Neuropsychologia 50 (2012) 3410–3418

Contents lists available at SciVerse ScienceDirect

Neuropsychologia

journal homepage: www.elsevier.com/locate/neuropsychologia

Research Report

Discriminating Thatcherised from typical faces in a case of prosopagnosia

Natalie Mestry [a,*], Nick Donnelly [a], Tamaryn Menneer [a], Rosaleen A. McCarthy [a,b]

[a] Psychology, University of Southampton, Southampton SO17 1BJ, UK
[b] Wessex Neurological Centre, Southampton General Hospital, Tremona Road, Southampton SO16 6YD, UK

ARTICLE INFO

Article history:
Received 10 May 2012
Received in revised form
29 August 2012
Accepted 17 September 2012
Available online 23 September 2012

Keywords:
Prosopagnosia
Configural processing
Emotion
Thatcher illusion
Face processing

ABSTRACT

We report data from a prosopagnosic patient (PHD), and aged-matched control participants, from experiments where participants categorised individually presented emotional faces (Experiment 1) and Thatcherised (from typical) faces (Experiment 2). In Experiment 2 participants also discriminated between simultaneously presented Thatcherised and typical faces. PHD was at chance categorising Thatcherised from typical faces. He was, however, able to discriminate between Thatcherised and typical faces, and partially able to categorise emotional faces. The results are discussed in terms of a loss of configural processing but preserved feature processing in PHD. The loss of configural processing impacts his categorisation of Thatcherised and typical faces, and his emotion processing, while his preserved feature processing supports his ability to categorise some emotional faces and his ability to discriminate between Thatcherised and typical faces.

1. Discriminating Thatcherised from typical faces in a case of prosopagnosia

The Thatcher illusion (Thompson, 1980) refers to the change in perception of "Thatcherised" faces when they are rotated from upright to inverted orientations. Thatcherised faces are made by inverting the eyes and mouth. These faces appear grotesque when upright but when inverted the grotesqueness disappears and faces look more typical. There have been various accounts put forward to explain the Thatcher illusion. These accounts range from expression analysis (Valentine, 1988), to the phenomenological experience of conflicting reference frames for faces and facial features (Parks, Coss, & Coss, 1985). More recently, and following Bartlett and Searcy (1993), the Thatcher illusion is thought to reveal the orientation specific nature of configural face processing. Perception of grotesqueness in the upright face is attributed to the perception of unusual configural relationships between the features, whilst perception of the typical inverted face relies on feature based processing (Stürzel & Spillmann, 2000). Furthermore, the automaticity with which grotesqueness is experienced makes it a useful test of the presence of configural face processing in atypical populations (Rouse, Donnelly, Hadwin, & Brown, 2004) including congenital prosopagnosia (Carbon, Grüter, Weber, & Lueschow, 2007).

There are two (related) ways in which configural processing (defined here as the encoding of between-feature spatial relationships) might lead to the perception of the Thatcher illusion. First, by a poor match between representations of Thatcherised and prototypical faces. Second, by creating local difficulties in configural processing for inverted eyes and mouths which are in an unusual orientation relative to otherwise upright faces. By the first account, grotesqueness results from the fact that individuals rate average and not distinctive faces as attractive (Rhodes & Tremewan, 1996). By the second account, grotesqueness results from low processing fluency associated with processing Thatcherised faces as faces (Reber, Winkielman, & Schwartz, 1998). Thatcherised faces are poor examples of the face category and lead to a processing difficulty that is experienced as grotesqueness rather than slow processing.

The involvement of emotional coding, in addition to configural processing, in the Thatcher illusion is manifest in recent neuroimaging studies. Specifically, areas known to be involved in social and emotional processing are also involved in the perception of both single Thatcherised faces (Rothstein, Malach, Hadar, Graif, & Hendler, 2001), and when discriminating Thatcherised from typical faces (Donnelly et al., 2011). Therefore, despite being thought of as an illusion demonstrating configural processing in faces, and Thatcherised faces not representing standard emotional faces, the phenomenology of the Thatcher illusion depends on the response of diffuse socio-emotional cortices to Thatcherised faces. The consequence of this fact is that any use of the Thatcher illusion as a marker of configural processing should be accompanied by evidence of broadly intact emotional processing. Otherwise any failure in perception of the Thatcher illusion might result from deficits in socio-emotional processing.

But what do we mean by broadly intact emotional processing in the context of a deficit in configural processing? The issue is

* Corresponding author. Tel.: +02380 595078.
E-mail addresses: natalie.mestry@soton.ac.uk (N. Mestry), n.donnelly@soton.ac.uk (N. Donnelly), t.menneer@soton.ac.uk (T. Menneer), rosaleen.mccarthy@uhs.nhs.uk (R.A. McCarthy).

0028-3932/$ - see front matter © 2012 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.neuropsychologia.2012.09.034

Table 1
Performance of PHD across a range of cognitive assessments.

Task	Score
WAIS III[a]	
Verbal scale	
Vocabulary	9
Similarities	9
Arithmetic	5
Digit span	10
Performance scale	
Picture completion	6
Digit symbol coding	6
Block design	9
Matrix reasoning	10
Camden memory tests[b]	
Faces	12/25 (chance)
Words	21/25 (2nd percentile)
Scenes	20/30 (7–9th percentile)
Paired Associates	17/48 (< 1st percentile)
Graded naming test[c]	8/30 (< 1st percentile)

[a] Scores represent age – related scores, M = 10, SD = 3, Wechsler (1997);
[b] Warrington (1996);
[c] McKenna and Warrington (1983).

complicated by the fact that configural processing contributes to the perception of some facial emotions. For example, Calder, Young, Keane, and Dean (2000) measured response times to aligned and misaligned composite faces where the face composites are formed from the same or different emotions. By determining which emotions were responded to more quickly when the top and bottom halves of faces were aligned relative to misaligned, Calder et al. were able to determine which emotions have their perception facilitated by configural processing. Table 1 of Calder et al. (2000) indicates the perception of anger, fear and sadness from whole faces cannot be predicted from that of part faces. In contrast, the perception of disgust, happiness and surprise from whole faces can be predicted from that of part faces. This means that the detection of anger, fear and sadness is improved by computing configural information from whole faces.

The role of configurality in emotion processing was also explored by McKelvie (1995). McKelvie compared categorisation accuracy to emotional faces shown in upright and inverted orientations. Inversion led to less accurate categorisation of sadness, fear, anger and disgust than when upright. Happiness, neutrality and surprise were unaffected by orientation. Like Calder et al., McKelvie suggested emotional expressions rely on configurality to different degrees. Together these studies show configurality was important in the perception of anger, fear and sadness. Only in respect of the perception of disgust did the two studies differ.

