Accepted by Cortex on 8th April 2016

Impaired perception of facial emotion in developmental prosopagnosia

Federica Biotti* and Richard Cook

Department of Psychology, City University London

*Corresponding author: Federica.Biotti@city.ac.uk

Department of Psychology
City University London
Whiskin Street
London, U.K.
EC1R 0JD

Abstract

Developmental prosopagnosia is a neurodevelopmental condition characterised by difficulties recognising faces. Despite severe difficulties recognising facial identity, expression recognition is typically thought to be intact in developmental prosopagnosia; case studies have described individuals who are able to correctly label photographic displays of facial emotion, and no group differences have been reported. This pattern of deficits suggests a locus of impairment relatively late in the face processing stream, after the divergence of expression and identity analysis pathways. To date, however, there has been little attempt to investigate emotion recognition systematically in a large sample of developmental prosopagnosics using sensitive tests. In the present study, we describe three complementary experiments that examine emotion recognition in a sample of 17 developmental prosopagnosics. In Experiment 1, we investigated observers' ability to make binary classifications of whole-face expression stimuli drawn from morph continua. In Experiment 2, observers judged facial emotion using only the eye-region (the rest of the face was occluded). Analyses of both experiments revealed diminished ability to classify facial expressions in our sample of developmental prosopagnosics, relative to typical observers. Imprecise expression categorisation was particularly evident in those individuals exhibiting apperceptive profiles, associated with problems encoding facial shape accurately. Having split the sample of prosopagnosics into apperceptive and non-apperceptive subgroups, only the apperceptive prosopagnosics were impaired relative to typical observers. In our third experiment, we examined the ability of observers' to classify the emotion present within segments of vocal affect. Despite difficulties judging facial emotion, the prosopagnosics exhibited excellent recognition of vocal affect. Contrary to the prevailing view, our results suggest that many prosopagnosics do experience difficulties classifying expressions, particularly those with apperceptive profiles. These individuals may have difficulties forming view-invariant structural descriptions at an early stage in the face processing stream, before identity and expression pathways diverge.

Keywords: Developmental Prosopagnosia; expression recognition; facial emotion; vocal emotion; face perception.

1. Introduction

Developmental prosopagnosia[1] (DP) is a lifelong neurodevelopmental disorder associated with impaired face recognition, thought to affect as many as one in every 50 people (Kennerknecht et al., 2006; Kennerknecht, Ho, & Wong, 2008). Individuals with DP exhibit deficits recognising personally familiar faces as well as problems discriminating unfamiliar faces, despite normal intelligence, typical low-level vision, and an absence of manifest brain injury (Behrmann & Avidan, 2005; Duchaine & Nakayama, 2006b; Susilo & Duchaine, 2013). Due to characteristic problems with face recognition, individuals with DP often utilise cues derived from hairstyle, voice, and gait, for person recognition. Nevertheless, recognising familiar people encountered out of context or following changes in external appearance, can prove challenging (Shah, Gaule, Sowden, Bird, & Cook, 2015).

The precise origin of the face recognition deficits seen in DP remains unclear. Cognitive accounts have argued that, relative to typically developing (TD) individuals, DPs exhibit reduced holistic processing of faces – whereby individual features (eyes, nose, mouth) are integrated into a coherent unified whole – compromising the accuracy and efficiency of their face recognition (Avidan, Tanzer, & Behrmann, 2011; Liu & Behrmann, 2014; Palermo et al., 2011). At the neurological level, differences in cortical structure (Behrmann, Avidan, Gao, & Black, 2007; Garrido et al., 2009), structural (Gomez et al., 2015; Thomas et al., 2009) and functional connectivity (Avidan & Behrmann, 2009; Avidan et al., 2013) have been observed in inferotemporal regions including the fusiform gyrus, a region thought to be crucial for face processing (Kanwisher, 2000). Strikingly, DP often runs in families (Duchaine, Germine, & Nakayama, 2007; Johnen et al., 2014; Lee, Duchaine, Wilson, & Nakayama, 2010; Schmalzl, Palermo, & Coltheart, 2008), suggestive of a genetic component.

The characteristic deficits of facial identity recognition seen in DP have attracted substantial research attention (Susilo & Duchaine, 2013). However, there has also been considerable interest in the expression recognition abilities of individuals with DP. The facial expressions of others are a rich source of social information, conveying cues to affective and mental states (Adolphs, 2002; Frith, 2009; Parkinson, 2005). The ability to interpret facial expressions correctly is therefore important for fluent social interaction and wider socio-cognitive development. Moreover, the question of emotion recognition in DP also has important implications for neurocognitive accounts of the condition (Bate & Bennetts, 2015; Kress & Daum, 2003a). Where observed together, difficulties recognising facial identity and facial

emotion are suggestive of apperceptive prosopagnosia (De Renzi, Faglioni, Grossi, & Nichelli, 1991); difficulties may arise early on in the face processing stream, leaving observers unable to form an accurate, view-invariant description of face shape (Bruce & Young, 1986; Haxby, Hoffman, & Gobbini, 2000). Alternatively, intact expression recognition despite impaired recognition of facial identity suggests a locus of impairment relatively late in the face processing stream, after the divergence of expression and identity analysis pathways (Bruce & Young, 1986; Duchaine, Parker, & Nakayama, 2003; Haxby et al., 2000).

Presently, difficulties recognising facial expressions are thought to be relatively uncommon in DP. Palermo et al. (2011) examined the performance of twelve DPs on three emotion recognition tests: *The Ekman 60 Faces Test,* in which participants label 60 greyscale images of prototypical basic emotions (Young, Perrett, Calder, Sprengelmeyer, & Ekman, 2002); *The Emotion Hexagon Test,* in which participants label expressions drawn from morph continua constructed from the six basic emotions[2] (Young et al., 2002); and *The Reading the Mind in the Eyes Test,* in which participants identify subtle social emotions from cues present around the eye region (Baron-Cohen, Wheelwright, Hill, Raste, & Plumb, 2001). Strikingly, the twelve DPs were unimpaired at both the group and single-case level, relative to aged-matched controls, on all three tasks (Palermo et al., 2011). Dobel, Bölte, Aicher & Schweinberger (2007) described intact emotion recognition in six DPs, having administered the Tübingen Affect Battery – a 4 alternative-forced-choice (AFC) emotion labelling task. Similar findings were reported by Humphreys, Avidan and Behrman (2007), having administered *The Emotion Hexagon Test* to three DPs[2], and Lee, Wilson, Duchaine and Nakayama (2010), having tested three DPs using *The Reading the Mind in the Eyes Test* and a 3AFC match-to-sample task. Several further studies of single cases have described intact emotion recognition in DP (Bentin, Degutis, D'Esposito, & Robertson, 2007; Duchaine et al., 2003; Kress & Daum, 2003b; Nunn, Postma, & Pearson, 2001). Moreover, a study of four DPs indicated that they made typical judgements of facial trustworthiness (Todorov & Duchaine, 2008), an inference thought to be mediated by subtle emotion cues.

Nevertheless, many DPs report problems recognising facial expressions in their daily lives (e.g. Lee et al., 2010), and case studies have described individuals with DP, who *do* exhibit deficits of expression recognition (Ariel & Sadeh, 1996; De Haan & Campbell, 1991;

Duchaine, Murray, Turner, White, & Garrido, 2009; Duchaine, Yovel, Butterworth, & Nakayama, 2006; Minnebusch, Suchan, Ramon, & Daum, 2007; Schmalzl et al., 2008). For example, Duchaine et al. (2006) described a 53-year-old male DP, Edward, who exhibited clear expression recognition impairments on *The Reading the Mind in the Eyes Test* and on a 3-AFC match-to-sample task. Similarly, De Haan and Campbell (1991) tested AB, the original case of DP first described by McConachie (1976), and found that as an adult she exhibited problems labelling prototypical basic emotions. Importantly, however, these reports are relatively infrequent (regarded as 'the exception' rather than 'the norm'), and no systematic investigation has found evidence for a group difference.

