates the abovementioned elements of emotion perception.

To study the underlying processes of emotion perception it is essential that stimuli represent natural emotional expressions, composed of facial patterns in constant motion combined with speech prosody. Static emotional images are inadequate, non-canonical stimuli, i.e., non-moving and devoid of auditory information (Kilts *et al.*, 2003). However, audio-visual displays of dynamic emotion

present more resemblance to real-life emotional expressions and elicit enhanced psychological responses (Yoshikawa and Sato, 2006; Collignon *et al.*, 2008; Lambrecht *et al.*, 2014) and neural activity in the emotion brain network (Robins *et al.*, 2009; Arsalidou *et al.*, 2011), and, as such, provide greater ecological validity. Direct comparisons of dynamic and static emotion perception provide evidence for significant differences in cognitive processes. For example, Kilts *et al.* (2003) found that dynamic emotion perception recruited neural correlates implicated in the evaluation of emotional messages in the constant changing of facial patterns. More importantly, a number of studies have demonstrated that, compared to static emotional expressions, the changing dynamics in naturalistic emotional expressions stimulate greater predictive processes (Palumbo and Jellema, 2013; Kaufman and Johnston, 2014). An explanation for this phenomenon is that dynamic emotions evolve over time and involve greater input of sensory information. Due to the greater sensory input, cognitive efficiency relies on predictive mechanisms from prior expectations, which reduce the incoming sensory information by directing attention to cues that are aligned with expectations (Friston *et al.*, 2006). As prior expectations are a vital process in dynamic emotion perception, the second aim of the study was to examine the behavioural and neural correlates of the effect that prior expectancy has on perception of dynamic emotion.

Previous emotion perception studies investigating the effect of prior expectancy have induced expectations with a preceding emotion instruction cue or a sentence describing an emotionally relevant context. For example, Barbalat *et al.* (2013) manipulated prior expectations by instructing participants to identify a specific emotion in a sequence of static emotional faces, and decide whether the emotional faces were congruent or incongruent with the emotion in the instruction. It is argued

that the instruction cue created an internal "template" against which to match the following stimuli (Summerfield and Koechlin, 2008), and in this way directed attention and facilitated response to emotional faces, which were congruent with the instruction cue. In Dieguez-Risco *et al.*'s (2015) study, on the other hand, prior expectations were induced by sentences describing different emotionally-relevant contexts, prior to the presentation of static emotional faces. Participants were instructed to indicate whether the expression shown in the face matched the context of the situation described in the preceding sentence. Interestingly, in this task, congruency with prior expectations did not facilitate responses to angry faces, whilst responses to angry faces were faster on incongruent trials. The authors explained that slower response times during the congruent angry conditions occurred due to greater uncertainty in evaluating whether a specific negative emotion matches a specific negative context, as a variety of different negative emotions may be related to different negative contexts. However, incongruent trials would be comparatively easier, because when a negative face is presented after a positive context there is a clear difference in valence. Notwithstanding the valuable findings reported by Barbalat *et al.* (2013) and Dieguez-Risco *et al*. (2015), only static displays of emotion were explored, despite the fact that facial motion preferentially invokes predictive mechanisms leading to prior expectations (Kaufman and Johnston, 2014). As such, utilization of static stimuli in experimental paradigms limits the scope of these results and, more importantly, may confound the predictive processes that are naturally involved in dynamic emotion perception.

Studies using *dynamic* social stimuli have induced expectations by increasing the occurrence probability of a particular stimulus. For example, Chambon *et al.* (2011a) examined the influence that prior expectations have on the ability to

appreciate other people's intentions. Prior expectations were induced by selectively manipulating the likelihood of a particular intention to occur within a sequence of different types of intentions. The intentions were communicated with an actor's hand, which was presented in a video format. The authors found that when the intentions were social and more complex, prior expectations were more likely to bias the inference of intention type. In a different study investigating change detection in dynamic presentations of emotional expressions, the likelihood of a particular emotional expression to occur was manipulated (Kreegipuu *et al.*, 2013). This emotional oddball paradigm involved a sequence of standard expressions (such as neutral), infrequently interspersed with a deviant emotional expression (such as angry). The perception of the incongruent emotional expression was found to rely on 'visual automatic change detection' processes, which automatically reorientate attention to changes in the environment (Clery *et al.*, 2013). Unfortunately, there are notable differences between oddball paradigms and tasks that involve a dynamic flow of emotion displayed in naturalistic emotion expression, and, as such, ecological validity of oddball paradigms is limited, as it fails to reflect natural change detection in emotion perception.

The aims of the current study were thus twofold: (1) to design and validate a novel dynamic, audio-visual emotion perception task, suitable for the fMRI environment; and (2) to investigate the differences in brain activity underlying perception of dynamic emotions, which are congruent as opposed to incongruent with prior expectations. In order to induce prior expectations we employed emotional instruction cues (Barbalat *et al.*, 2013) and we increased the likelihood for a congruent emotion to occur (Chambon *et al.*, 2011a). Furthermore, as the consensus is that emotion perception requires a large-scale neural network with distributed activity

(Vuilleumier and Pourtois, 2007), we used a multivariate, Partial Least Squares (PLS; McIntosh and Lobaugh, 2004) analytic approach to investigate the coordinated activity of brain regions. We expected to find activations in brain regions consistent with previous prior expectancy literature, such as the ventromedial prefrontal cortex (vmPFC), during conditions congruent with prior expectations (Summerfield and Koechlin, 2008; Barbalat *et al.*, 2013), and brain regions associated with prediction error and visual automatic change detection (Clery *et al.*, 2013) during incongruent conditions. In addition, we hypothesized that due to the greater complexity of naturalistic emotion, key regions of the 'social brain' would be involved during emotion perception (Van Overwalle, 2009; Kennedy and Adolphs, 2012). Specifically, we predicted that: (1) conditions congruent with prior expectations would engage the amygdala, an area critical in emotion learning (Hamann, 2001; Hooker *et al.*, 2006), as prior expectations are reinforced by learning from past experience (Dolan, 2007; Chambon *et al.*, 2011a); and (2) conditions incongruent with prior expectations would engage the right temporoparietal junction (rTPJ), an area known to be important in efficient detection and reorienting towards unexpected change in a social environment (Decety and Lamm, 2007; Geng and Mangun, 2011).

2. Methods

2.1 Participants

Twenty-eight healthy, right-handed males (age range = 23-46; mean age = 31.79, SD = 4.95) were recruited through on-line advertising to staff and students across the University of Queensland. We recruited only males in order to reduce heterogeneity; it has been found that males and females may differ in their emotion perception (Stevens and Hamann, 2012; Lambrecht *et al.*, 2014). Screening was

conducted over the phone prior to the recruitment, to confirm that participants had no eye disease, were not currently taking medication, and were without a history of neurological disorders or metal implants in their body. The Mini International Neuropsychiatric Interview version 5.0.0 (Sheehan *et al.*, 1997), was used to ensure that participants did not have current alcohol dependence and were not experiencing a major depressive episode. Intelligence quotient (IQ) was estimated using 2 subsets (vocabulary and matrix reasoning) of the Wechsler abbreviated scale of intelligence (WASI; Wechsler, 1999). On the WASI, the mean standard score was 107.33 (SD = 11.79), indicating that the sample displayed, on average, normal IQs.