What impact would the absence of configural processing have on emotion perception in general and the Thatcher illusion in particular? We propose that an absence of configural processing should be manifest in (1) a specific pattern of modest deficit in the recognition of facial emotions that partially rely on configural processing and (2) an inability to perceive Thatcherised faces. In this study we report on a series of experiments conducted on a brain-damaged patient with prosopagnosia. PHD has been reported previously, in an ERP study comparing unfamiliar faces and houses, and does not generate an N170 component in response to faces (Eimer & McCarthy, 1999). If a failure to generate a face effect at N170 is linked to a failure of configural processing in faces through an absence of face categorisation (e.g., Eimer, 2000), we should predict two related findings. First, PHD will be able to categorise facial emotion but only when this can be done from features. Second, PHD will be unable to categorise Thatcherised from typical faces.

We start our exploration of PHD by testing his ability to categorise facial emotions and to discriminate between faces exhibiting different levels of emotion. We predict PHD's categorisation of emotional faces will be impaired in those conditions that rely on configural processing. Nevertheless, we also predict PHD will have intact categorisation and discrimination of emotions and emotional intensity when this can be achieved through featural analysis. The goal of Experiment 1 is, therefore, to show that PHD is able to perform categorisation and discrimination of faces with emotional valence in at least some conditions. In Experiment 2, we go on to explore his ability to categorise and discriminate Thatcherised from typical faces. His ability to perform in these tasks is compared with that of controls.

2. Experiment 1

In Experiment 1, the sensitivity of PHD and age-matched controls participants to categorising emotions was measured. While we were interested in PHD's overall ability to categorise emotions, we were also interested in his ability to do so when only two emotions are possible. In this case, the task demands of categorization are the same as when categorizing Thatcherised from typical faces.

2.1. Method

2.1.1. Participants

An individual with prosopagnosia (PHD) volunteered to participate in studies regarding his deficit. PHD is a left-handed male, who was aged 48 to 51 over the course of the current experiments. He sustained a closed head injury as a result of a road traffic accident at the age of 17. Structural MRI in 2005 (Fig. 1) showed a unilateral lesion in the ventral temporal lobe in the region of the fusiform gyrus on the left with no other macroscopic areas of damage. PHD suffers significant cognitive deficits including apperceptive prosopagnosia and some category specific visual agnosia, especially for the living things domain (animals and fruit and vegetables). He has persistent difficulties recognising people from their faces without context or other supporting information. PHD has a mild deuteranomaly and corrected-to-normal visual acuity with eye-glasses and his visual fields are full.

PHD's most recent cognitive assessment showed him to be functioning at an average level on most subtests of the WAIS-III (Wechsler, 1997, Table 1). On the Visual Object and Spatial Perception Test battery (VOSP, Warrington, & James, 1991) he scored within the normal range on Screening, Fragmented Letters, Object

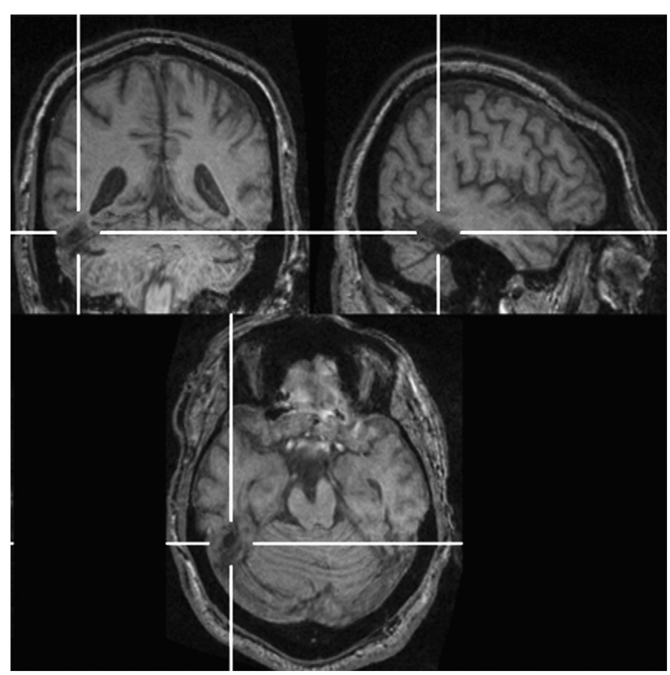

Fig. 1. Structural MRI taken from PHD showing a focal area of injury in the inferior temporal lobe of the left hemisphere in the region of the fusiform gyrus.

Decision, Dot Counting, Position Discrimination, Cube Analysis and Number Location. On the Warrington Recognition Memory Tests for Words and Faces (Warrington, 1984), PHD's recognition memory for words was above chance but within the clinical range for his age. His recognition memory for faces was at chance but he was within the normal range on the Camden test of memory for Topographic Scenes (Warrington, 1996, Table 1). When confronted with portraits of contemporary famous people, PHD was only able to identify the Queen and President Obama. He indicated some familiarity with some of the faces (e.g., asking "is he an entertainer?" for Bruce Forsyth). If the most lenient criterion for recognition is adopted, PHD scored 6/22 correct: a score below the poorest achieved by patients attending a clinic for people with moderate levels of Alzheimer's disease (McCarthy, personal data).

The Benton Face Recognition Test (Benton, Sivan, Hamsher, Varney, & Spreen, 1994) was used to assess the status of face perception. PHD was able find a single face in the array of six that is identical to the target, scoring 6/6 on this first part. The remaining 16 items require matching for facial identity across viewpoints: the participant must find three examples of the target face amongst an array of six items. PHD only succeeded on 30/48 choices, scoring within the severely impaired range for the test as a whole. PHD's score deteriorated further when the task was presented in an inverted orientation (28). He tried to solve the puzzle of matching the inverted faces by using an overt feature-naming strategy ("have they got the same eyebrows?" or "I can tell by the cheeks").

PHD's ability to extract emotion cues from faces was initially evaluated clinically using a paper based version of the Ekman 60 Faces Test, based on six facial expressions generated by 10 different people from Ekman and Friesen (1976) facial expression series (with stimuli kindly made available by Andy Young). The stimuli were selected so that each emotional expression was as well recognised as possible according to the Ekman and Friesen norms. They were presented singly, in a pseudo-random sequence, with the labels "happiness, fear, surprise, sadness, disgust, anger" printed underneath each face. PHD was asked "which word is closest to the emotion being experienced by the person in the picture". A preliminary series of six expressions was presented (with feedback) followed by 60 test faces, with 10 examples of each emotion, one posed by each person. PHD's total score was worse than that of a sample of 10 control participants (age range 50–68) from the Cambridge subject panel (PHD scored 41/60). He was very impaired on fear, only recognising 2/10 examples (controls: $M=8.6$, $SD=1.17$). PHD tended to misidentify fear as anger or surprise. He was also significantly impaired on sadness, scoring 6/10 (controls: $M=8.7$, $SD=1.34$) suggesting disgust, anger and fear as possible options. In the case of anger, he was within one standard deviation of controls, scoring 7 (controls: $M=7.7$, $SD=1.42$) despite misidentification of sad faces as angry. He was also within the range of controls for disgust (PHD scored 9; controls: $M=9.0$, $SD=1.25$); for surprise (PHD scored 8; controls: $M=8.5$, $SD=1.58$); and was not impaired at recognising happiness (PHD scored 9; controls: $M=9.9$, $SD=0.32$). Despite typical performance categorising angry faces, PHD was likely to mislabel other emotional faces as angry.