The present study sought to re-examine the expression recognition abilities of individuals with DP. As discussed above, this question offers critical insight into the locus of the perceptual difficulties seen in this condition. In particular, we sought to test systematically a large sample of DPs using sensitive tests. The ability of different tests to detect emotion recognition deficits varies widely. For example, Edward, the DP described by Duchaine et al. (2006), was substantially impaired on *The Reading the Mind in the Eyes Test* (4.1 standard deviations below the TD mean), but only mildly impaired on *The Emotion Hexagon Test* (1.4 standard deviations below the TD mean). In Experiment 1, we investigated observers' ability to make binary classifications of whole-face expression stimuli drawn from morph continua. In Experiment 2 observers judged facial emotion using only the eye-region (the rest of the face was occluded). In our third experiment, we examined the ability of observers' to classify the emotion present within segments of vocal affect.

2. Neuropsychological testing

A group of 17 (11 females) individuals with DP participated in the study (Table 1). DP participants were recruited through *www.troublewithfaces.org*. All members of the DP sample described lifelong face recognition problems. None of the DPs had a history of brain injury or psychiatric disorder (e.g., Schizophrenia, Autism Spectrum Disorder). Convergent diagnostic evidence for the presence of DP was collected using the Cambridge Face Memory Test (CFMT; Duchaine & Nakayama, 2006a), the Cambridge Face Perception Test (CFPT; Duchaine, Germine, & Nakayama, 2007) and the Twenty-Item Prosopagnosia Index questionnaire (PI20; Shah, Gaule, Sowden et al., 2015). When administered in the upright orientation, performance on the CFMT correlated closely with scores on the CFPT (r = -.73, p < .001) and the PI20 (r = -.82, p < .001). There was also strong correlations between the

PI20 and the CFPT (r = .61, p < .001). The prosopagnosics' scores on the CFMT and CFPT were compared with a comparison group of 35 age- and gender-matched TD controls. All but one of the DPs scored at least two standard deviations below the control mean on the CFMT (the remaining DP participant was 1.77 standard deviations below the TD mean). In addition to the face recognition tests, participants completed the Cambridge Car Memory Test (CCMT; Dennett et al., 2011) and the Cambridge Bicycle Memory Test (CBMT; Dalrymple, Garrido, & Duchaine, 2014) to assess their wider object recognition ability. In addition, the DPs were screened for colour blindness using Ishihara's Tests for Colour-deficiency (Ishihara, 1993).

Table-1

3. Experiment 1

Measuring individual differences in expression recognition ability is not straight-forward. In particular, tasks that require participants to label prototypical emotional expressions (e.g., happy, sadness, fear, disgust, anger, surprise) may lack sensitivity due to ceiling effects or noise introduced by differences in guessing base-rates (Ipser & Cook, 2015). In our first experiment we sought to determine whether DPs are impaired at making binary categorisations of whole-face emotional expression stimuli drawn from morph continua. Psychophysical modelling of categorisation probability yields sensitive and reliable estimates of expression recognition ability. Previous studies suggest that this approach can reveal group effects that may go undetected by simple labelling paradigms (Cook, Brewer, Shah, & Bird, 2013). Should individuals with DP exhibit subtle expression recognition deficits, we reasoned that a psychophysical approach may be most likely to reveal these problems.

3.1 Methods

The performance of the DPs was compared with a group of 23 TD controls (6 males; M_{age} = 42.65, SD_{age} = 13.44). All TD participants were screened for DP (Table 2). All participants had normal or corrected-to-normal visual acuity. Ethical clearance was granted by the local ethics committee and the study was conducted in line with the ethical guidelines laid down in the 6th (2008) Declaration of Helsinki. All participants gave informed consent.

Table-2

Three morph continua (happiness-anger, disgust-sadness, fear-surprise) were produced by blending incrementally two greyscale photographs of emotional facial expressions, produced by a single actor, selected from Ekman and Friesen's (1975) *Pictures of Facial Affect*. Image morphing was performed using Morpheus Photo Morpher Version 3.11 (Morpheus Software, Indianapolis, IN). Each continuum consisted of seven stimuli which varied in emotion intensity between 20% and 80% in equidistant 10% increments. Stimuli were cropped to exclude external features (e.g., ears, hairline) and presented in greyscale (Figure 1a).

Participants completed a computer-based task written in MATLAB (The MathWorks, Natick, MA) using Psychophysics Toolbox (Brainard, 1997; Pelli, 1997). Experimental trials presented a single image centrally for 1200 ms. Each stimulus subtended approximately 6.5° × 4.0° of visual angle when viewed at 60 cm. Following stimulus offset, participants were asked to make a binary categorisation about the stimulus image (e.g., happiness or anger?). Each of the 21 expression stimuli (3 continua × 7 levels of morph intensity) was presented 20 times in a randomised order. Participants completed 6 practice trials before starting the experimental task. No feedback was provided during the practice or experimental procedures. In total, the procedure consisted of 420 trials and took approximately 20-25 minutes to complete.

Participants' responses were modelled by fitting cumulative Gaussian functions to estimate separate psychometric functions for the three continua. Function fitting was carried out in MATLAB using the Palamedes Toolbox (Prins & Kingdom, 2009). The key parameter of interest, inferred from the psychometric function, was the estimate of categorisation threshold. The threshold estimate is a measure of the precision with which stimuli are categorized and was defined as the standard deviation of the symmetric Gaussian distribution underlying each cumulative Gaussian function (subject to a log transform to attenuate positive skewing). Threshold estimates are inversely related to the slope of the psychometric function; steep and shallow slopes are associated with low and high threshold estimates, respectively. Lower threshold estimates indicate that observers can perceive subtle differences in stimulus strength and vary their responses accordingly. Higher threshold estimates reveal that participants' responses are relatively invariant to physical changes in stimulus strength, indicative of imprecise categorization.

Figure-1

3.2 Results and discussion

The threshold estimates obtained for the DP and TD groups are shown in Figure 1b. Threshold estimates were analysed using ANOVA with Continuum (happiness-anger, disgust-sadness, fear-surprise) as a within-subjects factor and Group (TD, DP) as a between-subjects factor. The analysis revealed a main effect of Continuum [$F(1.46, 55.35) = 46.68$, $p < .001$, $\eta_p^2 = .55$]. Contrasts indicated that fear-surprise categorisations were associated with higher thresholds ($M = 2.81$, $SD = .58$) than happiness-anger ($M = 1.59$, $SD = .87$) [$t(39) = 8.15$, $p < .001$] and disgust-sadness ($M = 2.33$, $SD = .48$) categorisations [$t(39) = 5.74$, $p < .001$]. Disgust-sadness categorisations were also associated with higher thresholds than happiness-anger categorisations [$t(39) = 5.25$, $p < .001$]. Crucially, the analysis also revealed a main effect of Group [$F(1,38) = 4.19$, $p = .04$, $\eta_p^2 = .10$]. Collapsing across the three continua, the DPs exhibited higher thresholds ($M = 7.26$, $SD = 1.54$) than the TD controls ($M = 6.36$, $SD = 1.24$). No Continuum × Group interaction was observed [$F(1.46, 55.35) = 1.02$, $p = .33$, $\eta_p^2 = .03$]. However, simple contrasts indicated a significant difference between the groups only in their fear-surprise thresholds, where the thresholds of the DP group ($M = 3.11$, $SD = .69$) were higher than those of the controls ($M = 2.59$, $SD = .37$) [$t(38) = 3.07$, $p = .004$]. Eight of the DPs scored at least one SD below the TD mean, and three (M3, F5, F11) were significantly impaired at single-case level (Figure 2a).

Figure-2

Clear correlations were observed between participants' categorisation thresholds for the fear-surprise continuum and their CFMT ($r = -.57$, $p < .001$) and PI20 ($r = .51$, $p = .001$) scores. Crucially, no significant correlations were observed with the CCMT ($r = .00$, $p = .98$) and the CBMT ($r = -.03$, $p = .88$) scores (Table 3). However, a striking correlation was found between participants' fear-surprise thresholds and their performance on the CFPT ($r = .78$, $p <.001$; Figure 1c). To investigate this relationship further, the DP sample was split into two sub-groups based on their performance on the CFPT. Eight DPs who scored at least two standard deviations below the control mean on the CFPT (Table 1), and the remaining nine DPs, were categorised as apperceptive and non-apperceptive, respectively. Simple contrasts revealed a significant difference in fear-surprise categorisation thresholds between the

apperceptive subgroup (M = 3.54, SD = .73) and TD controls (M = 2.59, SD = .37) [$t(29)$ = 4.8, $p < .001$]. Interestingly, however, the fear-surprise categorisation thresholds of the non-apperceptive subgroup (M = 2.72, SD = .33) did not differ significantly from the TD sample [$t(30)$ = .95, p = .35].