Participants were provided with an information sheet, which included a full description of the study and Magnetic Resonance Imaging (MRI) procedure. After reading the document, written informed consent was obtained. This research was approved by the Medical Research Ethics Committee of the University of Queensland. Participants received $30 as reimbursement.

2.2 Materials

The novel "Dynamic emotion perception" (DEP) task used in this study involved viewing of audio-visual video clips that carry emotional information congruent or incongruent with prior expectations. Prior expectations were induced by displaying an emotional instruction cue (Barbalat *et al.*, 2013) before the video clips, and by increasing the occurrence likelihood of emotions congruent with the emotion in the instruction (Chambon *et al.*, 2011b), for example by increasing the occurrence of 'anger' within a block where an angry instruction cue was displayed.

2.2.1 Video clips

The emotional video clips were recorded by a Canon EOS 70D video camera and edited (size, audio, and luminance correction) in iMovie on a MacBook Air

laptop. The video clips presented a 38-year old Caucasian female actor, wearing a black shirt. The recording captured the head, neck, and shoulders of the actor, against a beige background. The actor had consistent direct gaze towards the camera and was seated in the same position throughout filming.

The actor was instructed to speak a list of 24 sentences whilst acting three different emotions; each sentence was first spoken in a happy expression, next a neutral expression and last an angry expression. The actor repeated each sentence five times in each emotion (i.e., a sentence was spoken in a happy expression five times, next in a neutral expression five times and lastly in an angry expression five times). Before the sentences were spoken, the actor was asked to think of a situation where she would speak the sentence in that certain emotion; she was instructed to convey the emotion through tone of voice and facial expressions. The sentences spoken by the actor were emotionally ambiguous (i.e., the semantic content made sense for multiple emotions), and had consistent level of difficulty, as quantified by word length, frequency, reaction time, and accuracy (The English Lexicon Project; Balota et al., 2007). The context of the sentences related to everyday familiar activities, such as dining (e.g. "You are eating the steak"), transport (e.g. "We'll catch the ferry"), cleaning (e.g. "You are cleaning the kitchen"), outdoors (e.g. "We'll sit under that tree"), mass media (e.g. "We will watch this TV show") and school (e.g. "Let's start our homework").

A total of 360 video clips were recorded. We attempted to keep the video clip recordings as close to 3 seconds each. All video clips obtained were later edited to 3 seconds; any video clips with a duration greatly deviating from 3 seconds were discarded. The video clips were assessed by the researchers for quality, and video clips with visual or sound issues were discarded, this resulted in a total of 98 edited

video clips. We conducted a pilot study with thirteen healthy participants (10 males, 3 females), in order to rate the 98 edited video clips. Participants viewed the video clips and labelled the emotion using a Likert scale: 5 = angry, 4 = slightly angry, 3 = neutral, 2 = slightly happy, and 1 = happy. The final 48 video clip stimuli obtained the highest emotional ratings, for happy and angry these were the videos that were rated as close to 1 and 5, respectively; for neutral videos this was closest to the average rating. There were 16 examples of happy, angry and neutral each, all three emotions were expressed with the same 16 sentences spoken.

The baseline video clips were created from three of the final video clips described above. The baseline video clips had scrambled visual and auditory features to control for low-level perception and motor movement. We scrambled the video clips using Matlab 7.14 (The Mathworks Inc.,) by first extracting frames from the original video clip, next converting the frames from RGB images into indexed images, then scrambling the pixels in each frame and lastly combining the scrambled frames into a video clip. For audio scrambling, waveforms were extracted from the original video clip and scrambled in Matlab, and then re-inserted into the scrambled video clip.

2.2.2 Cues

The emotional cues were created in Adobe Photoshop CS5.1; the cues presented a still picture of the actor expressing an emotion, with the expressed emotion written in white text underneath the picture; the picture and the writing were presented in the centre of the cue and overlaid on a black background. The still pictures of the actor were screenshots from discarded video clips (discarded due to sound issues), there were 3 still pictures of the actor, and in each the actor expressed either an angry, happy or neutral emotion. The baseline cue was the same design as

the emotional cue, with a picture and writing in the centre of the cue, overlaid on a black background. However, the baseline cue contained a scrambled still picture of the actor; we scrambled the picture using Matlab by converting the RGB image into an indexed image, and then scrambling the pixels in the image. The writing underneath the baseline picture was "baseline". At the start of each block, the emotional or baseline cue contained white text above the picture, instructing to either make an "index finger press" or "any finger press", respectively.

The emotion cue displayed the emotion both visually (static facial expression of the actor), as well as verbally (written emotion word). The addition of visual features was done to aid memory for the emotion, as there is evidence that males, compared to females, have a verbal disadvantage, which extends into decreased memory for verbal content (Andreano and Cahill, 2009). The emotion cue containing the instruction, presented at the beginning of each emotion block, was used to induce the prior expectation (as per Barbalat *et al.* (2012)), thereafter the emotion cue (without the instruction) was displayed before each video clip, to reinforce the instruction and to aid memory and expectation for the particular emotion.

2.3 Design

The experimental procedure consisted of three runs of the DEP task and its nine experimental conditions; 3 emotional cues (happy, angry or neutral) x 3 emotional videos (happy, angry or neutral). When the emotions in the instruction cue and video matched, this was a 'congruent' condition, whereas when the emotions in the instruction cue and video did not match, this was an 'incongruent' condition. Please note that we included neutral video clips to increase the difficulty of the task by increasing the variety of video stimuli. In the Emotion task by Barbalat *et al.*

(2013), the authors also included 3 emotional stimuli, reasoning that a choice of 3 rather than 2 alternatives would produce an additional cognitive constraint, which would increase the congruency effect. Neutral conditions were removed from the final analysis, as neutral emotion is ambiguous and contentious in the emotion literature (Kesler et al., 2001; Cooney et al., 2006; Sabatinelli et al., 2011). Likewise, we did not use the neutral video clips as the baseline task, as neutral faces may elicit activation in social-emotion regions (Blasi *et al.*, 2009) and are sometimes evaluated as emotional (Lee *et al.*, 2008).