PHD's recognition of emotion was also assessed using body postures (kindly provided by Beatrice de Gelder). He was asked to judge whether a posture was happy, sad, angry or fearful. PHD was not as good as the controls reported by De Gelder and Van den Stock (2011), but his score of 20/24 was well above chance.

PHD was also assessed on the TASIT (McDonald, Flanagan, Rollins, & Kinch, 2003): a series of video vignettes that have been designed to evaluate the ability to extract emotional and social cues from short interactions. PHD performed well on those items evaluating positive emotions (happiness, surprise and neutral items) but was at or below the 5th percentile for the negative emotions of anger, anxiety/fear and revulsion. More complex social interactions were mostly understood well. See Eimer and McCarthy (1999) for further details of the abilities of PHD across a variety of measures.

In Experiment 1, a control group of four age-matched, males (age range 31 to 63, $M=52.25$ years, $SD=14.41$) were recruited. All participants had normal or corrected-to-normal vision and were asked to complete the Edinburgh Handedness Inventory (Oldfield, 1971) prior to participation. Three were left-handed. No controls had any history of neurological problems.

2.1.2. Stimuli

Grey-scale face morph stimuli were used from the Facial Expression of Emotion test (Young, Perrett, Calder, Sprengelmeyer, & Ekman, 2002). Angry, fearful, disgusted and happy faces morphed with neutral were shown (see Fig. 2). The intensity of the emotion in the morph was varied creating a range of morphs for each emotion with different intensities of emotion relative to neutral. All morphs were based on two images of the same individual. Forty-five morphs were used for each emotion: range 10–98% emotion at 2% intervals (180 morphs in total). Each stimulus appeared at a size of 6.20 cm × 8.80 cm on the screen. Therefore, the visual angle for each image was 7.10 × 10.06 degrees, when viewed on a desktop computer screen from a distance of 50 cm.

2.1.3. Procedure

Participants judged which of two possible emotions was being presented on each trial in six separate blocks (paired combinations of happy, angry, disgusted and fearful). A fixation cross was displayed for 300 ms followed by the stimulus presented centrally on a black background until response. When each stimulus face was removed it was replaced by a rectangle mask of monochrome Gaussian noise displayed for 500 ms. Ninety trials were shown in each condition, with each morph face being shown once. Participants responded by using the mouse buttons in order to select one of the two emotion labels presented on the screen. Condition order and the emotion assigned to each mouse button were counterbalanced for controls and PHD. Controls each completed the task once and PHD completed the task three times over a period of several months.

Fig. 2. Examples for morphed face stimuli shown in Experiment 1. Angry, disgust, fear and happiness (left to right). Percentage of emotion morphed with neutral: 10%, 50%, 90% (top to bottom).

2.2. Results

The data were aggregated into four intensity ranges: (10–30%; 32–52%; 54–74%; and 76–98%). These aggregated data were converted into sensitivity (d') and bias (c) scores. These data for controls were analysed in two separate 6 (Emotion-pair) × 4 (Intensity) repeated measures ANOVAs. Separate ANOVAs being performed for sensitivity and bias data.

For sensitivity (d'), there was a main effect of emotion-pairing ($F(5, 15)=4.56$, $p=0.010$). Sensitivity was highest for the fear and disgust comparison ($M=3.13$, $SE=0.22$) and lowest for the anger and fear comparison ($M=2.11$, $SE=0.36$) but pairwise comparisons revealed there were no significant differences in sensitivity between the six emotion pairings. There was also a main effect of intensity ($F(3, 9)=27.88$, $p<0.001$). Sensitivity improved linearly from the 10–30% to the 54–74% intensity level (see Fig. 3) and reached asymptote at this point with pairwise comparisons revealing significant differences between the 10–30% intensity range and the three other intensity ranges (32–52%, $p=0.026$; 54–74%, $p=0.027$; and 76–98%, $p=0.033$), but not amongst any other comparisons ($ps>0.517$). The interaction of category and intensity for sensitivity was not significant ($F(15, 45)=1.10$, $p=0.39$). There were no significant effects of bias ($Fs<0.02$, $ps>0.996$).

PHD was compared to controls using Crawford and Howell's (1998) method which uses the control sample statistics as statistics in the test rather than as parameters, which is more appropriate when the control sample is modest in size. Bonferroni correction

