This is a repository copy of *Face matching impairment in developmental prosopagnosia*.

Face matching impairment in developmental prosopagnosia

David White[1], Davide Rivolta[2], A. Mike Burton[3], Shahd Al-Janabi[4], Romina Palermo[5]

[1]School of Psychology, University of New South Wales, Australia
[2] School of Psychology, University of East London, UK
[3] Department of Psychology, University of York, UK
[4] Department of Psychology, University of Wisconsin-Milwaukee, USA
[5] ARC Centre of Excellence in Cognition and its Disorders, and School of Psychology, University of Western Australia, Australia

Submitted to QJEP special issue:
"Advancing Developmental Prosopagnosia Research", S. Bate & J. Tree (Eds.).

Running Head: Face matching in developmental prosopagnosia
Keywords: face perception; face recognition; congenital prosopagnosia; image variability; unfamiliar face matching.
Word Count (excluding abstract, table & figure captions): 6591 words
Abstract: 211 words

Corresponding Author:
Dr. David White
School of Psychology
University of New South Wales
Kensington
Sydney
NSW 2052
david.white@unsw.edu.au
(+61) 2 9385 325

ABSTRACT

Developmental prosopagnosia (DP) is commonly referred to as 'face blindness', a term that implies a perceptual basis to the condition. However, DP presents as a deficit in face recognition and is diagnosed using memory-based tasks. Here, we test face identification ability in six people with DP, who are severely impaired on face memory tasks, using tasks that do not rely on memory. First, we compared DP to control participants on a standardised test of unfamiliar face matching using facial images taken on the same day and under standardised studio conditions (*Glasgow Face Matching Test;* GFMT). DP participants did not differ from normative accuracy scores on the GFMT. Second, we tested face matching performance on a test created using images that were sourced from the Internet and so vary substantially due to changes in viewing conditions and in a person's appearance (*Local Heroes Test;* LHT). DP participants show significantly poorer matching accuracy on the LHT relative to control participants, for both unfamiliar and familiar face matching. Interestingly, this deficit is specific to 'match' trials, suggesting that people with DP may have particular difficulty in matching images of the same person that contain natural day-to-day variations in appearance. We discuss these results in the broader context of individual differences in face matching ability.

INTRODUCTION

Developmental Prosopagnosia (DP) results from a failure to develop the cognitive mechanisms necessary for adequate face identity recognition (Dalrymple & Palermo, 2016; Rivolta, Palermo & Schmalzl, 2013; Susilo & Duchaine, 2013). Individuals with DP (also known as congenital or hereditary prosopagnosia) do not report brain injury, have typical vision and do not have general intellectual impairments, yet they report everyday difficulties recognizing familiar faces. Some have difficulty recognizing the faces of close friends, family and even themselves; for others the difficulty is limited to recognizing less frequently seen people in unexpected contexts, for example when meeting a neighbor at the supermarket.

Importantly, DP is characterized as a deficit in face *memory* and cases of DP are confirmed using memory-based tasks (see Dalrymple & Palermo, 2016).   In daily life, the condition primarily affects a person's ability to recognise faces of people they know. However, given the time consuming nature of constructing tests using personally familiar faces, tests of famous face recognition are typically used (e.g., Macquarie Famous Face Test- 2008, Palermo, Rivolta, Wilson & Jeffery 2011). Another common method for measuring face learning and memory ability in DP is the Cambridge Face Memory Test (CFMT, Duchaine & Nakayama, 2006). In this standardised test, participants study the images of six unfamiliar males and are then tested for their recognition accuracy across changes in viewpoint, lighting and with the addition of visual noise.

While face memory is impaired in DP by definition, it is less clear how often face *perception* is impaired in adults with this condition. Within individual studies, it is very common for some DP participants to show difficulties on tests of face perception, while other cases do not (Chatterjee & Nakayama, 2012; Dalrymple, Garrido, & Duchaine, 2014; Humphreys, Avidan, & Behrmann, 2007; Palermo, Willis, et al., 2011; Rivolta, Palermo, Schmalzl & Coltheart, 2012). Consistent with this heterogeneity, impairments in acquired prosopagnosia – where face recognition impairments are the result of brain injury – cluster into deficits that primarily disrupt face perception ("apperceptive" prosopagnosia: De Renzi1 Faglioni, Grossi & Nichelli, 1991; Young, Newcombe, Haan, Small & Hay, 1993; Dalrymple et al. 2011)

and deficits that appear to spare face perception ("associative" prosopagnosia: De Renzi et al., 1991; Barton, 2008; or "prosopamnesia": Tippett, Miller & Farah, 2000). Thus, neuropsychological evidence suggests that face perception and face recognition rely on dissociable stages of face processing (Bruce & Young, 1986), which may explain why development of normal face perception in DP can be independent of impairments in face memory.