Within each run there were nine experimental blocks and three baseline blocks. Each experimental block began with an instruction emotional cue (3 sec), followed by six or nine sequences of trials consisting of an emotional cue (1 sec), a black screen as the inter-stimulus interval (ISI; mean duration of 1sec) and an emotional video clip (3 sec) (see **Figure 1**). The ISI was jittered within a block, with a uniform distribution between 500ms and 1500ms, of either 6 x 200ms intervals (during blocks of 6 video clips) or 9 x 125ms intervals (during blocks of 9 video clips). The experimental blocks contained either six or nine video clips; this reduced the predictability of how many video clips each block contained, and thus the ability of participants to predict the number of incongruent or congruent video clips. Alternating the number of video clips was also done to eliminate repetitiveness, as this may result in fluctuations of attention. Within a block of six video clips, there were four congruent and two incongruent video clips (66.67% expectancy bias); and within a block of nine video clips, there were five congruent and four incongruent video clips (55.55% expectancy bias). The baseline blocks involved the same procedure as the experimental blocks, such that the block began with an instruction baseline cue (3 sec), followed by six sequences of trials consisting of a baseline cue (1

sec), a black screen as the ISI (mean duration 1 sec, jittered) and a baseline video clip (3 sec).

[Please insert Figure 1 here]

The experiment was a mixed design; meaning we used both blocked (emotional cues) and event related (video clips). Thus, within an experimental block, the cue was always one emotion (e.g., happy block: a happy instruction cue followed by a happy video (congruent), then a happy cue followed by, for example, another happy video (congruent), then a happy cue followed by an angry video (incongruent) etc.). However, video clips within a block would alternate in emotion (e.g. a happy block of six video clips, would contain four happy video clips, one angry video clip and one neutral video clip). The video clips within a block were randomized in Microsoft Excel, so that the appearance of congruent or incongruent video clips could not be predicted. The emotion blocks were counterbalanced between runs, as were the runs between participants, using the Balanced Latin Squares method.

2.4 Procedure

The participants were asked to respond to the video clips to indicate if the emotion presented in the instruction cue matched the emotion expressed in the video. Specifically, participants were told to press the button with their index finger when the video clip was concordant with the emotion in the instruction cue and press with their middle finger when it was not. In the baseline condition, the participants were told to press any button during the baseline video clips. Participants were instructed to respond as quickly and as accurately as possible during the video clip. Accuracy and reaction times (RTs) were recorded for each trial.

Prior to the fMRI experiment, participants were trained with a practice task outside the MRI scanner. Both the practice task and fMRI task were presented using E-Prime 2.0 software (https://www.pstnet.com/eprime.cfm, 2013; Schneider et al., 2012) on a Windows computer screen. The practice task consistent of 9 blocks and feedback was given if the correct/incorrect button was pressed. The goal was to ensure that participants understood the aim of the task and that the finger response became automated outside the scanner. During the fMRI experiment the DEP task was seen by participants through a tilted mirror attached to the head coil on the MRI scanner. Responses were made on a custom-built MR-compatible response box.

After the fMRI experiment participants completed the two questionnaires: WASI (Wechsler, 1999) and Mayer-Salovey-Caruso Emotional Intelligence Test (MSCEIT; Mayer et al., 2003), in a testing room outside the MRI scanner. The practice task, fMRI task, and questionnaires were completed at the Centre for Advanced Imaging, University of Queensland 3T scanner facility.

2.5 MRI Procedure and Preprocessing

Structural and functional MRI images were acquired by a 3T Siemens Magnetom TrioTim system using a 12-channel head coil. The scans collected for each participant were as follows: localizer, T1-weighted anatomical image MP2-RAGE sequence (repetition time (TR): 1900 ms, echo time (TE): 2.32 ms, resolution: 1 mm^3, FoV= 230 mm, 192 slices, inversion time (TI): 900 ms, flip angle: 9 degrees), whole-head T2*-weighted echo-planar sequence (TR: 3000 ms, TE: 30 ms, resolution: 2.5 mm^3, slices: 46, FoV: 192 mm, flip angle: 90 degrees), DWI (TR: 8400 ms, TE: 100 ms, resolution: 2.3 mm x 2.3 mm x 2.5 mm, slices: 60, FoV: 300 mm, b-value: 2000

s/mm^2, directions: 64), and resting-state (TR: 3000 ms, TE: 30 ms, resolution: 2.5 mm^3, slices: 46, FoV: 192 mm). The total scanning time per session was 45min.

Standard preprocessing of the images was carried out using SPM8 (http://www.fil.ion.ucl.ac.uk/spm/software/spm8, 2013; Friston, 2003). The preprocessing steps were as follows: slice timing on the functional images, to correct for differences in slice acquisition times within each volume using the middle slice as reference; realignment (estimate and reslice) on the functional images, to correct for inter-scan movement within each run (no participant was excluded for excessive movement (defined as >3 mm translation, >2 degrees rotation); co-registration of the functional and structural images; segmentation of the structural image, with heavy regularisation (0.1) recommended for MP2-RAGE sequence; normalization of the resliced images into a standardized, stereotaxic space (according to the Montreal Neurological Institute template); and smoothing of normalized images with a 6mm full-width-at-half-maximum isotropic Gaussian kernel.

2.6 Data analysis

2.6.1 Preliminary analysis

Convergent validity was assessed by examining the correlation between the new DEP task and the MSCEIT (Mayer *et al.*, 2003). The MSCEIT remains the flagship test for emotional intelligence, consistently providing high validity and reliability (Mayer *et al.*, 2012). We chose two components of the MSCEIT, Perceiving Emotions and Understanding Emotions, as these components aim to measure the perception and interpretation of emotion, similarly to the DEP task (Kee *et al.*, 2009). Within the two components of the MSCEIT were subcomponents (i) Perceiving Emotions, which included identifying emotions conveyed in 'faces' and

'pictures' of paintings and landscapes, and (ii) Understanding Emotions, which included understanding the 'blends' of emotions and recognizing how emotions may 'change' and develop (Mayer et al., 2003).

Significant correlations were observed between RTs and accuracy percentages within the DEP task and components of the MSCEIT. Specifically, both RTs and accuracy had significant correlations with the 'blends' MSCEIT subcomponent ($r = -.34$, $p = .04$; $r = .55$, $p = .002$, respectively). The 'faces' subcomponent of Perceiving Emotions, which particularly converges with the DEP task, had a significant correlation with DEP accuracy ($r = .34$, $p = .04$). Finally, DEP accuracy was also significantly correlated with both Perceiving Emotions ($r = .38$, $p = 0.03$) and Understanding Emotions ($r = .49$, $p = .01$).