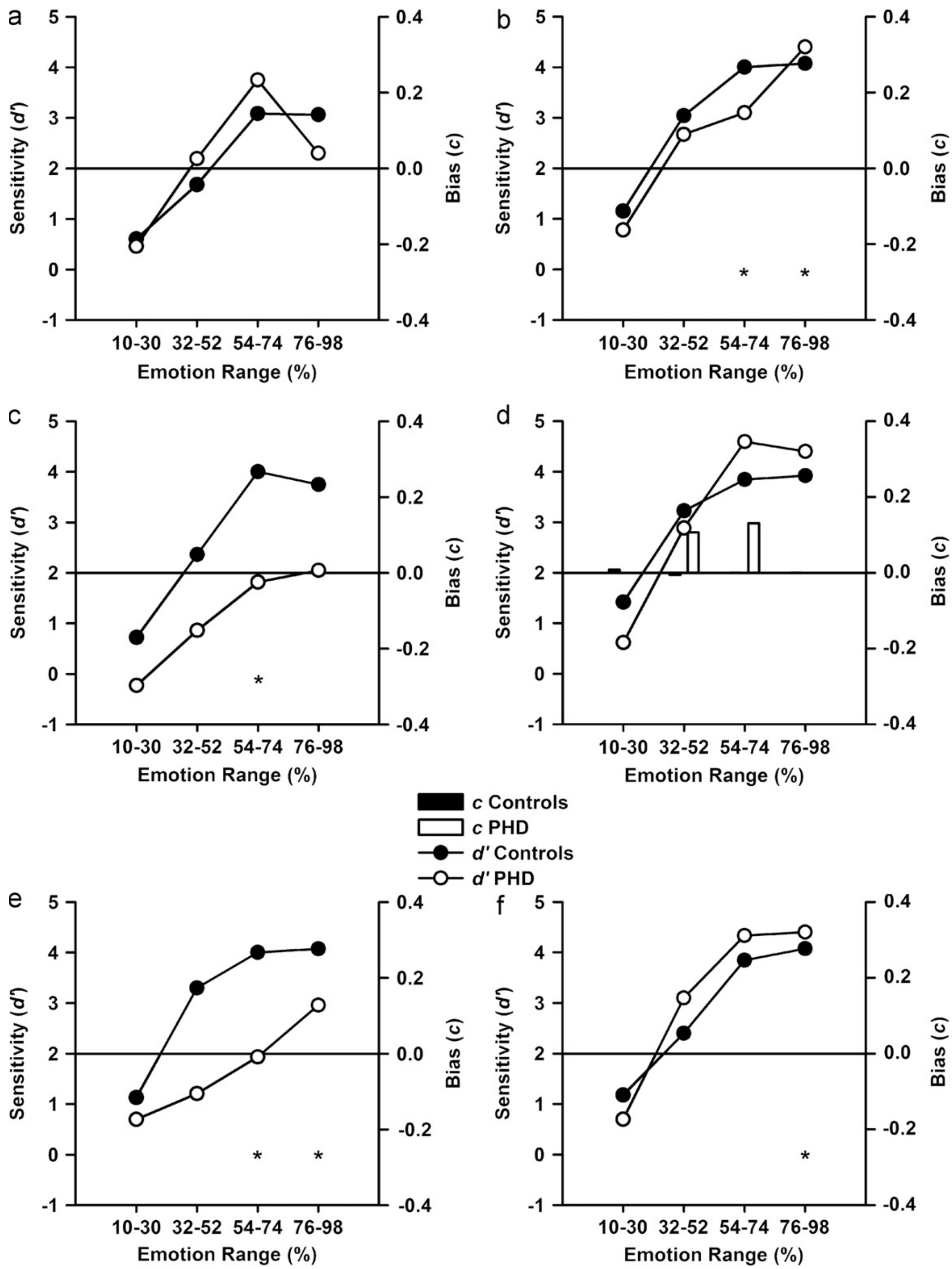

Fig. 3. Sensitivity (d') and bias (c) data for PHD and controls in all pair-wise categorisations in Experiment 1 with * to indicate significance at $p=0.01$ (two-tailed). (a) anger and fear (b) anger and happiness (c) anger and disgust (d) fear and happiness (e) fear and disgust (f) happiness and disgust. Note, bias values are very small and therefore appear absent in most figures (indicating no bias).

was used to correct for multiple comparisons. Significant differences at $p=0.01$ are indicated in Fig. 3. These demonstrated that PHD performed as controls except when discriminating fearful from disgusted faces in the 54–74% and 76–98% ranges ($t(3)=-184.88$, $p<0.001$; $t(3)=-111.50$, $p<0.001$), and angry from disgusted faces ($t(3)=-195.43$, $p<0.001$) and happy from angry faces ($t(3)=-80.77$, $p<0.001$) in the 54–74% range where his d' was significantly lower than for controls. Also for happy from angry ($t(3)=29.16$, $p<0.001$) and happy from disgust faces ($t(3)=29.16$, $p<0.001$) in the 76–98% range when his d' was significantly higher than for controls. Note the values for c in Fig. 3 are very small (indicating no bias).

2.3. Discussion

The clinical data showed PHD to be poor at categorizing emotions when recognising those emotions is facilitated by configural processing (i.e., fear and sadness, and in some respects angry if misidentification is considered). PHD's impoverished categorisation of these particular emotions may reflect an absence of configural processing. However, these clinical data are difficult to interpret. First, it is possible that performance reflects a difficulty using such information in the context of a six-alternative forced choice decision. Second, the categorisations may be subject to criterion shifts, such as over-estimating the relative frequency of some emotions, e.g., angry faces.

To overcome these issues we tested PHD in conditions where the alternatives were limited to two and sensitivity could be determined independently of bias. These data show that, given sufficient intensity, discriminating between pairs of emotions is rather easy for control participants. More importantly, the data show that PHD can make pair-wise categorisations of facial emotions, especially if one of the face categories is happiness. However, PHD is markedly less sensitive than controls when categorising anger from disgust and fear from disgust, especially at high levels of stimulus intensity.

By themselves, these data can be interpreted in a number of different ways. We explored them further in a follow-up study: PHD and controls performed a psychophysical intensity discrimination threshold task using the same face set as in Experiment 1. We sought to establish the magnitude of the difference in intensity between two faces of the same emotion that was required before PHD and controls could reliably report faces of the higher intensity. We reasoned that uncertainty in relation to faces of specific emotions would translate into high discrimination thresholds (a higher percentage difference in emotion required to make the discrimination). Using a one-up, two-down threshold paradigm, we measured the point at which face intensity could be reliably discriminated on 71% of occasions. In this staircase paradigm, correct discrimination leads to a reduced difference in emotion intensity between the stimuli on the next trial, whilst incorrect discrimination leads to increased difference. The threshold for reliably discriminating between pairs of simultaneously presented angry, happy, fearful and disgusted faces was measured. Again, Crawford and Howell's (1998) method for comparing a single participant to a small control group with Bonferroni correction was used. The data show PHD to be significantly poorer than controls at within-emotion intensity judgements of anger ($t(3)=9.22$, $p=0.001$) and disgust ($t(3)=8.934$, $p=0.001$), but not of happiness ($t(3)=2.32$, $p=0.103$) or fear ($t(3)=1.77$, $p=0.174$, see Fig. 4).

Together, the data from Experiment 1 suggest PHD to have difficulties in the perception of anger and disgust, some difficulty in the categorisation of anger from fear and no difficulty in the perception or categorisation of happiness. This pattern of results is consistent with a loss of configural information, where some emotions will be more affected than others (McKelvie, 1995; Calder et al., 2000).

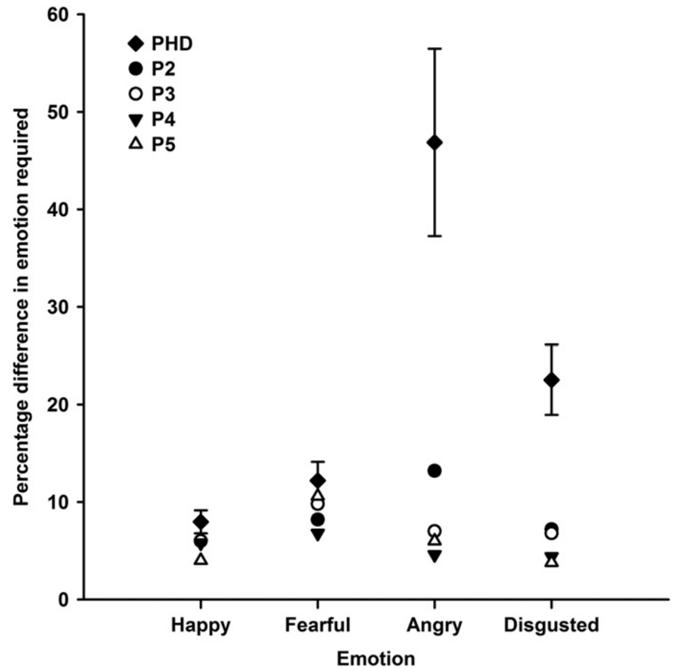

Fig. 4. Percentage difference in emotion content required to discriminate magnitude of emotion for PHD and each control participant (P2-P5). Values for PHD represent the mean difference across four repeats and therefore, include standard error bars of the mean.