Face *perception* abilities in DP are commonly assessed using the Cambridge Face Perception Test (CFPT, Duchaine et al. 2007, Figure 1: see also Bowles et al., 2009; Dalrymple, Garrido & Duchaine, 2014). In the CFPT, participants are given one minute to arrange an array of six facial images according to their similarity to a target face. The CFPT is designed as a perceptual task and so all images appear on the screen at the same time, therefore minimising demands on memory. However, unlike tests of face memory, the CFPT does not measure face identification ability directly, but rather indexes participants' perception of facial similarity *between* identities. Stimulus arrays are created by morphing from the target face to six other identities, which introduces changes to the appearance of facial structure (i.e. changes to the face that signal changes in identity). Performance on this test is indexed by the degree to which subjective perceptions of facial similarity aligns with objective similarity, as defined by the relative weighting of the six foil identities in the morphed images. Although it may be argued that veridical perception of similarity is necessary for successful identification, it is not clear that this test recruits face identification processes.

Face identification requires determining whether two images are of the same person, while accounting for *within*-identity changes in facial appearance caused by variables such as camera-to-subject distance, lighting, head orientation and expression (see Bruce, 1994; Jenkins, White, Van Monfort & Burton, 2011; Burton, 2013). Importantly, the few studies that have tested face matching ability in DP have used tests created from images taken in a single studio session and with a single camera. This approach has important consequences, as it enables identification to be achieved by comparing image-specific parameters and so may not reflect a person's skill in matching identity across variable input stimuli (Burton, 2013; Duchaine & Nakayama, 2004, 2006). In support of this, tests created in this way often produce ceiling levels

of performance in DP participants, even when external features such as hair are removed (e.g. Humphreys, Avidan & Behrmann, 2007). Similarly, the Benton Facial Recognition Test (BFRT: Benton, Sivan, Hamsher, Varney & Spreen, 1994) requires participants to match identity of images that are presented simultaneously on the screen, but which are also highly standardised in terms of lighting and capture settings. Some studies show DP participants are impaired in the BFRT (Huis in 't Veld, Van den Stock & de Gelder, 2012), while others report that individuals with DP can perform well by adopting a feature matching strategy (Duchaine & Nakayama, 2004, 2006).

Ascertaining whether people with DP are impaired in face identification tasks that do not involve memory is important in determining which stages of face processing are impaired. It is therefore surprising that studies have not used a wider range of tests to examine perceptual impairments. In the context of the broader population, perceptually-based identification tasks have been studied extensively, primarily due to the importance of reliably verifying the identity of facial images in applied settings (e.g. Bruce et al. 1999; Burton, White & McNeill, 2010; O'Toole, An, Dunlop, Natu & Phillips, 2012). This work has consistently shown that matching identity of unfamiliar faces, in the absence of memory demands, is difficult – even for participants with otherwise typical face recognition abilities (e.g. Bruce et al. 1999; Burton et al. 2010; White, Kemp, Jenkins, Matheson & Burton, 2014) and with professional experience in the task (White et al. 2014a; White, Dunn, Schmid & Kemp, 2015; White, Phillips, Hahn, Hill & O'Toole, 2015).