Table-3

The results of Experiment 1 suggest that our emotion categorisation task and the CFPT may tap very similar processes. The CFPT requires participants to rank order test faces according to their resemblance to a target face. Because the test and target faces are presented throughout each trial, the test is thought to measure observers' ability to form perceptual descriptions of faces, under conditions of minimal working memory load. However, because the physical differences between test faces are subtle, the CFPT provides a demanding test of observers' face encoding. Where perceptual descriptions are compromised, observers may be left unable to detect and interpret subtle physical differences between stimuli, resulting in i) poor sorting performance on the CFPT and ii) judgements of expressions that vary less closely with physical stimulus changes.

4. Experiment 2

The results of Experiment 1 suggest that relative to TD controls, individuals with DP are less able to categorise whole-face expression stimuli drawn from continua that morph emotional facial expressions. While our analyses suggest a trend for less precise categorisation overall, difficulties were particularly clear when observers were required to detect the subtle physical differences between stimuli drawn from the fear-surprise continuum. At least two accounts may be advanced to explain the group difference observed in Experiment 1. First, difficulties integrating information from disparate facial regions may prevent observers with DP forming unified perceptual descriptions of facial expressions. Consistent with this possibility, some observers with DP exhibit reduced composite interference for facial expressions (Palermo et al., 2011), suggestive of reduced holistic processing of facial emotion. Second, observers with DP may have a fundamental difficulty encoding the shape of local facial features. For example, cases of acquired prosopagnosia have been described who appear to have particular problems using information from around the eye region to discriminate (Bukach, Le Grand, Kaiser, Bub, & Tanaka, 2008) and recognise (Caldara et al., 2005) facial identities.

Interestingly, problems using cues from the eye-region are thought to be associated with particular problems recognising facial expressions of fear (Adolphs et al., 2005).

In Experiment 2 we sought to distinguish these rival explanations by examining participants' ability to judge facial emotion using cues from the eye-region alone (i.e., a local region), using a variant of the *Reading the Mind in the Eyes Test* (Baron-Cohen et al., 2001). If the impairments observed in Experiment 1 arise from diminished integration of information from disparate facial regions, we reasoned that the DP group should perform typically on a task that does not require whole-face processing. However, if the impairment in emotion recognition is due to difficulties encoding the shape of local features, the group difference should still be evident.

4.1 Methods

The performance of the DPs was compared with a group of 23 TD controls (7 males; M_{age} = 44.26, SD_{age} = 13.59). All TD participants were screened for DP (Table 2). All participants had normal or corrected-to-normal visual acuity. Ethical clearance was granted by the local ethics committee and the study was conducted in line with the ethical guidelines laid down in the 6th (2008) Declaration of Helsinki. All participants gave informed consent.

The original *Reading the Mind in the Eyes Test* requires observers to recognise complex 'social emotions' (e.g., concerned vs. unconcerned, sympathetic vs. unsympathetic), and may therefore tax both mentalizing and perceptual processes. To minimize any mentalizing demands, our novel variant included different exemplars of four commonly encountered facial emotions. Stimuli were constructed from six Caucasian identities (3 females) selected from the Radboud Faces Database (Langner et al., 2010). For each identity, we produced four morph continua by blending images of the actor exhibiting a neutral expression, with images of the same actor expressing happiness, anger, fear and sadness[3]. The expression morphs containing 30%, 50% and 70% of each emotion (corresponding to low, moderate and high intensity) were cropped so that only the eye-region was visible, and presented in greyscale (Figure 3a). The position of the eyes in the resulting 72 images (6 identities × 4 emotions × 3 levels of emotion intensity) was standardised to ensure similar cues were available in each stimulus. Stimulus images subtended approximately 2.5° × 6.5° of visual angle when viewed at 60 cm.

Experimental trials presented a single stimulus centrally for 1200 ms, followed by a prompt to make a 4-AFC response (happiness, anger, fear, or sadness). The 72 stimuli were presented three times each, in a randomised order, yielding a total of 216 trials. The experiment was preceded by 6 practice trials. No feedback was provided during the practice or experimental procedures. The task lasted approximately 20 minutes. The experimental program was written in MATLAB using the Psychophysics Toolbox (Brainard, 1997; Pelli, 1997).

Figure-3

4.2 Results and discussion

The performance (% correct responses) of the DP and TD groups in the three intensity conditions is depicted in Figure 3b. Results were analysed using ANOVA with Intensity (30%, 50%, 70%) as a within-subjects factor and Group (TD, DP) as a between-subjects factor. The analysis revealed a main effect of Intensity [$F(2,74) = 453.95$, $p < .001$, $\eta_p^2 = .92$]. Accuracy scores were lower in the 30% (low intensity) condition ($M = .47$, $SD = .07$) than in the 50% (moderate intensity) ($M = .69$, $SD = .10$) [$t(38) = 18.39$, $p < .001$] and 70% (high intensity) conditions ($M = .80$, $SD = .08$) [$t(38) = 27.20$, $p < .001$]. The 50% condition was also harder than the 70% condition [$t(38) = 10.60$, $p < .001$]. The analysis also revealed a main effect of Group [$F(1,37) = 6.49$, $p = .01$, $\eta_p^2 = .15$], indicating that the DP group ($M = .62$, $SD = .07$) correctly identified fewer emotions than the TD group ($M = .68$, $SD = .06$), when performance was collapsed across emotion intensity. Interestingly, however, Intensity interacted significantly with Group [$F(2,74) = 4.43$, $p = .01$, $\eta_p^2 = .12$]. Simple contrasts indicated that the TD group ($M = .82$, $SD = .05$) outperformed the DP group ($M = .77$, $SD = .09$) in the 70% condition [$t(37) = 2.04$, $p = .04$]. A similar difference was seen between the DP ($M = .64$, $SD = .09$) and TD ($M = .73$, $SD = .08$) groups for the 50% condition [$t(37) = 3.24$, $p = .003$], but not for the 30% condition [$t(37) = 1.14$, $p = .26$]. Eight of the DPs scored at least one SD below the TD mean, and three (M1, M3, F10) were significantly impaired at single-case level (Figure 2b).

Significant correlations were found between participants' overall performance (collapsing across Group and Intensity) and their scores on the PI20 ($r = -.48$, $p = .003$), the CFMT ($r = .48$, $p = .002$) and the CFPT ($r = -.58$, $p < .001$). However, no significant correlations were observed with the CCMT ($r = -.02$, $p = .93$) and the CBMT ($r = .11$, $p = .54$) (Table 4). Once again, the DP sample was split into apperceptive and non-apperceptive sub-groups based on their performance on the CFPT. Simple contrasts revealed a significant difference in emotion

recognition ability of the apperceptive subgroup (M = .60, SD = .09) and TD controls (M = .68, SD = .06) [$t(28) = 2.74$, $p = .01$]. Interestingly, however, the performance of the non-apperceptive subgroup (M = .64, SD = .06) did not differ significantly from the TD sample [$t(29) = 1.48$, $p = .15$]. The inability of the apperceptive DPs to judge facial emotion using cues from the eye-region alone does not appear to be a product of diminished integration of information from the eye and mouth regions (Palermo et al., 2011), or to a strategic failure to use information from the eye region (Adolphs et al., 2005).

Table-4

5. Experiment 3

The results of the first two experiments indicate that DP individuals are less able to categorise ambiguous facial expressions than TD controls. In Experiment 3 we sought to determine whether this affect recognition deficit was specific to faces, or whether these difficulties extend to other domains. Crucially, aberrant limbic functionality may leave observers unable to interpret emotion *per se* (Calder & Young, 2005). For example, individuals with developmental alexithymia – a neurodevelopmental condition associated with problems interpreting emotional experiences and other forms of interoceptive sensation (Bird & Cook, 2013; Brewer, Happe, Cook, & Bird, 2015) – exhibit a range of emotion recognition difficulties, including problems categorizing facial (Cook et al., 2013), vocal (Heaton et al., 2012), and musical affect (Allen, Davis, & Hill, 2012). To determine whether DPs exhibit face-specific emotion recognition difficulties, we examined their ability to recognise vocal affect. Typical performance on this task would suggest that the poor categorisation exhibited by the DP group in the first two experiments is a product of face, not emotion, perception deficits.