2.6.2 Behavioural analysis

Mean RTs and percentage accuracy from all responses acquired during scanning were calculated for each participant, across 12 conditions (3 runs x 2 video x 2 cue). These analyses were conducted using the SPSS package (SPSS Inc.). We conducted a factorial ANOVA with a 3 (experimental run) x 2 (emotion) x 2 (video) within-subjects design to assess the effect of congruency on RTs and accuracy with factors: experimental run (Run 1, Run 2, Run 3), emotion (happy and angry) and video (congruent and incongruent). Specifically, congruent video indicates the pairing of happy cues with happy video clips and angry cues with angry video clips, whereas incongruent video indicates the pairing of angry cues with happy video clips and happy cues with angry video clips. Degrees of freedom were corrected using Greenhouse-Geisser estimates of sphericity where the assumption of sphericity had been violated. Significant interactions were further analyzed using Bonferonni-

corrected simple contrasts and paired-samples t-tests, after running 1000 permutations.

2.6.3 fMRI analysis

In the present study, we employed a multivariate approach PLS, which investigates the distributed patterns of neural activity rather than the independent activity of a single brain region, as we believe that this analysis is more appropriate for our data as naturalistic emotion perception engages a widespread and interactive brain network (Vuilleumier and Pourtois, 2007; Arsalidou *et al.*, 2011). In addition, we wanted to avoid a-priori contrasts and forced cognitive subtraction, as it assumes that there are no interactions among cognitive components in a task (Amaro Jr and Barker, 2006). PLS does not constrict the brain activation based on researcher-defined contrasts; instead, it identifies the fundamental relations (latent variables: LVs) between brain activity and experimental conditions that account for maximum covariance in the data. Similar to principal component analysis, PLS decomposes the data into orthogonal dimensions by conducting singular value decomposition (SVD). Furthermore, PLS provides enhanced statistical power and sensitivity for paradigms such as ours, which involve event-related features (McIntosh *et al.*, 2004).

Because our interest was in brain activity during the emotional video clips, not activity during the emotional cues, we isolated activity during the video clips by conducting the analysis across five TRs (TR 0 – TR 5) starting at the onset of the video clips. Activity at each time point was normalized to the first TR (labelled TR 0 in the figures) and thus, our measure of dynamic emotion perception activity was relatively uninfluenced by cue activity. In an event-related paradigm, PLS provides a set of brain regions related to the experimental conditions for each TR on each LV. At each TR, for each participant a 'brain score' is calculated by multiplying the

'salience' (i.e. the degree of covariance of activity with the task condition on each LV) of each voxel by the signal of each brain voxel, and summing these across the entire brain. We plotted the mean brain scores at each TR to show overall brain activity fluctuations across the different conditions expressed over the 15 sec period, which is analogous to hemodynamic response functions that are typically plotted for individual brain regions.

For all analyses, we ran 500 permutations, consistent with the literature (McIntosh and Lobaugh, 2004; Burianova and Grady, 2007), to determine significant LVs at $p < 0.002$. In addition, we ran 100 bootstraps, estimating the standard errors of the salience for each voxel in order to assess the reliability and robustness of each voxel's contribution to a pattern of brain activity. We used the mean-centering approach in PLS, which involves subtracting the grand mean of the data matrix from the task means. We restricted the bootstrap ratio threshold to ± 3 ($p < 0.0001$) and reported areas with a cluster size of 100 or more voxels. Confidence intervals (95%) were calculated from the bootstrap; for the mean brain scores in each condition across the five TRs, significant differences between conditions were determined by a lack of overlap in the confidence intervals.

3. Results

3.1 Behavioural findings

We assessed the effect of congruency on RTs across different emotional valences and across experimental runs. A three-way analysis of variance yielded a main effect of run, $F(1.61, 43.42) = 5.31$, $p = 0.01$, RTs in the first run (Mean (M) = 864.45; Standard Deviation (SD) = 196.12) were slower than run 2 ($M = 796.22$; $SD = 189.55$), $p = 0.02$. No significant difference was found between the first and third run

(M = 805.90; SD = 227.64) or second and third run. Additionally, a main effect of emotion was found, $F(1, 27) = 11.38$, $p < 0.0001$. Angry videos (M = 802.34; SD = 189.47) were faster to detect than happy videos (M = 842.05; SD = 201.47), p = 0.002. There was also a main effect of video, $F(1, 27) = 24.71$, $p < 0.0001$, such that videos congruent with the instruction cue (M = 858.59; SD = 219.29) obtained significantly slower RTs compared to videos incongruent with the instruction cue (M = 785.80; SD = 171.65).

There was a significant interaction between experimental run and video ($F(1.61, 43.53) = 10.43$, $p < 0.0001$), indicating that the difference between RTs for video clips congruent and incongruent with the instruction cue was depended on which experimental run was being performed. To break down this interaction, simple contrasts were conducted comparing experimental runs and videos, an RT difference was revealed between run 1 and both run 2 and run 3, for congruent versus incongruent videos, (Run 1 vs Run 2: $F(1,27) = 10.78$, $p = 0.003$; Run 1 vs Run 3: $F(1,27) = 13.62$, $p = 0.001$). Displayed in the RT interaction graph (see **Figure 2**) and results from the paired-samples t-tests, these effects reflect that incongruent conditions (compared to congruent) decreased RTs more in the first run ($p = 0.001$) compared to run 2 ($p = 0.018$), and not in run 3 ($p = 0.69$). Paired-samples t-tests were then performed, showing that RTs significantly decreased during congruent conditions in run 2 (M = 825.64; SD = 219.72) and run 3 (M = 809.87; SD = 235.40) compared to run 1 (M = 940.24; SD = 256.02), $p = 0.001$ and 0.002, respectively. While remaining constant for incongruent conditions across experimental runs (*Ms (SD)*: Run 1 = 788.66 (155.92), Run 2 = 766.79(171.33), Run 3 = 801.93(229.73); *ns*). We did not find a significant interaction between emotion and video, which

suggests that the congruency effect on RTs was consistent across the different valences.

Finally, we assessed the effect of congruency on accuracy across different emotional valences and across experimental runs. A three-way analysis of variance did not produce any significant main effects of run, emotion or video, however, there was a significant interaction between experimental run and video, $F(2,54) = 7.42$, $p = 0.001$. The significant interaction was further analyzed using contrasts. There was an accuracy difference between run 1 and both run 2 and run 3, for congruent versus incongruent videos, (Run 1 vs Run 2: $F(1,27) = 8.10$, $p = 0.008$; Run 1 vs Run 3: $F(1,27) = 10.68$, $p = 0.003$). The accuracy interaction graph (see **Figure 2**) and paired-samples t-tests, indicate that incongruent conditions (compared to congruent) increased accuracy significantly in the first run ($p = 0.04$), but not in run 2 ($p = 0.30$) or run 3 ($p = 0.17$). Similar to the RT results, it was demonstrated by paired-samples t-tests that during congruent conditions accuracy increased in run 2 ($M = 99.31\%$; $SD = 1.51\%$) and run 3 ($M = 99.05\%$; $SD = 2.23\%$) compared to run 1 ($M = 97.32\%$; $SD = 3.90\%$), $p = 0.02$ (for both). While remaining constant during incongruent conditions across experimental runs (*Ms (SD)*: Run 1 = 99.36% (1.90%), Run 2 = 98.60% (3.77%), Run 3 = 98.31% (3.11%); *ns*). Again, we did not find a significant interaction between emotion and video, which suggests that the congruency effect on accuracy was consistent across the different valences.