To test unfamiliar face matching ability, many recent studies have used the *Glasgow Face Matching Test* (GFMT: Burton et al. 2010, See Figure 2A). In this test, participants decide whether two images presented side-by-side on a computer monitor are the same person or two different people. All images are taken on the same day, under similar lighting conditions and in the same neutral pose – but crucially with different cameras. Although superficial, this image change introduces subtle differences in aspect ratio and metric distances across face images (Burton, Schweinberger, Jenkins & Kaufmann, 2015), resulting in nontrivial variations across images of the same face that must be tolerated when matching identity (see Figure 2A, top row). Studies reporting performance on this test in the general population

show average error rates of 20% (where chance is 50%). In other tests created from photos captured in unconstrained environmental conditions, referred to as 'ambient' images because they contain natural day-to-day variations in a person's appearance, even poorer accuracy has been reported (e.g. O'Toole et al. 2012; White et al. 2014a, White, Kemp, Jenkins & Burton, 2014; see Figure 2B, 2C for examples of ambient stimuli).

Here, we test the face perception abilities of a group of adults with DP who report everyday face recognition difficulties, as well as showing deficits in recognition of famous faces (MFFT-08; Palermo et al., 2011) and memory for previously unfamiliar faces (CFMT, Duchaine & Nakayama, 2006). First, we test their ability on the CFPT (Duchaine, Germine & Nakayama, 2007), a standard test used to determine whether adults with DP also show a face perception deficit. However, as discussed above, the CFPT does not explicitly test for ability to identify faces. Therefore, we also tested face *identity matching* in two tasks that do not involve memory: the *GFMT* (Burton et al. 2010), and the *Local Heroes Test*, the latter being a more challenging test of face identity matching created from 'ambient' images, as described above (see also Jenkins & Burton, 2011).

The *Local Heroes Test* (LHT) follows the same format as the GFMT – participants decide if two images are of the same person or of different people. However it differs from the GFMT in two ways. First, as discussed above, images were collected from the Internet and so in unconstrained, 'ambient' capture conditions. Second, the LHT involves matching identity of familiar as well as unfamiliar faces. The beneficial effect of familiarity to face matching accuracy in typical participants has been well documented (e.g. Clutterbuck & Johnston, 2004; Megreya & Burton, 2006; White, Burton, Jenkins & Kemp, 2014) and enables typical participants to match identity across substantial variation in appearance (Jenkins et al., 2011; White et al., 2014c). Therefore, we expected that typical participants would be more accurate on the *familiar* condition of the Local Heroes test as compared to the *unfamiliar* condition. Because DP participants are impaired in forming memory representations of familiar faces, we predicted that this enhancement would be reduced in DP participants.

METHOD

Control Participants

The LHT consists of local celebrities in the UK and Australia that are selected to be familiar to participants in only one of these locations. Therefore, we recruited control participants from both the UK (n = 11; Mean age = 48.5; SD = 9.0) and Australia (n = 12; Mean age = 39.9; SD = 10.0). The purpose of recruiting two groups was to verify a benefit on familiarity in typical participants that was independent of the particular stimuli used in each portion of the test. This also enabled comparison of DP performance on familiar and unfamiliar matching tasks with control groups that were both unfamiliar and familiar with each set of faces. These same control participants also completed the GFMT but did not complete the full battery of assessment tests completed by people with DP (see below).

People with DP

Six participants (4 female) reporting lifelong difficulties in face recognition were recruited via the Australian Prosopagnosia Register[1] (Mean age = 46.2 years; SD = 11.6). Visual acuity was assessed with a visual acuity test using Sloan font (see Dalrymple & Palermo, 2016) and/or the Functional Acuity Contrast Test (FACT- Vision Sciences Research Corporation, 2002) with all participants performing within typical limits. These participants were confirmed as having impairment in face recognition by a combination of poor performance on both the Macquarie Famous Face Test 2008 (MFFT-08, Palermo et al., 2011) and the Cambridge Face Memory Test (CFMT, Duchaine & Nakayama, 2006). Initial screening selected participants who scored below 2 standard deviations on age-adjusted z-scores for the MFFT. Consistent with recent work, the criteria for final inclusion of DP participants in the study was that the participant scored below 1.7 standard deviations on age-adjusted z-scores for the CFMT (see DeGutis, Cohan & Nakayama, 2014). In addition, we measured non-face object memory using the Cambridge Car Memory Task (CCMT; Dennett et al. 2012). Age-adjusted z-scores were computed using data from Bowles et al. (2009) for all diagnostic tests, and are presented in Figure 3. Raw scores are available in Supplementary Materials (Table S1).