3. Experiment 2

In Experiment 2a, we explored categorisation of Thatcherised from typical faces in PHD and controls. Using a task similar to that used in Experiment 1, participants were asked to determine if individually presented facial stimuli were Thatcherised or typical faces. Given the salience of the Thatcher illusion for typical participants we anticipated that controls would perform as if discriminating between highly salient emotions. However, this assumption needs to be stated more formally. We assume the d′s of Experiment 1 reflect the perceived difference in emotional valence of face categories. We can match the d's of controls when categorising Thatcherised from typical faces with those from Experiment 1 when categorising emotional faces. The critical question is how PHD performs in the condition that has been matched for valence. In particular, can his performance in Experiment 2a be predicted from his performance in Experiment 1?

We also ran all participants in an additional study (Experiment 2b). In Experiment 2b, participants made 'same' and 'different' decisions to pairs of faces composed of orthogonal combinations of Thatcherised and typical faces shown in Experiment 2a.

In addition to testing participants in Experiments 2a and 2b with faces, we also tested participants using images of churches that were manipulated ("Thatcherised") in the same way as faces.

3.1. Method

3.1.1. Participants

PHD and eight left handed male controls were tested. Two controls were age-matched (aged 54 and 56) and six were students (age range 20–29, $M=23.17$ years, $SD=3.66$). All participants had normal or corrected-to-normal vision and were asked to complete the Edinburgh Handedness Inventory (Oldfield, 1971) prior to participation, all were left-handed. No controls had any history of neurological problems.

3.1.2. Stimuli

Ten grey-scale face stimuli from the NimStim face set (Tottenham et al., 2009) were used to create Thatcherised stimuli by inverting the eyes and mouths. Grayscale images of churches were used as control stimuli and were manipulated in a similar way to faces by inverting the windows and the door (see Fig. 5). Individual face stimuli appeared at a size of 10 cm × 13 cm and individual church stimuli at a width of 8.30 cm, although the height varied between 7.59 cm and 16.98 cm. Therefore, the

Fig. 5. Examples for face and church stimuli shown in Experiment 2a (categorisation task) and Experiment 2b (discrimination task).

visual angle was 7.63 by 9.91 degrees for faces, and 6.33 by between 5.79 and 12.92 degrees for churches, when viewed on a desktop computer screen from a distance of 75 cm.

3.1.3. Categorisation task (Experiment 2a)

Participants were instructed to decide if the stimulus (either a church or a face) was 'typical' or 'odd'. Odd was defined by explaining how they had been changed to look grotesque and by showing examples of Thatcherised faces versus typical faces (and the equivalent versions for churches). Stimuli were presented centrally until a response was made. The experimental design incorporated 10 individual faces, each in a Thatcherised and typical form. Upright and inverted versions of each were repeated four times, creating 160 trials. The same design was used for churches, creating 160 church trials. The order of stimulus type was counterbalanced across participants. Controls completed the task once and PHD completed the task three times over a period of several months.

3.1.4. Discrimination task (Experiment 2b)

Participants were instructed to decide whether pairs of stimuli were 'the same' or 'different'. Stimuli were displayed until a response was made by pressing one of two designated mouse buttons. The stimuli from the categorisation task were combined into pairs and shown centrally, separated by a 2 cm gap for simultaneous comparison. Equal numbers of matching and mismatching pairs were created using the same face/church identity and orientation, with the only difference in a mismatched pair being one stimulus was Thatcherised and one was typical. Each stimulus appeared eight times in equal numbers of upright and inverted presentations. There were a total of 160 face comparison trials and 160 church comparison trials. Trials were blocked by orientation and object type, and block order was counterbalanced across participants. Controls completed the task once and PHD completed the task three times over a period of several months.

3.2. Results

As in Experiment 1, all data were analysed in terms of signal detection measures, sensitivity (d') and bias (c). The age-matched controls were always within the range of student controls and therefore the data from all control participants was combined. The results were analysed using 2 (Stimulus: Faces versus churches) $\times$ 2 (Orientation: Upright versus inverted) repeated measures ANOVAs. Separate analyses were computed for the categorisation and discrimination tasks as well as for d' and c.

The main effect of orientation was significant for d' sensitivity in both the categorisation and discrimination tasks ($F(1, 7)=57.64$, $p<0.001$; $F(1, 7)=11.02$, $p=0.013$, respectively). Participants were more sensitive to upright ($M=3.03$, $SE=0.14$; $M=4.24$, $SE=0.26$) than inverted stimuli ($M=2.34$, $SE=0.13$; $M=3.80$, $SE=0.26$). The main effect of stimulus type was significant in the categorisation task ($F(1,7)=6.34$, $p=0.040$) with sensitivity higher to churches ($M=2.94$, $SE=0.17$) than to faces ($M=2.43$, $SE=0.16$). There was no significant effect of stimulus type in the discrimination task ($F(1, 7)=3.20$, $p=0.117$).

Orientation and stimulus type yielded a significant interaction in the categorisation task thus replicating many previous studies of the Thatcher illusion ($F(1, 7)=13.55$, $p=0.008$). Inversion reduced sensitivity to faces in the categorisation task (upright, $M=3.15$, $SE=0.16$; inverted, $M=1.72$, $SE=0.25$), but there was no difference for churches (upright, $M=2.92$, $SE=0.15$; inverted, $M=2.95$, $SE=0.20$). There was no significant interaction between orientation and stimulus type in the discrimination task ($F(1, 7)=4.41$ $p=0.074$).

With respect to the bias data, there was a main effect of stimulus type on the categorisation task ($F(1, 7)=7.08$, $p=0.032$), with bias towards the 'odd' response being greater for faces ($M=-0.27$, $SE=0.09$) than for churches ($M=-0.02$, $SE=0.09$). There was no main effect of stimulus type on bias in the discrimination task ($F(1, 7)=0.00$, $p=0.999$). The main effect of orientation was not significant in either the categorisation or the discrimination tasks ($F(1, 7)=4.09$, $p=0.08$; $F(1, 7)=0.82$, $p=0.396$). However, there was an interaction between stimulus type and orientation in both the categorisation and discrimination tasks ($F(1, 7)=16.30$, $p=0.005$; $F(1, 7)=10.03$, $p=0.016$, respectively). Responses to faces were more biased towards 'odd' and 'different' when inverted ($M=-0.60$, $SE=0.16$) than when upright ($M=0.07$, $SE=0.08$), whereas the reverse was true with churches (upright, $M=-0.08$, $SE=0.09$; inverted, $M=0.04$, $SE=0.14$).