[Please insert Figure 2 here]

3.2 fMRI-PLS Results

Our focus was on spatiotemporal activity of brain regions that was differentially involved in congruent and incongruent dynamic emotion perception. The congruent conditions were happy videos preceded by the happy instruction cue

and angry videos preceded by the angry instruction cue. The incongruent conditions were angry videos preceded by the happy instruction cue and happy videos preceded by the angry instruction cue. For further analyses, involving baseline and neutral stimuli see the Supplementary Materials. Specifically, **Supplementary Materials 1** show the analysis with baseline conditions, confirming that the novel DEP task successfully activates regions known to be involved in emotion perception; and **Supplementary Materials 2** include the neutral conditions.

The significant LV accounted for 63.60% of covariance in the data and differentiated congruent and incongruent conditions (see **Figure 3**). In line with previous literature, the vmPFC and dorsal medial prefrontal (dmPFC) regions known to be involved in congruency with prior expectations (Summerfield and Koechlin, 2008; Barbalat *et al.*, 2013) showed significantly greater activation for congruent compared to incongruent conditions. Additional areas active for dynamic emotions under congruent conditions were the, bilateral inferior frontal gyri, right amygdala, left putamen, and right insula.

Incongruent emotions recruited a wide network of regions, including the left middle occipital gyrus, cuneus, left precuneus, left superior parietal lobule, bilateral inferior parietal lobule, left middle frontal gyrus, medial frontal gyrus and left precentral gyrus, areas known to be involved in visual automatic change detection (Clery *et al.*, 2013). Additional areas included the anterior mid-cingulate gyrus, left middle temporal gyrus, and bilateral insula, regions critical in prediction error (Summerfield and Koechlin, 2008; Barbalat *et al.*, 2013). Finally, we identified the following regions during incongruent trials: the rTPJ, posterior mid-cingulate, right thalamus, and left claustrum (see **Table 1**).

[Please insert Figure 3 and Table 1 here]

A critical advantage of PLS is that it provides temporal information for each of the activation peaks, enabling extraction of regional activity for each TR and plotting of time courses (McIntosh *et al.*, 2004). To investigate the nature of changes in activity, the percent signal change was examined over time for specified regions of interest: vmPFC, right amygdala, rTPJ and posterior region of the mid-cingulate gyrus. We identified the time point of maximal difference for congruent conditions from incongruent conditions at TR 5, while incongruent conditions showed greatest divergence from congruent conditions at TR1 (see **Figure 4**).

[Please insert Figure 4 here]

5. Discussion

The purpose of the current study was to develop and validate a novel fMRI paradigm with improved ecological validity, using dynamic, audio-visual video clips, instead of static pictures as stimuli. The DEP was designed to investigate the prior expectancy effect on response rate and spatiotemporal neural activity during naturalistic emotion perception. Our study using dynamic emotional stimuli, found some results consistent with previous literature, which used static stimuli (Summerfield and Koechlin, 2008; Barbalat et al., 2013; Clery et al., 2013); however, our study also revealed significant differences in brain activity and behaviour. Our findings show that: (1) recognition of emotions congruent with prior expectations is initially slower than recognition of incongruent emotions, engaging the prefrontal cortices, which suggests greater executive functioning processes; (2) over time, recognition of emotions congruent with prior expectations becomes faster, engaging the amygdala and putamen, which suggests emotional learning (Hamann, 2001; Hooker *et al.*, 2006) and conditioning (Tricomi *et al.*, 2009; Brovelli *et al.*, 2011); and (3) recognition of incongruent emotions is consistently fast, recruiting visual

automatic change detection neural correlates and posterior mid-cingulate gyrus, which suggests quick adjustments of attention to change (Decety and Lamm, 2007; Vogt, 2014).

The newly developed DEP task provides convergent validity, suitability for use in fMRI, and increased ecological validity. The DEP task has convergent validity with the Perceiving Emotions and Understanding Emotions components of the MSCEIT, which is a reliable and valid measure of emotional intelligence (Mayer *et al.*, 2012). In addition, the DEP task has increased ecological validity, as the presentation of emotion includes both biological movement and multimodal sensory information, which are critical in the investigation of emotion perception; previous work has relied on static pictures of emotional expressions, only representing a single facial pattern, resulting in decreased activity in key social-emotion regions (Arsalidou *et al.*, 2011). In the current study, we replicated the findings of previous studies investigating prior expectation on emotion, such that the vmPFC was recruited for congruent emotion processing, whereas the anterior mid-cingulate gyrus, middle temporal gyrus, insula (Summerfield and Koechlin, 2008; Barbalat *et al.*, 2013), occipital regions, posterior parietal regions and premotor gyri (Clery *et al.*, 2013) were recruited during incongruent emotion processing. More importantly, however, the advantage of increased ecological validity enabled the recruitment of key regions of the social brain, namely amygdala and rTPJ (Van Overwalle, 2009; Kennedy and Adolphs, 2012). Thus, we were able to examine the contribution of these regions for conditions congruent and incongruent with prior expectations.

Congruency with prior expectation involved greater top-down processing of emotional content and emotion learning and conditioning, with regions peaking at later time points of hemodynamic response. Activations included the prefrontal cortex

(superior, medial and inferior regions) and right amygdala, which are areas implicated in higher-order emotion processing, such as evaluation and decision-making during emotion processing, encoding of emotional experiences in memory and emotion-related learning (Hamann, 2001; Baker and Kim, 2004; Koenigs *et al.*, 2007; Lindquist *et al.*, 2012). Activations were also found in the putamen, an area consistently active during conditioning and habituation induced by visual cues (Tricomi *et al.*, 2009; Brovelli *et al.*, 2011). The engaged network of activated regions was found to peak at later time points (i.e., TR 5), in addition we identified slower response times for congruent conditions, indicative of the need for greater cognitive engagement (Theeuwes *et al.*, 2000). Slower response times for congruency with prior expectations have been reported previously in the literature (Dieguez-Risco *et al.*, 2015), particularly for stimuli that carry greater uncertainty and require evaluation of congruency. It is argued that when asked to judge emotional congruency in dynamic displays of emotion there is initially higher uncertainty because the emotional expression may suddenly change, and therefore the evaluator takes longer than when judging emotional incongruency, which is immediately evident to differ from the emotional cue. With time, however, the response time to emotional congruent stimuli decreases as the nature of the emotional contingency is learned. This is demonstrated in our study as response times became faster over time (i.e., across the experimental runs), only during congruent conditions, giving support to temporal cue-related emotion learning. Thus, in contrast to previous prior expectancy studies (Barbalat *et al.*, 2013), the results of the current study demonstrate that congruently cued emotions require greater cognitive engagement and processes such as evaluation and decision-making, and only after repeated cues and conditioning, prior expectations are reinforced to facilitate emotion perception. These findings are

akin to real life where prior expectations are not instantaneous and develop through recurrent experience and learning.