---

[1] https://www.cogsci.mq.edu.au/research/projects/prosopagnosia/register/

Cambridge Face Perception Test (CFPT, Duchaine et al. 2007)

During initial screening for DP, participants also completed the CFPT. An example trial from the CFPT is shown in Figure 1. In the CFPT, participants are shown eight separate arrays that contain one target face (top) and six array images (bottom). Participants must rank the array images in order of their relative similarity to a target face. Array images are created by morphing the target face to images of six different identities, with varying contributions of the target face to each morph. Proportion of contribution of the target face to the array image is taken as an index of similarity between the target image and the array image, and performance is calculated as the number of ranking placements made by participants that do not match the morph-based ranking. Figure 1 shows the correct arrangement of target faces for one array. Previous work has shown high internal reliability of the CFPT (Cronbach's alpha = .74; Bowles et al., 2009). Z-scores for DP participants on the CFPT scores are shown in Figure 3 (see Table S1 for raw scores).
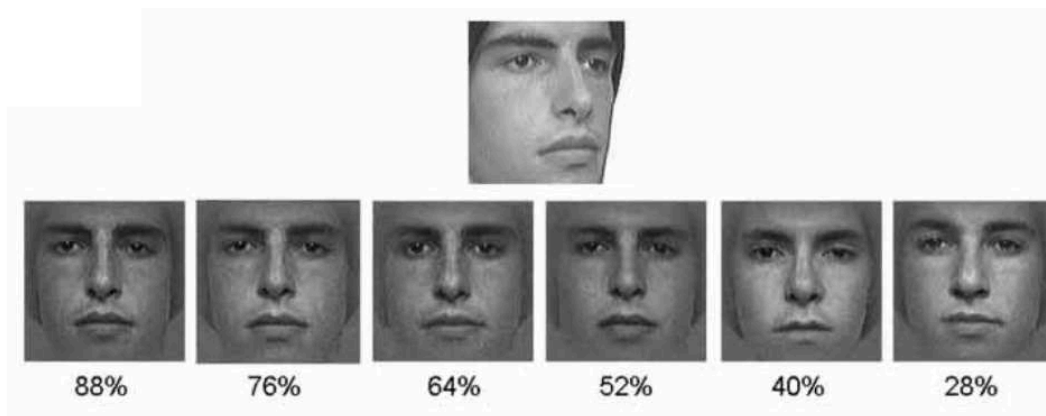
Figure 1. An example trial from the CFPT (see text for details).

Glasgow Face Matching Test (GFMT; Burton et al. 2010)

Stimuli for the short version of the GFMT consisted of 20 same- and 20 different-identity image pairs. Same-identity pairs show two images of the same person taken under similar lighting conditions, on the same day, but using different digital cameras. For different-identity pairs, one of these images was paired with a similar looking person from the database, so that each identity appears once in a same-identity pair and once in a different-identity pair. For each image pair, participants responded "same" or "different" identity. The task was self-paced and image pairs remain on the computer monitor until participants make their response, at which point the next

image pair was presented. Performance on the GFMT does not vary as a function of age (Burton et al. 2010; cf. Megreya & Bindemann, 2015) hence the z-scores for this test, which are presented in Figure 3, have not been age-adjusted. Internal reliability for this test based on data from Burton et al (2010) is very high (Cronbach's alpha = 0.91).