PHD was compared against chance ($d'=0$) and the pooled control groups using one sample t-tests with Bonferroni correction. PHD was above chance in all conditions of the discrimination task ($ts(2)>9.76$, $ps<0.010$). In the categorisation task he was above chance in the upright ($t(2)=15.05$, $p=0.004$) and inverted ($t(2)=5.96$, $p=0.027$) church conditions but he was completely unable to perform the face task scoring at chance for both the upright ($t(2)=0.20$, $p=0.858$) and inverted ($t(2)=-0.37$, $p=0.746$) face conditions (see Fig. 6). When comparing PHD's sensitivity to controls (Crawford & Howell, 1998, with Bonferroni correction), there were no significant differences in any condition of the discrimination task (magnitude of $t(7)s<1.42$, $ps>0.198$) or the upright and inverted church conditions of the categorisation task ($t(7)=-1.03$, $p=0.337$; $t(7)=-1.97$, $p=0.090$). However, PHD's inability to categorise faces was evident, with controls significantly more sensitive than PHD in the upright face condition of the categorisation task ($t(7)=-6.38$, $p<0.001$), and marginally more significant in the inverted condition ($t(7)=-2.28$, $p=0.056$).

3.3. Discussion

The categorisation task of Experiment 2a revealed that controls showed, as expected, good sensitivity to upright and inverted faces and churches. The sensitivity of controls to upright faces was similar to the levels achieved when categorising highly intense emotions in Experiment 1. Although equally sensitive when categorising upright faces and churches, controls were more sensitive when categorising inverted churches than inverted faces.

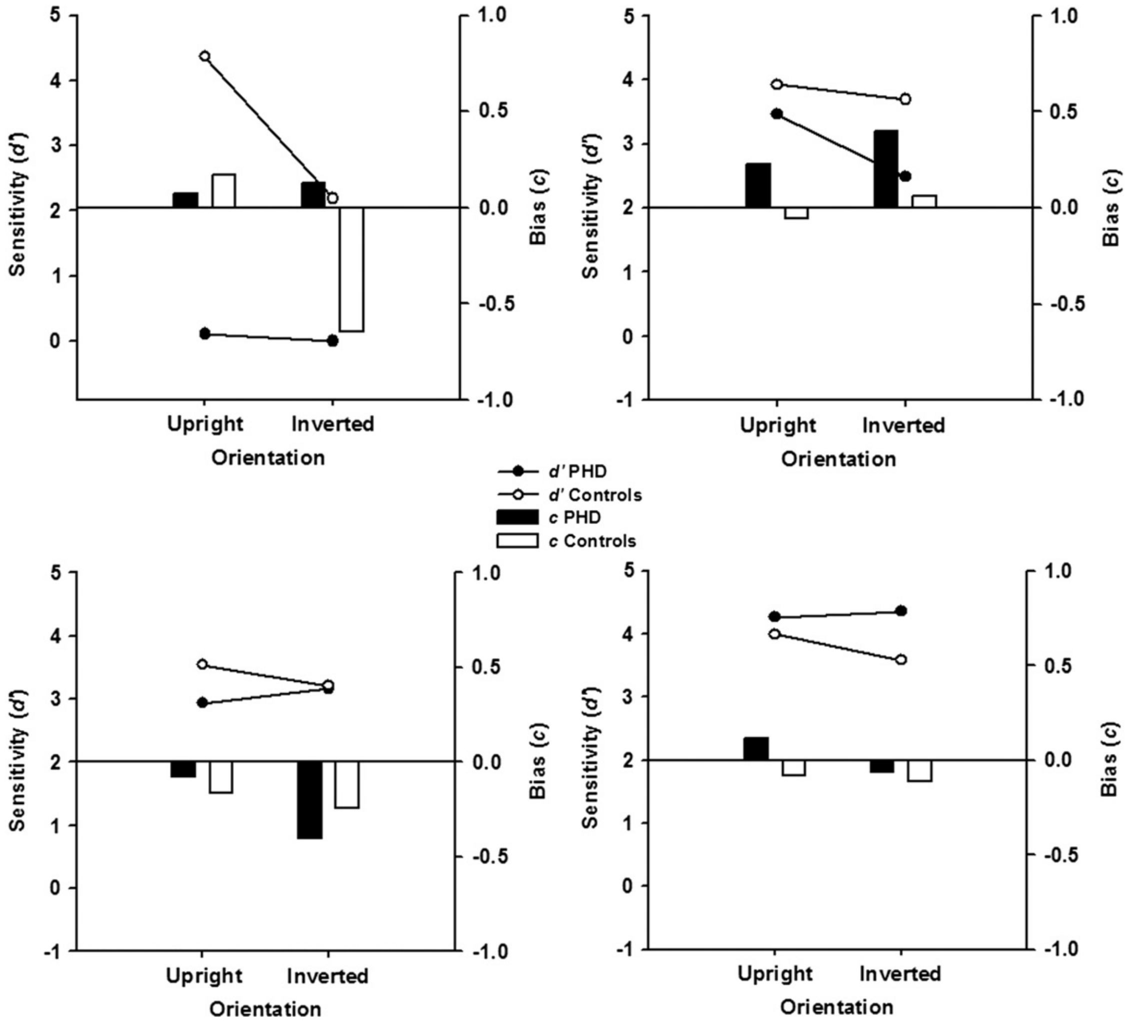

Fig. 6. Sensitivity (d') and bias (c) data for face (right) and church (left) conditions in Experiment 2a (top) and Experiment 2b (bottom).

In contrast, PHD was at chance categorising both upright and inverted faces but not churches, where his performance was above chance and no different to controls. It might have been that PHD was unable to perceive the stimulus alterations that differentiated Thatcherised from typical faces. However, the simultaneous discrimination task of Experiment 2b shows this to not be the case. PHD was sensitive to differences in both faces and churches, when both upright and inverted.

Comparing across discrimination and categorisation tasks, the key contrast is that PHD cannot categorise Thatcherised from typical faces, despite being able to discriminate the very same faces when shown them simultaneously.

4. General discussion

Controls were sensitive to categorising facial emotions, discriminating intensity of emotional expressions, and categorising and discriminating Thatcherised from typical faces. They also demonstrated the expected inversion effect for faces when categorising Thatcherised from typical faces.

PHD's results differ from controls' in important ways. First, while PHD can perform pair-wise categorisations of emotional expressions, providing emotions are sufficiently salient, he is poor at judging relative intensity within some categories (anger and disgust). Second, PHD is at chance categorising Thatcherised from typical faces, despite being able to discriminate differences between these faces and being able to categorise churches.

We have hypothesized that this conjunction of findings is consistent with an absence of configural face processing. There are, however, two competing hypotheses. First, that Thatcherised faces should be thought variants of angry faces (as suggested by a reviewer) as PHD was impaired with angry faces in Experiment 1, and he fails with Thatcherised faces in Experiment 2a. Second, that the Thatcherised faces shown in Experiment 2 were less intense in their perceived emotion than the emotional faces shown in Experiment 1: if so, it might be that PHD's failure in Experiment 2 reflects a mere intensity effect.

With respect to the hypothesis that Thatcherised faces are closely matched to angry faces, we asked PHD to categorise angry, neutral and Thatcherised faces with a view to exploring the resulting confusion matrix. PHD was not very good at this task (78.13% correct on neutral trials, 58.13% on angry face trials and 62.50% correct on Thatcherised trials). More importantly, he mistook Thatcherised faces as neutral faces (24.38%) more often than as angry faces (13.13%). We conclude that PHD does not see Thatcherised faces as angry faces and that his difficulties perceiving both classes of face, although related, are different.