In contrast to congruency with prior expectations, incongruent trials engaged regions previously associated with visual automatic change detection and rapid cognition. An area active during automatic change detection (Clery *et al.*, 2013), which is also a key hub of the social brain (Van Overwalle, 2009; Kennedy and Adolphs, 2012), is the TPJ. This region was recruited during incongruent emotion conditions and it is known to be particularly involved in rapid processing of unexpected cues and enabling of immediate reactions to social situations (Decety and Lamm, 2007; Ciaramelli *et al.*, 2008). Additionally, we found activations in the posterior parts of the mid-cingulate gyrus, an area critically important in rapid cognitive adjustments and cognitive efficiency (Leech and Sharp, 2014; Vogt, 2014). The activity of rTPJ and posterior mid-cingulate gyrus peaked early, demonstrating evidence for their roles in efficient and rapid detection of incongruent emotion. The suggested quick detection for incongruent emotion is in line with the consistently fast response rates during incongruent conditions, relative to the initial slower responding during congruent conditions. Our findings for fast detection of incongruent emotion are novel and make sense from an evolutionary perspective, because in a dynamic and complex environment, advantageous or dangerous situations (critical to our survival) may arise unexpectedly, requiring automatic detection and attention directed towards a potential opportunity or threat (Juth *et al.*, 2005; Bishop, 2008; Öhman, 2009).

Critically, we provide crucial new insights into the influence of prior expectations on perception of naturalistic emotion. In the case of dynamic audio-visual emotional stimuli, initial congruence detection may be slower because the participant is evaluating, with certainty, whether the emotional content (which is in

constant motion, with changing audio-visual features) is congruent with previous information. The requirement to evaluate congruency is in line with our fMRI findings, as key brain regions in emotion evaluation and decision-making were associated with the congruent conditions. During incongruent trials, on the other hand, the participant reaches a decision sooner because once a different emotional valence is detected one can be certain that it is an incongruent emotion. We suggest that this quick detection of change in emotional valence during incongruent conditions may in part be driven by visual automatic change detection processes, as supported by the fMRI results. One limitation of this study is that the results pertain only to a male population. The male sample was recruited in order to reduce heterogeneity, as males and females may differ in their emotion perception (Lambrecht et al., 2014; Stevens and Hamann, 2012). The validity and reliability of this novel paradigm will be the focus of follow-up studies in which we will investigate the role of prior expectancy in healthy females and in patients suffering from disorders associated with dysfunctional emotion perception; specifically, in patients with schizophrenia who have been reported to have an overreliance on prior expectations (Chambon *et al*., 2011b).

5.1 Conclusion

In conclusion, we present a novel fMRI paradigm, which gauges the influence of prior expectations on dynamic emotion perception with increased ecological validity. The improved methodology enabled us to provide new insights into the behavioural and neural mechanisms of prior expectations and delineate the spatiotemporal activity of brain regions active during dynamic emotion perception. Specifically, prior expectations seem to be reinforced by top-down processes, which direct attention to emotional events congruent with expectations, and which, with

repeated exposure, facilitate perception. Conversely, emotional events incongruent with prior expectations gain attention via change detection processes, an adaptive behaviour, which allows one to automatically detect a change in one's environment and identify a potential opportunity or threat.

Acknowledgements

References

Amaro Jr, E. & Barker, G. J. (2006) 'Study design in fMRI: Basic principles', *Brain and Cognition*, **60**(3), pp. 220-232.

Andreano, J. M. & Cahill, L. (2009) 'Sex influences on the neurobiology of learning and memory', *Learning & Memory*, **16**(4), pp. 248-266.

Arsalidou, M., Morris, D. & Taylor, M. J. (2011) 'Converging Evidence for the Advantage of Dynamic Facial Expressions', *Brain Topography*, **24**(2), pp. 149-163.

Baker, K. B. & Kim, J. J. (2004) 'Amygdalar lateralization in fear conditioning: Evidence for greater involvement of the right amygdala', *Behavioral Neuroscience*, **118**(1), pp. 15-23.

Balota, D. A., Yap, M. J., Cortese, M. J., Hutchison, K. A., Kessler, B., Loftis, B., Neely, J. H., Nelson, D. L., Simpson, G. B. & Treiman, R. (2007) 'The English Lexicon Project', *Behavior Research Methods*, **39**(3), pp. 445-459.

Barbalat, G., Bazargani, N. & Blakemore, S. J. (2013) 'The Influence of Prior Expectations on Emotional Face Perception in Adolescence', *Cerebral Cortex*, **23**(7), pp. 1542-1551.

Barbalat, G., Rouault, M., Bazargani, N., Shergill, S. & Blakemore, S. J. (2012) 'The influence of prior expectations on facial expression discrimination in schizophrenia', *Psychol Med*, **42**(11), pp. 2301-2311.

Bishop, S. J. (2008) 'Neural mechanisms underlying selective attention to threat', *Molecular and Biophysical Mechanisms of Arousal, Alertness, and Attention*, **1129**, pp. 141-152.

Blasi, G., Hariri, A. R., Alce, G., Taurisano, P., Sambataro, F., Das, S., Bertolino, A., Weinberger, D. R. & Mattay, V. S. (2009) 'Preferential Amygdala Reactivity to the Negative Assessment of Neutral Faces', *Biological Psychiatry*, **66**(9), pp. 847-853.

Brovelli, A., Nazarian, B., Meunier, M. & Boussaoud, D. (2011) 'Differential roles of caudate nucleus and putamen during instrumental learning', *Neuroimage*, **57**(4), pp. 1580-1590.

Burianova, H. & Grady, C. L. (2007) 'Common and unique neural activations in autobiographical, episodic, and semantic retrieval', *Journal of Cognitive Neuroscience*, **19**(9), pp. 1520-1534.

Chambon, V., Domenech, P., Pacherie, E., Koechlin, E., Baraduc, P. & Farrer, C. (2011a) 'What Are They Up To? The Role of Sensory Evidence and Prior Knowledge in Action Understanding', *PLoS One*, **6**(2).

Chambon, V., Pacherie, E., Barbalat, G., Jacquet, P., Franck, N. & Farrer, C. (2011b) 'Mentalizing under influence: abnormal dependence on prior expectations in patients with schizophrenia', *Brain*, **134**, pp. 3725-3738.