A   **B**   **C**
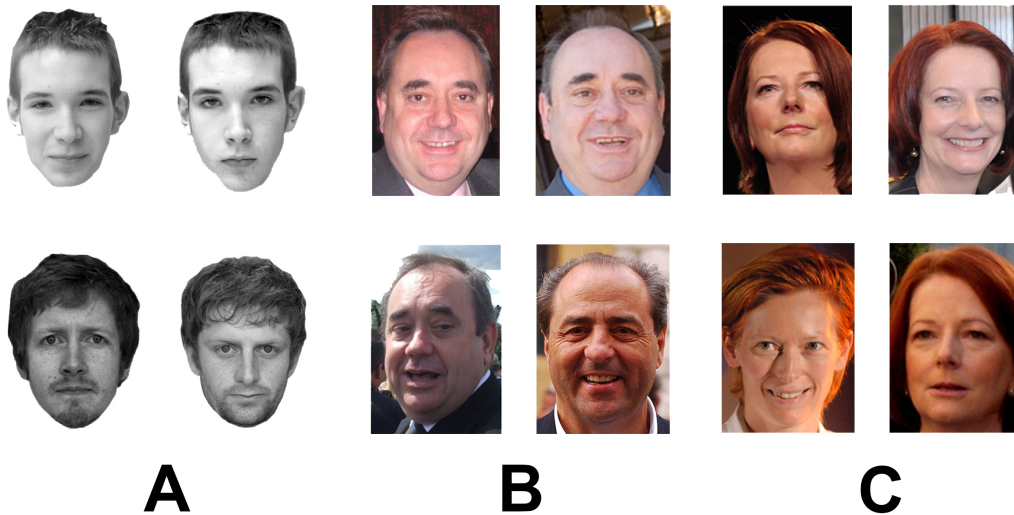
Figure 2. Example image pairs from the different face matching tests. Image pairs on the top row are of the same person, and image pairs on the bottom row are of different people. We tested DP face matching performance with unfamiliar (A: GFMT; B: LHT Unfamiliar celebrities) and familiar faces (C: LHT Familiar celebrities); and with studio-captured (A: GFMT) and ambient images (B & C: LHT).

***Local Heroes Test* (LHT).**

As with the GFMT, the LHT required participants to decide if two simultaneously presented images were of the same person or of two different people. This test was constructed from a set of 40 faces that we expected to be *familiar* to Australian participants (Australian public figures, such as Julia Gillard) and 40 that were *unfamiliar* to these participants (UK public figures, such as Alex Salmond). Importantly, all identities were 'local heroes' such that control participants in the UK were familiar with the UK set but not the Australian set and vice versa. Thus we could examine the benefit of familiarity conferred to DP participants by comparing performance to both Australian and UK control groups (see White et al. 2014b, Experiment 2 for details).

Images in this test were downloaded from the Internet and so are typical of the types of

images retuned by a Google Image search. All images showed a full colour face in roughly frontal pose, with no occlusions, and an inter-ocular distance of at least 100 pixels. These were the only selection criteria. The images were unconstrained with respect to facial (e.g. expression, age), environmental (e.g. lighting, distance-to-camera) and image variables (e.g. camera characteristics). Using these images, we created one *match* and one *mismatch* pair for each face. *Match* pairs were made by pairing two randomly chosen photos of one individual, and *mismatch* pairs were made by pairing randomly chosen photos of two individuals who matched the same basic verbal description (e.g. middle aged male with black hair).

In total, the test comprised of 80 match and 80 mismatch pairs that were presented in a different random order for each participant. To verify DP and control participants' familiarity with the familiar faces, participants then viewed printed names of the Australian and UK celebrities, and classified these as familiar or unfamiliar. Afterwards, participants were again shown the faces and asked to indicate whether the face was familiar or unfamiliar. We calculated internal reliability for the LHT based on data from 96 participants in a previous study (White et al., 2014b) and found reliability to be high (Cronbach's alpha = .834).

RESULTS

1. DP performance on normative tests

Z-scores for individual DP participants were calculated using existing normative data (GFMT: Burton et al., 2010; CFPT: Bowles et al., 2009; CFMT: Bowles et al., 2009; MFFT: Palermo et al., 2011; CCMT: Dennett et al., 2012), and are presented individually and as group summary scores in Figure 3. Overall, z-scores show deficits for DP participants in face memory tasks (MFFT-08; CFMT), and somewhat impaired performance in a standard test of face perception (CFPT). Notably however, group DP performance on the GFMT fell well within the normal range. Further, at the individual level, five of the six participants were less than one standard deviation below normative GFMT performance, suggesting that the ability to match identity of simultaneously presented faces is less impaired in DP when compared to identification tasks that involve memory. Individual performance on the CFPT was