With respect to the hypothesis that our Thatcherised faces shown in Experiment 2 were less intense than the emotional faces shown in Experiment 1, we can compare the salience of faces across Experiments 1 and 2 directly. For controls, the sensitivity (d's) in Experiment 2 matched that of high and high-medium

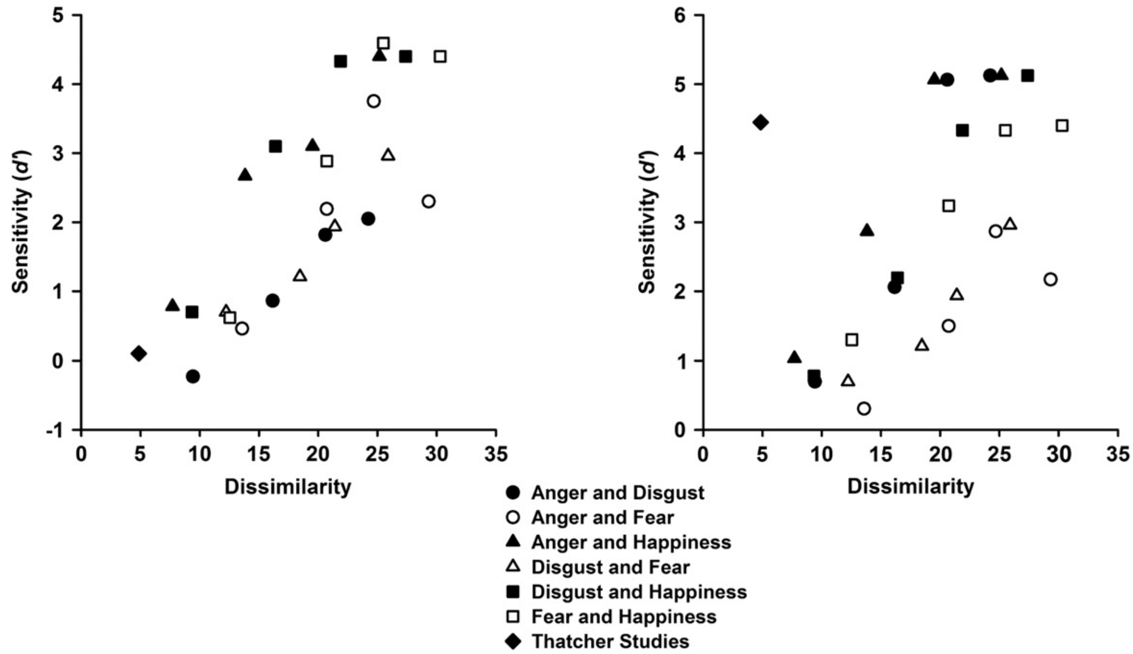

Fig. 7. Graphs showing the regression of d' against image similarity for PHD (left) and controls (right), using the d' values from Experiment 1 and the upright face conditions in experiment 2a.

emotion conditions in Experiment 1. PHD's failure to categorise Thatcherised faces from typical faces cannot, therefore, result from Thatcherised faces being of low intensity.

We conclude that these findings are consistent with PHD being able to use facial features to map some emotional faces into certain emotion categories. However, doing this requires being able to compute the similarity of actual facial features to those that define each category. At the limit, featural similarity can be estimated in terms of the similarity of pixels across images in a way that is not true for similarity determined by configural relations. If PHD categorises faces (Thatcherised and emotional) with reference to simple features then featural similarity between the sets of images forming the categories should predict his behavior in a way not true for controls.

We tested this idea by computing the similarity between pairs of sets of images used in the categorisation tasks of Experiments 1 and 2. The sets of images were determined by the sets over which sensitivity was calculated. For the emotion task, for example, sensitivity was calculated between the set of low intensity (10% to 30%) happy faces and the set of low intensity angry faces. There were 16 sets altogether (four emotion ranges for four different emotions). For the Thatcher task, sensitivity was computed between the set of typical faces and the set of Thatcherised faces, providing two sets of images (typical and Thatcherised). For each set of images, an average image was created by taking the mean of RGB values for each pixel location across all images in the set of faces. This average image formed a single representation of all images in the corresponding set. The difference between a pair of average images was then computed by taking the difference in RGB values at each pixel location, summing the squares of these differences, and taking the square root of the sum to provide a measure of dissimilarity between the pair. These dissimilarity measures therefore represented the overall difference between two sets of images. These dissimilarity data were then regressed against the mean sensitivity data for all pair-wise categorisation decisions.

The results show that PHD's sensitivity, across all categorisation conditions of Experiments 1 and 2, is predicted by this simple feature similarity model (see Fig. 7). PHD was more sensitive to categories when the images drawn from each category were physically very different. His sensitivity reduces as the difference between images reduces. The regression model was significant, $F(1, 23)=42.22$, $p<0.001$. Similarity is a significant predictor of sensitivity ($\beta_1=0.17$, $t(23)=6.50$, $p<0.001$, $R^2=0.65$).

In contrast, whilst the regression model is significant for controls ($F(1, 23)=12.77$, $p=0.002$ with similarity as a significant predictor of sensitivity, $\beta_1=0.15$, $t(23)=3.57$, $p=0.002$, $R^2=0.38$), the data emphatically show that control participant's responses to Thatcherised faces are not based on simple feature similarity, as is the case for PHD (see Fig. 7). Controls have an exquisite sensitivity to the manipulations used to create Thatcherised faces.

These data suggest that, for controls, the categorization and discrimination of Thatcherised from typical faces results from the computation of some configural feature around eyes and mouths, or from the difficulty of attempting to compute such features. These two classes of explanation were raised in the Introduction. Here we make one further point. The current data cannot determine which of these accounts is correct. Other papers have explored the nature of configurality in the Thatcher illusion (Donnelly, Cornes, & Menneer, 2012; Mestry, Menneer, Wenger, & Donnelly, in press). Moreover, the results suggest an absence of configural processing in upright Thatcherised faces (Mestry et al., in press). These data would suggest, therefore, that it is PHD's absence of configural processing in typical faces, alongside the general absence of configural processing of Thatcherised faces, that leaves him unable to categorise Thatcherised from typical faces.

This study has shown that discrimination of Thatcherised and emotional faces is possible in a patient who lacks the N170. However, the ability to categorise Thatcherised faces is lost. A deficit in configural processing impacts PHD's categorisation of Thatcherised faces and to some extent his emotion processing while his preserved feature processing supports his ability to categorise emotional faces and his ability to discriminate between Thatcherised and typical faces. These results support the role of configural processing in both perception of the Thatcher illusion and typical emotion processing, but cannot speak to sources of configurality being present in the upright Thatcherised face. The Thatcher illusion is considered to be a face specific illusion where configural processing is dependent on the automatic categorisation of faces (see Boutsen, Humphreys, Praamstra, & Warbrick, 2006; Milivojevic, Clapp, Johnson, & Croballis, 2003; Carbon, Schweinberger, Kaufman, & Leder, 2005 for explorations of the N170 in relation to Thatcher illusion). PHD does not show an N170 face effect and does not experience the oddity of Thatcherised faces. Therefore, we suggest that the neural substrate for automatic face categorisation, indexed by the N170, is a prerequisite for the experience of discomfort and oddity that characterises neurotypical individual's encounters with "Thatcherised" faces.