Ciaramelli, E., Grady, C. L. & Moscovitch, M. (2008) 'Top-down and bottom-up attention to memory: A hypothesis (AtoM) on the role of the posterior parietal cortex in memory retrieval', *Neuropsychologia*, **46**(7), pp. 1828-1851.

Clery, H., Andersson, F., Fonlupt, P. & Gomot, M. (2013) 'Brain correlates of automatic visual change detection', *Neuroimage*, **75**, pp. 117-122.

Collignon, O., Girard, S., Gosselin, F., Roy, S., Saint-Amour, D., Lassonde, M. & Lepore, F. (2008) 'Audio-visual integration of emotion expression', *Brain Research*, **1242**, pp. 126-135.

Cooney, R. E., Atlas, L. Y., Joormann, J., Eugene, F. & Gotlib, I. H. (2006) 'Amygdala activation in the processing of neutral faces in social anxiety disorder: Is neutral really neutral?', Psychiatry Research-Neuroimaging, 148(1), pp. 55-59.

Corp., I. (Released 2013). IBM SPSS Statistics for Windows (Version Version 22.0). Armonk, NY: IBM Corp.

Decety, J. & Lamm, C. (2007) 'The role of the right temporoparietal junction in social interaction: How low-level computational processes contribute to meta-cognition', *Neuroscientist*, **13**(6), pp. 580-593.

Dieguez-Risco, T., Aguado, L., Albert, J. & Hinojosa, J. A. (2015) 'Judging emotional congruency: Explicit attention to situational context modulates processing of facial expressions of emotion', *Biological Psychology*, **112**, pp. 27-38.

Dolan, R. J. (2007) 'The human amygdala and orbital prefrontal cortex in behavioural regulation', *Philosophical Transactions of the Royal Society B-Biological Sciences*, **362**(1481), pp. 787-799.

Friston, K., Kilner, J. & Harrison, L. (2006) 'A free energy principle for the brain', *J Physiol Paris*, **100**(1-3), pp. 70-87.

Friston, K. J. (2003) 'Statistical parametric mapping', *Neuroscience Databases*, Springer.

Geng, J. J. & Mangun, G. R. (2011) 'Right temporoparietal junction activation by a salient contextual cue facilitates target discrimination', *Neuroimage*, **54**(1), pp. 594-601.

Hamann, S. (2001) 'Cognitive and neural mechanisms of emotional memory', *Trends in Cognitive Sciences*, **5**(9), pp. 394-400.

Hooker, C. I., Germine, L. T., Knight, R. T. & D'Esposito, M. (2006) 'Amygdala response to facial expressions reflects emotional learning', *Journal of Neuroscience*, **26**(35), pp. 8915-8922.

Juth, P., Lundqvist, D., Karlsson, A. & Ohman, A. (2005) 'Looking for foes and friends: Perceptual and emotional factors when finding a face in the crowd', *Emotion*, **5**(4), pp. 379-395.

Kaufman, J. & Johnston, P. J. (2014) 'Facial Motion Engages Predictive Visual Mechanisms', *PLoS One*, **9**(3).

Kee, K. S., Horan, W. R., Salovey, P., Kern, R. S., Sergi, M. J., Fiske, A. P., Lee, J. H., Subotnik, K. L., Nuechterlein, K., Sugar, C. A. & Green, M. F. (2009) 'Emotional intelligence in schizophrenia', *Schizophrenia Research*, **107**(1), pp. 61-68.

Kennedy, D. P. & Adolphs, R. (2012) 'The social brain in psychiatric and neurological disorders', *Trends in Cognitive Sciences*, **16**(11), pp. 559-572.

Kesler, M. L., Andersen, A. H., Smith, C. D., Avison, M. J., Davis, C. E., Kryscio, R. J. & Blonder, L. X. (2001) 'Neural substrates of facial emotion processing using fMRI', Cognitive Brain Research, 11(2), pp. 213-226.

Kilts, C. D., Egan, G., Gideon, D. A., Ely, T. D. & Hoffman, J. M. (2003) 'Dissociable neural pathways are involved in the recognition of emotion in static and dynamic facial expressions', *Neuroimage*, **18**(1), pp. 156-168.

Koenigs, M., Young, L., Adolphs, R., Tranel, D., Cushman, F., Hauser, M. & Damasio, A. (2007) 'Damage to the prefrontal cortex increases utilitarian moral judgements', *Nature*, **446**(7138), pp. 908-911.

Kreegipuu, K., Kuldkepp, N., Sibolt, O., Toom, M., Allik, J. & Naatanen, R. (2013) 'vMMN for schematic faces: automatic detection of change in emotional expression', *Frontiers in Human Neuroscience*, **7**.

Lambrecht, L., Kreifelts, B. & Wildgruber, D. (2014) 'Gender differences in emotion recognition: Impact of sensory modality and emotional category', *Cognition & Emotion*, **28**(3), pp. 452-469.

Lee, E., Kang, J. I., Park, I. H., Kim, J. J. & An, S. K. (2008) 'Is a neutral face really evaluated as being emotionally neutral?', *Psychiatry Research*, **157**(1-3), pp. 77-85.

Leech, R. & Sharp, D. J. (2014) 'The role of the posterior cingulate cortex in cognition and disease', *Brain*, **137**(Pt 1), pp. 12-32.

Lindquist, K. A., Wager, T. D., Kober, H., Bliss-Moreau, E. & Barrett, L. F. (2012) 'The brain basis of emotion: A meta-analytic review', *Behavioral and Brain sciences*, **35**(3), pp. 121-143.

MATLAB Release 2012b, The MathWorks, Inc., Natick, Massachusetts, United States.

Mayer, J. D., Salovey, P. & Caruso, D. R. (2012) 'The Validity of the MSCEIT: Additional Analyses and Evidence', *Emotion Review*, **4**(4), pp. 403-408.

Mayer, J. D., Salovey, P., Caruso, D. R. & Sitarenios, G. (2003) 'Measuring Emotional Intelligence With the MSCEIT V2.0', *Emotion*, **3**(1), pp. 97-105.

McIntosh, A. R., Chau, W. & Protzner, A. B. (2004) 'Spatiotemporal analysis of event-related fMRI data using partial least squares', *Neuroimage*, **23**(2), pp. 764-775.

McIntosh, A. R. & Lobaugh, N. J. (2004) 'Partial least squares analysis of neuroimaging data: applications and advances', *Neuroimage*, **23**, pp. S250-S263.

Öhman, A. (2009) 'Of snakes and faces: An evolutionary perspective on the psychology of fear', *Scandinavian Journal of Psychology*, **50**(6), pp. 543-552.

Palermo, R. & Rhodes, G. (2007) 'Are you always on my mind? A review of how face perception and attention interact', *Neuropsychologia*, **45**(1), pp. 75-92.

Palumbo, L. & Jellema, T. (2013) 'Beyond Face Value: Does Involuntary Emotional Anticipation Shape the Perception of Dynamic Facial Expressions?', *PLoS One*, **8**(2).