5.1 Methods

The performance of the DPs was compared with a group of 22 TD controls (8 males; M_{age} = 42.86, SD_{age} = 12.89). All TD participants were screened for DP (Table 2). All participants had normal or corrected-to-normal hearing. Ethical clearance was granted by the local ethics committee and the study was conducted in line with the ethical guidelines laid down in the 6th (2008) Declaration of Helsinki. All participants gave informed consent. All participants spoke English as first language.

The stimuli employed in Experiment 3 were short (< 3000 ms) audio sequences of British actors (2 males, 2 females) uttering 3-digit numbers ("two-hundred-and-fifty-five" and "five-hundred-and-twenty-eight") with different emotional inflections (happiness, disgust, fear, sadness, anger and surprise). Stimuli were recorded in a soundproof studio. Having cropped the audio files, and removed background noise using Audacity sound-editing software (http://audacity.sourceforge.net/), stimuli were validated in an online rating study. To create exemplars with varying degrees of ambiguity, we sought to manipulate the pitch of the stimuli, a vital component of vocal affect (e.g., Scherer, 1986). Different amounts (0%, 30%, 60%; corresponding to low, moderate and high noise) of jitter – variability in pitch over the course of the sound – were added to the audio tracks using the 'Raspiness' function in *Praat* (Boersma & Weenink, 2015). In total, 144 stimuli were employed (2 exemplars × 6 emotions × 4 actors × 3 levels of degradation).

Experimental trials presented a single audio clip, followed by a prompt to make a 6-AFC response (happiness, anger, disgust, fear, sadness, or surprise). Each stimulus was presented once, in a randomised order, yielding a total of 144 trials. The task lasted approximately 15 minutes. Twelve practice trials (all with 0% jitter) preceded the experimental procedure to help familiarise participants with the actors' voices. No feedback was provided during the practice or experimental procedures. The experimental program was written in MATLAB using the Psychophysics Toolbox (Brainard, 1997; Pelli, 1997).

5.2 Results and discussion

The performance (% correct responses) of the DP and TD groups was analysed using ANOVA with Jitter (0%, 30%, 60%) as a within-subjects factor and Group (TD, DP) as a between-subjects factor. The analysis revealed a main effect of Jitter [$F(2,74) = 13.04$, $p < .001$, $\eta_p^2 = .26$]. As expected, greater pitch degradation was associated with poorer recognition: Accuracy scores were lower in the 60% (high noise) condition ($M = .61$, $SD = .09$) than in the 30% (moderate noise) ($M = .65$, $SD = .10$) [$t(38) = 3.21$, $p = .003$] and 0% (low noise) conditions ($M = .68$, $SD = .09$) [$t(38) = 4.95$, $p < .001$]. The 30% condition was also harder than the 0% condition [$t(38) = 2.29$, $p = .03$]. Crucially, however, we observed no main effect of Group [$F(1,37) = 1.90$, $p = .18$, $\eta_p^2 = .05$], nor a Group × Jitter interaction [$F(2,74) = .99$, $p = .38$, $\eta_p^2 = .03$], indicative of similar recognition accuracy in the TD and DP groups.

These results support the view that the emotion recognition difficulties exhibited by the DP group in the first two experiments are face-specific, and are not indicative of broader emotion processing impairments. The ability of the DP sample to interpret vocal signals accurately accords with anecdotal evidence that DPs often recognise familiar others using their voice (e.g., Cook & Biotti, 2016). We note, however, that recognition of vocal identity and vocal affect are thought to dissociate; for example, cases of developmental phonagnosia have been described who appear to exhibit broadly intact recognition of vocal affect, despite striking difficulties recognising vocal affect (Garrido et al., 2009).

Discussion

Despite severe difficulties recognising facial identity, emotion recognition deficits are thought to be relatively uncommon in DP (Bate & Bennetts, 2015; Humphreys et al., 2007; Palermo et al., 2011). Contrary to this view, however, we find evidence for widespread deficits in this population. In Experiment 1 we tested observers' ability to make binary classifications of whole-face expression stimuli drawn from morph continua. Psychophysical analyses revealed diminished ability to classify morphed facial expressions in our sample of DPs, relative to TD observers. We replicated this group difference in Experiment 2 when observers categorised facial emotion using only the eye-region. In our third experiment, we examined the ability of observers to classify the emotion present within segments of speech. Despite their difficulties judging facial emotion, the prosopagnosics exhibited excellent recognition of vocal affect, suggestive of a face-specific difficulty.

In our first two experiments, we observed striking correlations between expression classification accuracy and performance on the CFPT (Duchaine et al., 2007). The CFPT is thought to provide a demanding test of face encoding - observers' ability to represent and discriminate facial shape - in the absence of substantial demands on visual memory. Poor performance on this test is suggestive of an apperceptive form of prosopagnosia (Dalrymple, Garrido et al., 2014; Duchaine et al., 2007; Shah, Gaule, Gaigg, Bird, & Cook, 2015). Strikingly, when the DP sample was split into apperceptive and non-apperceptive subgroups based on CFPT performance, only the apperceptive subgroup exhibited impaired recognition of facial emotion. DPs with an apperceptive profile may have difficulties forming view-invariant structural descriptions of faces at an early stage in the face processing stream, before the divergence of identity and expression processing (Bruce & Young, 1986; Haxby et

al., 2000). Inaccurate descriptions of local feature shape may result in imprecise expression categorisation as well as severe problems recognising facial identity.

To our knowledge, these findings are the first evidence of impaired recognition of facial emotion in DP, at the group level. Importantly, our results suggest that the ability to detect emotion recognition difficulties in this population may be extremely sensitive to the procedure used. In our first experiment, the clearest group difference was observed when observers were required to categorise expressions containing different degrees of surprise and fear. Typical observers also found these categorisations more demanding, and the increased difficulty may be responsible for the clear group difference observed. Alternatively, DPs with an apperceptive profile may have particular problems encoding the shape of the eye-region, variation crucial for distinguishing emotions, notably fear and surprise (Adolphs et al., 2005). In our second experiment, a clear group difference was observed only when judging the eye-region stimuli containing intermediate emotion intensities. All three levels of emotion intensity (30%, 50%, 70%) yielded recognition performance comfortably above chance (floor) and below 100% (ceiling) when typical observers were tested. However, stimuli either side of the 50% 'sweet-spot' may i) contain sufficiently obvious cues to be detected by observers with apperceptive deficits, or ii) be difficult for some typical observers to categorise reliably.

In light of these results, we recommend that authors demand a high standard of evidence before concluding that cases of DP exhibit intact emotion recognition. With respect to methodology, task sensitivity is a crucial issue. Modelling the categorisation of stimuli drawn from morph continua, by fitting psychometric functions, offers a precise means to estimate perceptual sensitivity independently of response bias[4]. Where morph continua are employed, the use of 7 levels of stimulus intensity affords greater sensitivity than the 5 stimulus levels present in the 'morph hexagon' used previously (Humprhreys et al., 2007; Palermo et al., 2011). The use of shorter presentation durations and ambiguous expression stimuli may have also increased sensitivity in the present study. With respect to sample size and composition, it is important that group studies have sufficient power to detect impairments. As awareness of DP increases, it should be easier to run group designs with reasonable sample sizes. Our results also suggest that studies with larger numbers of apperceptive DPs may be more likely to find expression recognition deficits. Where samples include relatively few DPs with an apperceptive profile, authors may consider qualifying their conclusions accordingly.