Role of the funding source

The research and the preparation of this manuscript were supported by a Ph.D. studentship grant from the Economic and Social Research Council (ESRC) awarded to N. Mestry.

Acknowledgments

Thank you to Dr. Matt Garner for his assistance in sourcing the morphed emotion stimuli, to Dr. Andy Young for providing the stimuli for the emotion assessment, to Dr. Beatrice de Gelder for providing the body postures stimuli and to Dr. Matthew Brett and Dr. Gregory DiGirolamo who commissioned and performed the scans presented in Fig. 1.

References

Bartlett, J. C., & Searcy, J. (1993). Inversion and configuration of faces. *Cognitive Psychology*, *25*, 281–316.

Benton, A. L., Sivan, A. B., Hamsher, K. deS., Varney, N. R., & Spreen, O. (1994). *Contributions to neuropsychological assessment* (2nd ed.). New York: Oxford University Press.

Boutsen, L., Humphreys, G. W., Praamstra, P., & Warbrick, T. (2006). Comparing neural correlates of configural processing in faces and objects: an ERP study of the Thatcher illusion. *NeuroImage*, *32*, 352–367.

Calder, A. J., Young, A. W., Keane, J., & Dean, M. (2000). Configural information in facial perception. *Journal of Experimental Psychology*, *26*, 527–551.

Carbon, C. C., Grüter, T., Weber, J. E., & Lueschow, A. (2007). Faces as objects of non-expertise: processing of Thatcherised faces in congenital prosopagnosia. *Perception*, *36*, 1635–1645.

Carbon, C. C., Schweinberger, S. R., Kaufman, J. M., & Leder, H. (2005). The Thatcher illusion seen by the brain: an event-related brain potential study. *Cognitive Brain Research*, *24*, 544–555.

Crawford, J. R., & Howell, D. C. (1998). Comparing an individual's test score against norms derived from small samples. *The Clinical Neuropsychologist*, *12*, 482–486.

De Gelder, B., & Van den Stock, J. (2011). The bodily expressive action stimulus test (BEAST). Construction and validation of a stimulus basis for measuring perception of whole body expression of emotions. *Frontiers in Psychology*, *2*, 181.

Donnelly, N., Cornes, K., & Menneer, T. (2012). An examination of the processing capacity of features in the Thatcher illusion. *Attention, Perception & Psychophysics*, *74*, 1475–1487.

Donnelly, N., Zürcher, N. R., Cornes, K., Snyder, J., Naik, P., Hadwin, J. A., et al. (2011). Discriminating grotesque from typical faces: evidence from the Thatcher illusion. *PLoS One*, *6*, e23340.

Eimer, M. (2000). Effects of face inversion on the structural encoding and recognition of faces: evidence from event-related brain potentials. *Cognitive Brain Research*, *10*, 145–158.

Eimer, M., & McCarthy, R. A. (1999). Prosopagnosia and structural encoding of faces: evidence from event-related potentials. *NeuroReport*, *10*, 255–259.

Ekman, P., & Friesen, W. V. (1976). *Pictures of facial affect*. Palo Alto, CA: Consulting Psychology Press.

McDonald, S., Flanagan, S., Rollins, J., & Kinch, J. (2003). TASIT: A new clinical tool for assessing social perception after traumatic brain injury. *Journal of Head Trauma Rehabilitation*, *18*, 219–238.

McKelvie, S. J. (1995). Emotional expression in upside-down faces: evidence for configurational and componential processing. *British Journal of Social Psychology*, *34*, 325–334.

McKenna, P., & Warrington, E. K. (1983). *Graded naming test*. Windsor, UK: NFER-Nelson.

Mestry, N., Menneer, T., Wenger, M. J., & Donnelly, N. (in press). Identifying sources of configurality in three face processing tasks. *Frontiers in Psychology.*

Milivojevic, B., Clapp, W. C., Johnson, B. W., & Croballis, M. C. (2003). Turn that frown upside down: ERP effects of Thatcherization on misorientated faces. *Psychophysiology*, *40*, 967–978.

Oldfield, R. C. (1971). The assessment and analysis of handedness: the Edinburgh inventory. *Neuropsychologica*, *9*, 97–113.

Parks, T. E., Coss, R. G., & Coss, C. S. (1985). Thatcher and the cheshire cat: context and the processing of facial features. *Perception*, *14*, 747–754.

Reber, R., Winkielman, P., & Schwartz, N. (1998). Effects of perceptual fluency on affective judgements. *Psychological Science*, *9*, 45–48.

Rhodes, G., & Tremewan, T. (1996). Averageness, exaggeration, and facial attractiveness. *Psychological Science*, *7*, 105–110.

Rothstein, P., Malach, R., Hadar, U., Graif, M., & Hendler, T. (2001). Feeling of features: different sensitivity to emotion in high-order visual cortex and amygdala. *Neuron*, *32*, 747–757.

Rouse, H., Donnelly, N., Hadwin, J. A., & Brown, T. (2004). Do children with autism perceive second-order relational features? The case of the Thatcher illusion. *Journal of Child Psychology and Psychiatry*, *45*, 1246–1257.

Stürzel, F., & Spillmann, L. (2000). Thatcher illusion: dependence on angle of rotation. *Perception*, *29*, 937–942.

Thompson, P. (1980). Margaret Thatcher: a new illusion. *Perception*, *9*, 483–484.

Tottenham, N., Tanaka, J. W., Leon, A. C., McCarry, T., Nurse, M., Hare, T. A., et al. (2009). The NimStim set of facial expressions: judgments from untrained research participants. *Psychiatry Research*, *168*, 242–249.

Valentine, T. (1988). Upside-down faces: a review of the effect of inversion upon face recognition. *British Journal of Psychology*, *79*, 471–491.

Warrington, E. K. (1984). *Recognition memory tests*. Berkshire, UK: NFER-Nelson.

Warrington, E. K. (1996). *Camden Memory Tests*. Hove, UK: Psychology Press.

Warrington, E. K., & James, M. (1991). *The visual object and spacial perception battery*. Burry St. Edmunds, UK: Thames Valley Test Co.

Wechsler, D. (1997). *Wechsler adult intelligence scale—3rd Edition (WAIS-3®)*. San Antonio, TX: Harcourt Assessment.

Young, A. W., Perrett, D. I., Calder, A. J., Sprengelmeyer, R., & Ekman, P. (2002). Facial expressions of emotion: stimuli and tests (FEEST). (Bury St Edmunds Thames Valley Test Company, (Eds.) *Psychology, 126*, 420.