Robins, D. L., Hunyadi, E. & Schultz, R. T. (2009) 'Superior temporal activation in response to dynamic audio-visual emotional cues', *Brain and Cognition*, **69**(2), pp. 269-278.

Sabatinelli, D., Fortune, E. E., Li, Q. Y., Siddiqui, A., Krafft, C., Oliver, W. T., Beck, S. & Jeffries, J. (2011) 'Emotional perception: Meta-analyses of face and natural scene processing', Neuroimage, 54(3), pp. 2524-2533.

Schneider, W., Eschman, A., and Zuccolotto, A. (2012). E-Prime User's Guide. Pittsburgh: Psychology Software Tools, Inc.

Sheehan, D. V., Lecrubier, Y., Sheehan, K. H., Janavs, J., Weiller, E., Keskiner, A., Schinka, J., Knapp, E., Sheehan, M. F. & Dunbar, G. C. (1997) 'The validity of the Mini International Neuropsychiatric Interview (MINI) according to the SCID-P and its reliability', *European Psychiatry*, **12**(5), pp. 232-241.

Stevens, J. S. & Hamann, S. (2012) 'Sex differences in brain activation to emotional stimuli: A meta-analysis of neuroimaging studies', *Neuropsychologia*, **50**(7), pp. 1578-1593.

Summerfield, C. & Koechlin, E. (2008) 'A neural representation of prior information during perceptual inference', *Neuron*, **59**(2), pp. 336-347.

Theeuwes, J., Atchley, P. & Kramer, A. F. (2000) 'On the time course of top-down and bottom-up control of visual attention', *Control of Cognitive Processes: Attention and Performance Xviii*, pp. 105-124.

Tricomi, E., Balleine, B. W. & O'Doherty, J. P. (2009) 'A specific role for posterior dorsolateral striatum in human habit learning', *European Journal of Neuroscience*, **29**(11), pp. 2225-2232.

Van Overwalle, F. (2009) 'Social Cognition and the Brain: A Meta-Analysis', *Human Brain Mapping*, **30**(3), pp. 829-858.

Vogt, B. A. (2014) 'Submodalities of emotion in the context of cingulate subregions', *Cortex*, **59**, pp. 197-202.

Vuilleumier, P. & Pourtois, G. (2007) 'Distributed and interactive brain mechanisms during emotion face perception: Evidence from functional neuroimaging', *Neuropsychologia*, **45**(1), pp. 174-194.

Wechsler, D. (1999). Wechsler Abbreviated Scale of Intelligence. San Antonio, TX: The Psychological.

Yoshikawa, S. & Sato, W. (2006) 'Enhanced perceptual, emotional, and motor processing in response to dynamic facial expressions of emotion', *Japanese Psychological Research*, **48**(3), pp. 213-222.

TABLE 1

Differences in activity during congruent vs. incongruent emotion perception

Peak MNI	Hem	BA	MNI coordinates			Voxels	BSR
			x	y	z		
Congruent > Incongruent							
Inferior frontal	B	47	-30	6	-22	135	5.03
			32	18	-16	101	5.34
Superior frontal	B	8/9	10	66	20	139	4.71
			-6	58	20	134	4.24
vmPFC / sgACC		11	-4	34	-12	183	4.44
Amygdala	R	28	18	-8	-18	146	4.83
Insula	R	13	40	26	10	570	5.75
Putamen	L		-22	4	8	181	5.22
Incongruent > Congruent							
Precentral	L	6	-28	-6	70	285	5.90
Middle frontal	L	6	-28	-2	64	110	4.68
Medial frontal		6	-4	6	74	220	4.58
Anterior mid-cingulate		24	2	4	48	108	5.22
Posterior mid-cingulate		23	-2	-24	32	140	4.92
Insula	B	13	-58	-36	20	1551	6.75
			38	-26	16	114	4.65
Thalamus	R		10	-4	16	171	5.80
Claustrum	L		-36	-20	-2	172	4.43
Superior parietal lobule	L	7	-26	-52	50	911	7.59
Precuneus	L	7	-4	-66	40	2641	7.10
Inferior parietal lobule	B	40	-42	-46	62	733	5.77
			44	-28	40	177	5.52
Temporoparietal junction	R	40	60	-50	34	939	5.35
Postcentral	B	2/40	60	-22	20	222	5.29
		2	-38	-22	36	100	3.93
Superior temporal	L	22	-60	-50	14	131	5.29
Middle temporal	L	39	-32	-64	32	184	5.11
Middle occipital	L	18/19	-32	-80	24	11015	9.01
Cuneus		7	-2	-64	38	3394	7.21
Lingual		18	6	-74	8	1457	6.44

Abbreviations: Hem = hemisphere; BA = Brodmann area; R = right; L = left; B = bilateral; BSR = bootstrap ratio; voxels = number of voxels (one voxel volume=6 mm3). All reported activations are ≥100 voxels (600 mm^3). sgACC = subgenual anterior cingulate cortex; vmPFC = ventromedial prefrontal cortex.

Figure Captions

FIG 1. Dynamic emotion perception task. A schematic diagram showing an example of a happy emotion block. Participants were asked to press a button with their index finger for happy videos (congruent trial) and press a button with their middle finger for every other emotion video (incongruent trial).

FIG 2. Mean reaction times and accuracy for experimental runs (Run 1, Run 2, Run 3) and video (congruent and incongruent). Error bars show the standard error of the mean.

FIG 3. Results of the mean-centered PLS analysis, differentiating congruent and incongruent conditions. The graph shows the brain scores from TR1, TR 4 and TR 5 (TR = 3 sec; error bars = 95% CIs from bootstrapping). All maxima have BSR $\geq$ 3.0 and cluster size $\geq$ 100 voxels (600 mm^3). Regions corresponding to areas activated during dynamic emotion under congruent conditions: vmPFC = ventromedial prefrontal cortex; rAMY = right amygdala; lPUT = left putamen; prefrontal cortex (bilateral inferior frontal gyri, bilateral superior frontal gyri); and incongruent: PC = posterior mid-cingulate; rTPJ = right temporoparietal junction; bilateral insula; occipital (left middle occipital gyrus, cuneus); left middle temporal gyrus; parietal (left superior parietal lobule, left precuneus, bilateral inferior parietal lobule, bilateral postcentral); anterior mid-cingulate; premotor (left precentral, left middle frontal).

FIG 4. For visualization purposes: time courses of activity from the whole-brain analysis, in the ventromedial prefrontal cortex (x = 46, y = 24, z = -16); right amygdala (x = 28, y = -2, z = -18); right temporoparietal junction (x = 56, y = -50, z = 32); posterior mid-cingulate (x = -2, y = -26, z = 30). Error bars show the standard error of the mean.