Problems recognising facial identity – the defining feature of DP – can impact substantially on the social development and behaviour of sufferers. DPs often avoid social situations experiencing feelings of guilt and shame about actual or imagined offense caused to others (Davis et al., 2011). Long-term consequences can include reduced social circle, loss of self-confidence and limited work opportunities (Dalrymple, Fletcher et al., 2014; Fine, 2012; Yardley, McDermott, Pisarski, Duchaine, & Nakayama, 2008). In severe cases, DP can also contribute to the development of depression and anxiety (Yardley et al., 2008). Where observed, problems recognising the expressions of interactants will likely exacerbate these difficulties. Reduced ability to detect the emotional and mental states of others may prevent DPs responding appropriately and hinder social interaction, particularly in situations where vocal cues are unavailable. At present relatively little is known about the impact of DP during childhood (Dalrymple, Corrow, Yonas, & Duchaine, 2013). The present results suggest the possibility that reasoning about the mental states of others ('theory of mind') may sometimes develop atypically in DP.

In summary, having tested a group of 17 DPs on complementary emotion recognition tasks, we find evidence of widespread difficulties recognising facial affect. These findings are contrary to the view that emotion recognition deficits are relatively uncommon in this population (Humphreys et al., 2007; Palermo et al., 2011). Deficits were apparent when observers were asked to categorise emotion using cues from the whole-face or from the eye-region only, and thus do not appear to reflect diminished integration of information from disparate facial regions (i.e., aberrant holistic processing). Instead, individuals with apperceptive forms of DP appear to have difficulties encoding facial shape, at an early stage in the face processing stream, before the divergence of identity and expression pathways. More broadly, these findings serve to illustrate how existing theoretical frameworks can be used to make sense of the heterogeneity seen in this population.

Footnotes

[1] We use the term Developmental Prosopagnosia in preference to Congenital Prosopagnosia to reflect the possibility that the condition emerges during development, and may not necessarily be present from birth.

[2] While expression stimuli were drawn from morph continua, psychophysical analyses were not employed (e.g., psychometric functions were not estimated). The authors' analysis was restricted to proportions of correct responses, defined through reference to the dominant emotion signal present in each stimulus.

[3] Pilot testing of a 6-AFC procedure (happiness, anger, disgust, sadness, fear, and surprise) revealed that typical participants were unable to reliably distinguish i) angry and disgusted eyes, and ii) fearful and surprised eyes. One expression in each of the two problematic pairs was therefore dropped (i.e., disgust and surprise).

[4] In previous studies employing the morph hexagon, the authors have selected particular levels and analysed % correct. Fitting psychometric functions may yield more accurate measures of perceptual precision that allow for individual differences in response bias.

References

Adolphs, R. (2002). Recognizing emotion from facial expressions: psychological and neurological mechanisms. *Behavioral and Cognitive Neuroscience Reviews, 1*(1), 21-62.

Adolphs, R., Gosselin, F., Buchanan, T. W., Tranel, D., Schyns, P., & Damasio, A. R. (2005). A mechanism for impaired fear recognition after amygdala damage. *Nature, 433*(7021), 68-72.

Allen, R., Davis, R., & Hill, E. (2012). The effects of autism and alexithymia on physiological and verbal responsiveness to music. *Journal of Autism and Developmental Disorders, 43*(2), 432-444.

Ariel, R., & Sadeh, M. (1996). Congenital visual agnosia and prosopagnosia in a child: a case report. *Cortex, 32*(2), 221-240.

Avidan, G., & Behrmann, M. (2009). Functional MRI reveals compromised neural integrity of the face processing network in congenital prosopagnosia. *Current Biology, 19*(13), 1146-1150.

Avidan, G., Tanzer, M., & Behrmann, M. (2011). Impaired holistic processing in congenital prosopagnosia. *Neuropsychologia, 49*(9), 2541-2552.

Avidan, G., Tanzer, M., Hadj-Bouziane, F., Liu, N., Ungerleider, L. G., & Behrmann, M. (2013). Selective dissociation between core and extended regions of the face processing network in congenital prosopagnosia. *Cerebral Cortex, 24*(6), 1565-1578.

Baron-Cohen, S., Wheelwright, S., Hill, J., Raste, Y., & Plumb, I. (2001). The "Reading the Mind in the Eyes" test revised version: A study with normal adults, and adults with Asperger syndrome or high-functioning autism. *Journal of Child Psychology and Psychiatry, 42*(2), 241-251.

Bate, S., & Bennetts, R. (2015). The independence of expression and identity in face-processing: evidence from neuropsychological case studies. *Frontiers in Psychology, 6*(770), 1-7.

Behrmann, M., & Avidan, G. (2005). Congenital prosopagnosia: face-blind from birth. *Trends in Cognitive Sciences, 9*(4), 180-187.

Behrmann, M., Avidan, G., Gao, F., & Black, S. (2007). Structural imaging reveals anatomical alterations in inferotemporal cortex in congential prosopagnosia. *Cerebral Cortex, 17*(10), 2354-2363.

Bentin, S., Degutis, J. M., D'Esposito, M., & Robertson, L. C. (2007). Too many trees to see the forest: performance, event-related potential, and functional magnetic resonance imaging manifestations of integrative congenital prosopagnosia. *Journal of Cognitive Neuroscience, 19*(1), 132-146.

Bird, G., & Cook, R. (2013). Mixed emotions: the contribution of alexithymia to the emotional symptoms of autism. *Translational Psychiatry, 3*, e285.

Boersma, P., & Weenink, D. (2015). *Praat: doing phonetics by computer [Computer program]. Version 6.0.08, retrieved 1 2015 from http://www.praat.org/.*

Brainard, D. H. (1997). The psychophysics toolbox. *Spatial Vision, 10*(4), 433-436.

Brewer, R., Happe, F., Cook, R., & Bird, G. (2015). Commentary on "Autism, oxytocin and interoception": Alexithymia, not Autism Spectrum Disorders, is the consequence of interoceptive failure. *Neuroscience and Biobehavioral Reviews, 56*, 348-353.

Bruce, V., & Young, A. W. (1986). Understanding face recognition. *British Journal of Psychology, 77*, 305-327.

Bukach, C. M., Le Grand, R., Kaiser, M. D., Bub, D. N., & Tanaka, J. W. (2008). Preservation of mouth region processing in two cases of prosopagnosia. *Journal of Neuropsychology, 2*(1), 227-244.

Caldara, R., Schyns, P., Mayer, E., Smith, M. L., Gosselin, F., & Rossion, B. (2005). Does prosopagnosia take the eyes out of face representations? Evidence for a defect in representing diagnostic facial information following brain damage. *Journal of Cognitive Neuroscience, 17*(10), 1652-1666.

Calder, A. J., & Young, A. W. (2005). Understanding the recognition of facial identity and facial expression. *Nature Reviews Neuroscience, 6*(8), 641-651.

Cook, R. & Biotti, F. (2016). Developmental prosopagnosia. *Current Biology, 26*(8), R1-R2.

Cook, R., Brewer, R., Shah, P., & Bird, G. (2013). Alexithymia, not autism, predicts poor recognition of emotional facial expressions. *Psychological Science, 24*(5), 723-732.

Dalrymple, K. A., Corrow, S., Yonas, A., & Duchaine, B. (2013). Developmental prosopagnosia in childhood. *Cognitive Neuropsychology, 29*(5-6), 393-418.

Dalrymple, K. A., Fletcher, K., Corrow, S., das Nair, R., Barton, J. J., Yonas, A., et al. (2014). "A room full of strangers every day": the psychosocial impact of developmental prosopagnosia on children and their families. *Journal of Psychosomatic Research, 77*(2), 144-150.

Dalrymple, K. A., Garrido, L., & Duchaine, B. (2014). Dissociation between face perception and face memory in adults, but not children, with developmental prosopagnosia. *Developmental Cognitive Neuroscience, 10*, 10-20.

Davis, J. M., McKone, E., Dennett, H., O'Connor, K. B., O'Kearney, R., & Palermo, R. (2011). Individual differences in the ability to recognise facial identity are associated with social anxiety. *PLoS One, 6*(12), e28800.

De Haan, E. H., & Campbell, R. (1991). A fifteen year follow-up of a case of developmental prosopagnosia. *Cortex, 27*(4), 489-509.

De Renzi, E., Faglioni, P., Grossi, D., & Nichelli, P. (1991). Apperceptive and associative forms of prosopagnosia. *Cortex, 27*(2), 213-221.

Dennett, H. W., McKone, E., Tavashmi, R., Hall, A., Pidcock, M., Edwards, M., et al. (2011). The Cambridge Car Memory Test: a task matched in format to the Cambridge Face Memory Test, with norms, reliability, sex differences, dissociations from face memory, and expertise effects. *Behavior Research Methods, 44*(2), 587-605.

Dobel, C., Bölte, J., Aicher, M., & Schweinberger, S. R. (2007). Prosopagnosia without apparent cause: Overview and diagnosis of six cases. *Cortex, 43*(6), 718-733.

Duchaine, B., Germine, L., & Nakayama, K. (2007). Family resemblance: ten family members with prosopagnosia and within-class object agnosia. *Cognitive Neuropsychology, 24*(4), 419-430.

Duchaine, B., Murray, H., Turner, M., White, S., & Garrido, L. (2009). Normal social cognition in developmental prosopagnosia. *Cognitive Neuropsychology, 26*(7), 620-634.

Duchaine, B., & Nakayama, K. (2006a). The Cambridge Face Memory Test: results for neurologically intact individuals and an investigation of its validity using inverted face stimuli and prosopagnosic participants. *Neuropsychologia, 44*(4), 576-585.

Duchaine, B., & Nakayama, K. (2006b). Developmental prosopagnosia: a window to content-specific face processing. *Current Opinion in Neurobiology, 16*, 166-173.

Duchaine, B., Parker, H., & Nakayama, K. (2003). Normal recognition of emotion in a prosopagnosic. *Perception, 32*(7), 827-838.

Duchaine, B., Yovel, G., Butterworth, E., & Nakayama, K. (2006). Prosopagnosia as an impairment to face-specific mechanisms: Elimination of the alternative hypotheses in a developmental case. *Cognitive Neuropsychology, 23*(5), 714-747.

Ekman, P., & Friesen, W. V. (1975). *Pictures of Facial Affect*. Palo Alto, CA: Consulting Psychologists Press.

Fine, D. R. (2012). A life with prosopagnosia. *Cognitive Neuropsychology, 29*(5-6), 354-359.

Frith, C. (2009). Role of facial expressions in social interactions. *Philosophical Transactions of the Royal Society B: Biological Sciences, 364*(1535), 3453-3458.

Garrido, L., Eisner, F., McGettigan, C., Stewart, L., Sauter, D., Hanley, J. R., et al. (2009). Developmental phonagnosia: a selective deficit of vocal identity recognition. *Neuropsychologia, 47*(1), 123-131.

Garrido, L., Furl, N., Draganski, B., Weiskopf, N., Stevens, J., Tan, G. C. Y., et al. (2009). Voxel-based morphometry reveals reduced grey matter volume in the temporal cortex of developmental prosopagnosics. *Brain, 132*, 3443-3455.

Gomez, J., Pestilli, F., Witthoft, N., Golarai, G., Liberman, A., Poltoratski, S., et al. (2015). Functionally defined white matter reveals segregated pathways in human ventral temporal cortex associated with category-specific processing. *Neuron, 85*(1), 216-227.

Haxby, J. V., Hoffman, E. A., & Gobbini, M. I. (2000). The distributed human neural system for face perception. *Trends in Cognitive Sciences, 4*(6), 223-233.

Heaton, P., Reichenbacher, L., Sauter, D., Allen, R., Scott, S., & Hill, E. (2012). Measuring the effects of alexithymia on perception of emotional vocalizations in autistic spectrum disorder and typical development. *Psychological Medicine, 42*(11), 2453-2459.

Humphreys, K., Avidan, G., & Behrmann, M. (2007). A detailed investigation of facial expression processing in congenital prosopagnosia as compared to acquired prosopagnosia. *Experimental Brain Research, 176*(2), 356-373.

Ipser, A., & Cook, R. (2015). Inducing a concurrent motor load reduces categorization precision for facial expressions. *Journal of Experimental Psychology: Human Perception Performance*.

Ishihara, S. (1993). *Ishihara's Tests for Colour-Blindness*. Tokyo, Japan: Kanehara.

Johnen, A., Schmukle, S. C., Hüttenbrink, J., Kischka, C., Kennerknecht, I., & Dobel, C. (2014). A family at risk: Congenital prosopagnosia, poor face recognition and visuoperceptual deficits within one family. *Neuropsychologia, 58*, 52-63.

Kanwisher, N. (2000). Domain specificity in face perception. *Nature Neuroscience, 3*(8), 759-763.

Kennerknecht, I., Grüter, T., Welling, B., Wentzek, S., Horst, J., Edwards, S., et al. (2006). First report of prevalence of non-syndromic hereditary prosopagnosia (HPA). *American Journal of Medical Genetics, 140A*(15), 1617-1622.

Kennerknecht, I., Ho, N. Y., & Wong, V. C. N. (2008). Prevalence of heriditary prosopagonsia (HPA) in Hong Kong Chinese population. *American Journal of Medical Genetics, 146A*(22), 2863-2870.

Kress, T., & Daum, I. (2003a). Developmental prosopagnosia: a review. *Behavioral Neurology, 14*(3-4), 109-121.

Kress, T., & Daum, I. (2003b). Event-related potentials reflect impaired face recognition in patients with congenital prosopagnosia. *Neuroscience Letters, 352*(2), 133-136.

Langner, O., Dotsch, R., Bijlstra, G., Wigboldus, D. H. J., Hawk, S. T., & van Knippenberg, A. (2010). Presentation and validation of the Radboud Faces Database. *Cognition & Emotion, 24*(8), 1377-1388.

Lee, Y., Duchaine, B., Wilson, H. R., & Nakayama, K. (2010). Three cases of developmental prosopagnosia from one family: detailed neuropsychological and psychophysical investigation of face processing. *Cortex, 46*(8), 949-964.

Liu, T. T., & Behrmann, M. (2014). Impaired holistic processing of left-right composite faces in congenital prosopagnosia. *Frontiers in Human Neuroscience, 8*, 750.

McConachie, H. R. (1976). Developmental prosopagnosia. A single case report. *Cortex, 12*(1), 76-82.

Minnebusch, D. A., Suchan, B., Ramon, M., & Daum, I. (2007). Event-related potentials reflect heterogeneity of developmental prosopagnosia. *European Journal of Neuroscience, 25*(7), 2234-2247.

Nunn, J. A., Postma, P., & Pearson, R. (2001). Developmental prosopagnosia: Should it be taken at face value? *Neurocase, 7*(1), 15-27.

Palermo, R., Willis, M. L., Rivolta, D., McKone, E., Wilson, C. E., & Calder, A. J. (2011). Impaired holistic coding of facial expression and facial identity in congenital prosopagnosia. *Neuropsychologia, 49*(5), 1226-1235.

Parkinson, B. (2005). Do facial movements express emotions or communicate motives? *Personality and Social Psychology Review, 9*(4), 278-311.

Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: transforming numbers into movies. *Spatial Vision, 10*(4), 437-442.

Prins, N., & Kingdom, F. A. A. (2009). Palamedes: Matlab routines for analyzing psychophysical data. . *http://www.palamedestoolbox.org*.

Scherer, K. R. (1986). Vocal affect expression: a review and a model for future research. *Psychological Bulletin, 99*(2), 143-165.

Schmalzl, L., Palermo, R., & Coltheart, M. (2008). Cognitive heterogeneity in genetically based prosopagnosia: a family study. *Journal of Neuropsychology, 2*(Pt 1), 99-117.

Shah, P., Gaule, A., Gaigg, S. B., Bird, G., & Cook, R. (2015). Probing short-term face memory in developmental prosopagnosia. *Cortex, 64*, 115-122.

Shah, P., Gaule, A., Sowden, S., Bird, G., & Cook, R. (2015). The 20-item prosopagnosia index (PI20): a self-report instrument for identifying developmental prosopagnosia. *Royal Society Open Science, 2*(6), 140343.

Susilo, T., & Duchaine, B. (2013). Advances in developmental prosopagnosia research. *Current Opinion in Neurobiology, 23*, 423-429.

Thomas, C., Avidan, G., Humphreys, K., Jung, K. J., Gao, F., & Behrmann, M. (2009). Reduced structural connectivity in ventral visual cortex in congential prosopagnosia. *Nature Neuroscience, 12*, 29-31.

Todorov, A., & Duchaine, B. (2008). Reading trustworthiness in faces without recognizing faces. *Cognitive Neuropsychology, 25*(3), 395-410.

Yardley, L., McDermott, L., Pisarski, S., Duchaine, B., & Nakayama, K. (2008). Psychosocial consequences of developmental prosopagnosia: a problem of recognition. *Journal of Psychosomatic Research, 65*(5), 445-451.

Young, A. W., Perrett, D. I., Calder, A. J., Sprengelmeyer, R., & Ekman, P. (2002). *Facial expressions of emotion: Stimuli and tests (FEEST) [computer software]*. Bury St Edmunds, England: Thames Valley Test Company.

Table 1: Scores of each Developmental Prosopagnosic on the Twenty-Item Prosopagnosia Index (PI20), The Cambridge Face Perception Test (CFPT), The Cambridge Face Memory Test (CFMT), The Cambridge Bicycle Memory Test (CBMT), The Cambridge Car Memory Test (CCMT). The mean and standard deviation of the comparison sample (N = 35) are provided below. The z-scores provided for the CFPT are based on performance in the upright condition.

Participant	Age	PI20	CFPT upright	CFPT inverted	CFMT %	CBMT%	CCMT%	CFMT z-scores	CFPT z-scores	PI20 z-scores
F1	43	79	26	58	62.50	83.33	62.50	-2.18	-0.45	-4.16
F2	46	61	68	58	52.78	80.56	73.61	-3.13	-4.25	-2.35
F3	22	89	30	70	50.00	77.78	52.78	-3.40	0.00	-5.16
F4	42	92	62	82	45.83	94.44	65.28	-3.81	-3.58	-5.47
F5	70	95	100	92	36.11	-	76.39	-4.76	-7.83	-5.77
F6	40	85	40	82	40.28	59.72	63.89	-4.35	-1.12	-4.76
F7	50	78	34	52	58.33	91.67	86.11	-2.59	-0.45	-4.06
F8	21	59	30	64	63.89	75.00	-	-2.04	0.00	-2.15
F9	29	68	32	58	61.11	-	63.89	-2.31	-0.22	-3.06
F10	63	79	40	70	61.11	80.56	66.67	-2.31	-1.12	-4.16
F11	53	85	74	94	45.83	88.89	63.89	-3.81	-4.92	-4.76
F12	65	81	44	78	59.72	93.06	66.67	-2.45	-1.57	-4.36
M1	47	92	86	54	51.39	98.61	94.44	-3.26	-6.26	-5.47
M2	46	72	50	64	66.67	90.28	76.39	-1.77	-2.24	-3.46
M3	68	92	92	78	27.78	70.83	47.22	-5.57	-6.94	-5.47
M4	43	78	52	50	58.33	94.44	93.06	-2.59	-2.46	-4.06
M5	28	62	46	60	62.50	77.78	69.44	-2.18	-1.79	-2.45
TD mean	40.48	37.35	29.40	60.00	85.00	83.22	61.45			
TD SD	13.31	10.75	9.48	14.47	8.48	12.26	13.15			

Table 2: Performance of the Developmental Prosopagnosics (DPs) and the typically developing controls (TDs) used in Experiments 1-3 on The Cambridge Face Memory Test (CFMT), The Cambridge Face Perception Test (CFPT), and The Twenty-Item Prosopagnosia Index (PI20).

	DPs		TDs (Experiment 1)			TDs (Experiment 2)			TDs (Experiment 3)		
	M	*SD*	*M*	*SD*	*contrast*	*M*	*SD*	*contrast*	*M*	*SD*	*contrast*
CFMT	53.18	10.94	83.68	9.94	$p < .001$	83.13	10.25	$p < .001$	83.41	10.34	$p < .001$
CFPT	53.29	23.22	32.48	8.48	$p < .010$	32.67	7.39	$p < .010$	32.24	9.38	$p < .010$
PI20	79.24	11.52	37.05	9.52	$p < .001$	36.95	9.60	$p < .001$	38.05	9.88	$p < .001$

Table 3: Correlations between the expression categorisation thresholds observed in Experiment 1 and participants' scores on the Cambridge Face Memory Test (CFMT), the Cambridge Face Perception Test (CFPT), the 20-item Prosopagnosia Index (PI20), the Cambridge Car Memory Test (CCMT) and the Cambridge Bike Memory Test (CBMT).

	CFMT		CFPT		PI20		CCMT		CBMT	
	r	*p*	*r*	*p*	*r*	*p*	*r*	*p*	*r*	*p*
Happiness-anger	-.23	.16	.24	.14	.08	.62	.00	.99	-.19	.29
Disgust-sadness	-.22	.19	.33	.04	.27	.09	.05	.80	.02	.92
Fear-surprise	-.57	< .001	.78	< .001	.51	< .001	.00	.98	-.03	.88
Mean threshold	-.46	.004	.61	< .001	.35	.03	.02	.92	-.13	.46

Table 4: Correlations between the expression recognition accuracies scores observed in Experiment 2 and participants' scores on the Cambridge Face Memory Test (CFMT), the Cambridge Face Perception Test (CFPT), the 20-item Prosopagnosia Index (PI20), the Cambridge Car Memory Test (CCMT) and the Cambridge Bike Memory Test (CBMT).

	CFMT		CFPT		PI20		CCMT		CBMT	
	r	*p*	*r*	*p*	*r*	*p*	*r*	*p*	*r*	*p*
70% (high intensity)	.39	.01	-.55	<.001	-.44	.007	-.01	.97	.02	.90
50% (moderate intensity)	.49	.002	-.52	.002	-.56	<.001	.01	.95	.10	.58
30% (low intensity)	.36	.03	-.46	.005	-.29	.12	-.05	.77	.17	.35
Overall performance	.48	.002	-.58	< .001	-.48	.003	-.02	.93	.11	.54

Figure 1: (a) Morphed expression stimuli used in Experiment 1. (b) Mean categorisation thresholds for the three continua exhibited by the typical observers and the developmental prosopagnosics. Error bars represent ± 1 standard error of the mean. (c) Scatter plot of the relationship observed between participants' scores on the Cambridge Face Perception Test (CFPT) and their thresholds for the Fear-Surprise categorisations.

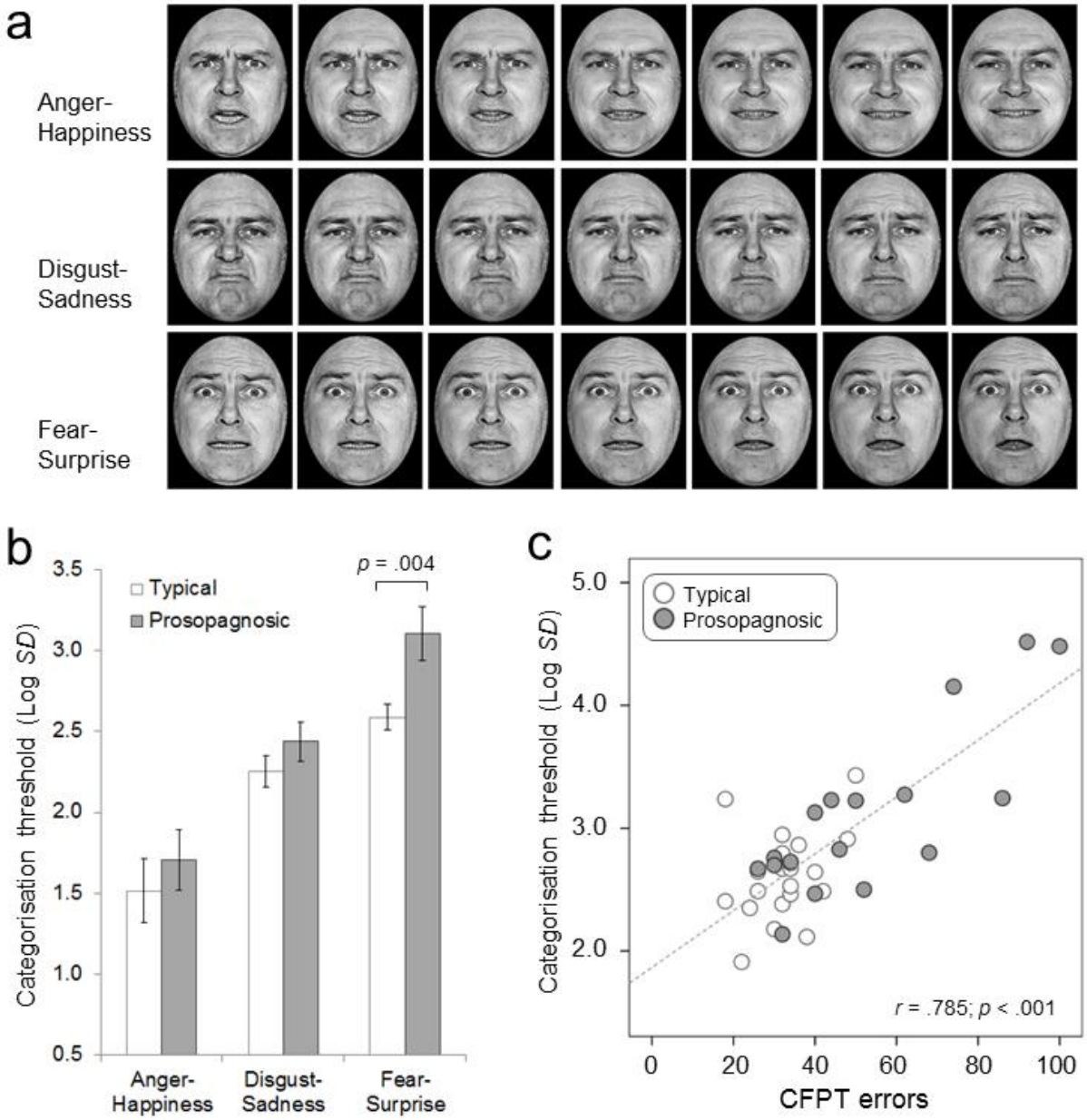

Figure 2: (a) Single-case analysis of the surprise-fear thresholds observed in Experiment 1. (b) Single-case analysis of the overall performance observed in Experiment 2. Error bars represent ± 1 standard deviation. * denotes performance < 1 standard deviation below the TD mean; ** denotes performance at least 2 standard deviations below the TD mean.

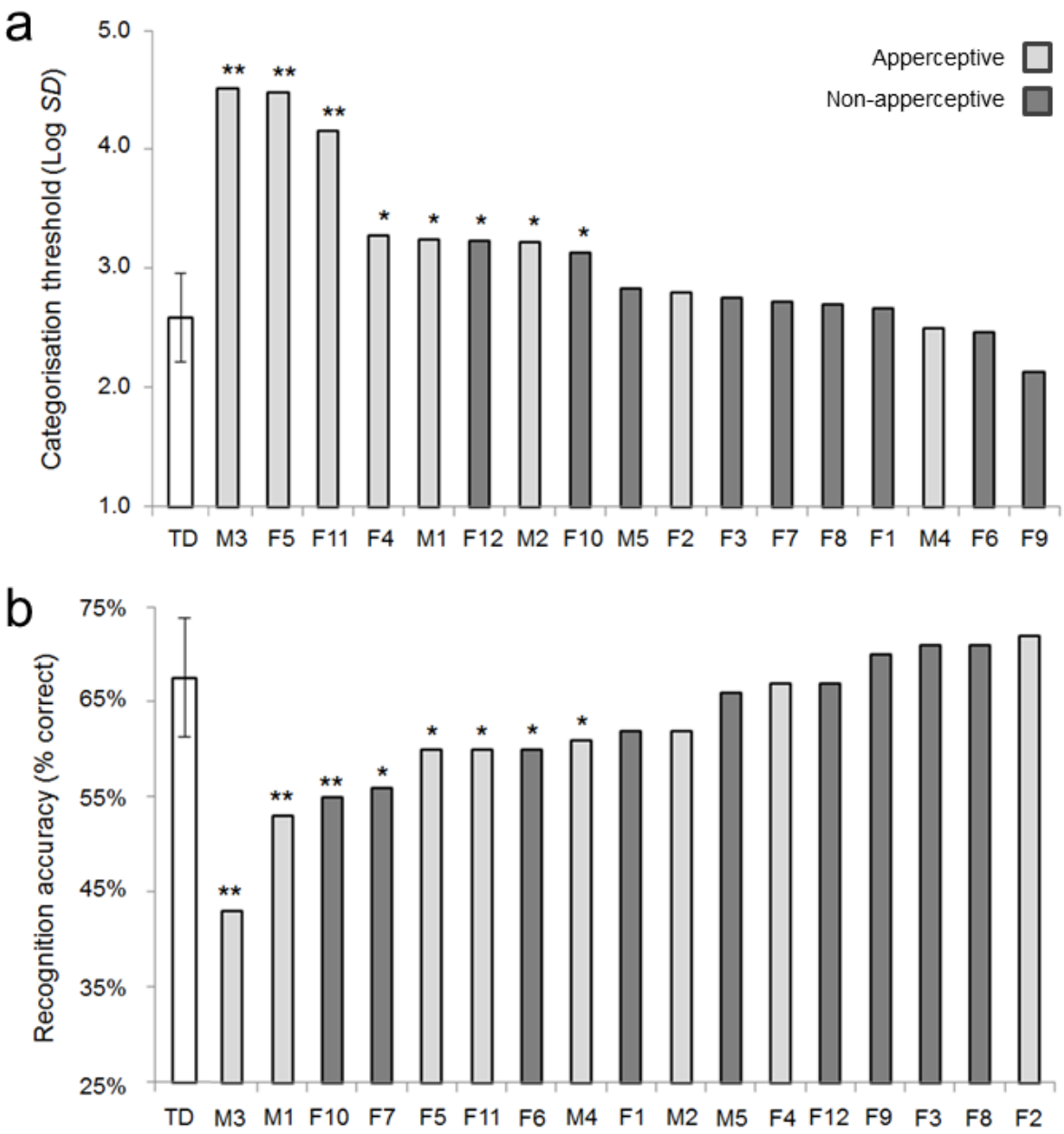
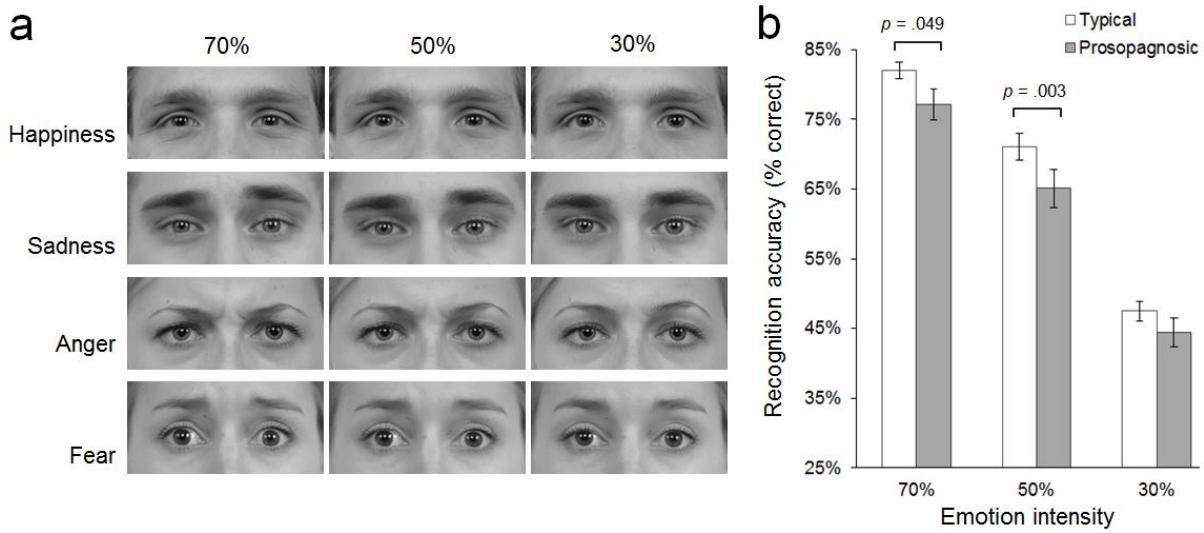
Figure 3: (a) Examples of the eye-region stimuli used in Experiment 2. (b) The mean recognition accuracy exhibited by the typical observers and the developmental prosopagnosics in the three emotion intensity conditions. Error bars represent ± 1 standard error of the mean.