more varied, consistent with previous studies showing that some people with DP are impaired on this task while others are not (e.g. Dalrymple et al., 2014).
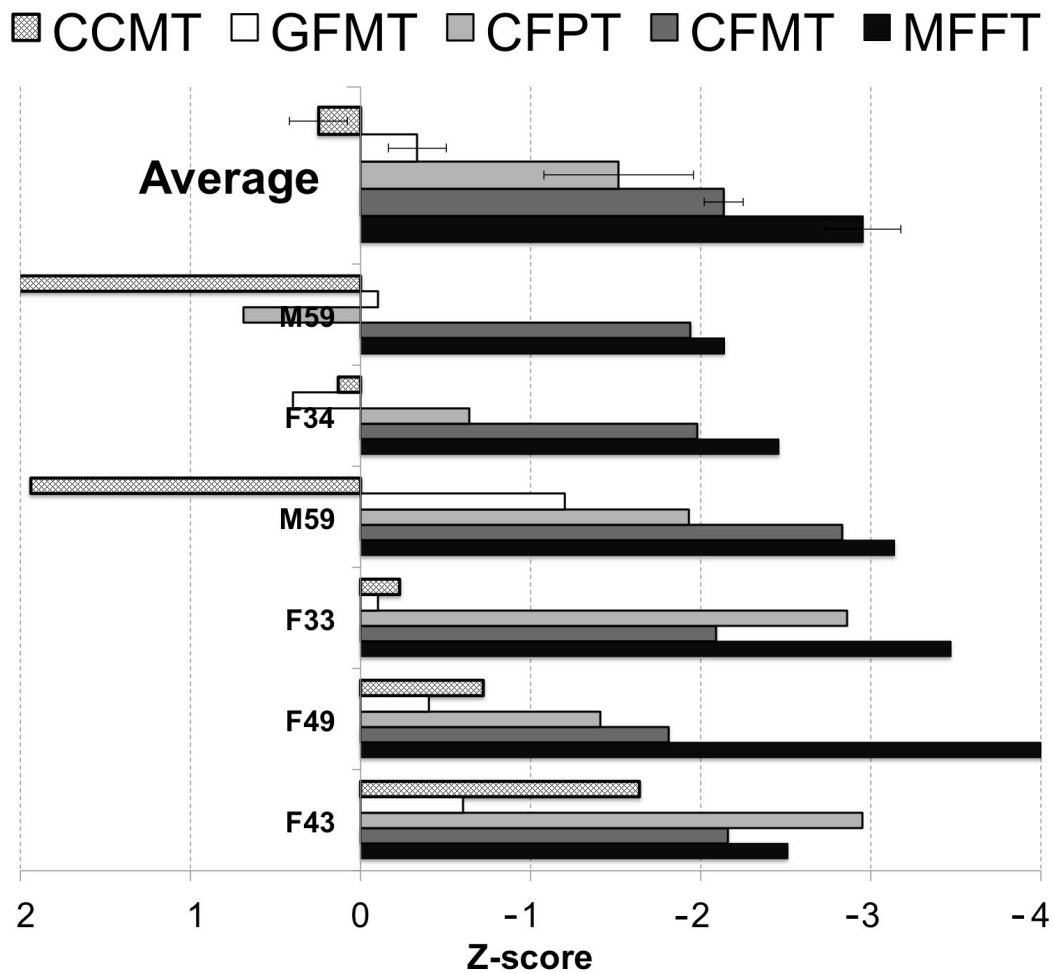
Figure 3. Z-scores for DP participants on standardised tests, computed from normative scores and age-adjusted where appropriate (see text for details). Error bars denote standard error.

2. The Glasgow Face Matching Test (GFMT)

Overall accuracy for the group of six DP participants on the GFMT was 77.9% (SD = 5.1%) and did not differ significantly from normative scores on the test (M = 81.3%; SD = 9.5%; from Burton et al. 2010), [t (198) = 0.85; p > 0.05, Cohen's d = 0.437]. Z-scores for individual DP participants are shown in Figure 3.

To compare performance of DP group to control participants we pooled data of UK and Australian participants, as performance did not differ between these groups [t (21) = 1.32; p > 0.05, Cohen's d = 0.548]. Previous research has shown a dissociation between ability on match and mismatch trials in unfamiliar face matching, raising the possibility that performance on these trial types may be driven by separate cognitive processes (Megreya & Burton, 2007; Attwood, Penton-Voak, Burton & Munafó, 2013). Therefore, when analysing differences between DP and control performance, we included the factor of Trial Type. Summary performance data is shown separately for match and mismatch trials in Table 1.

	Match trials		Mismatch trials	
	Control	*DP*	*Control*	*DP*
	(n = 23)	*(n = 6)*	*(n = 23)*	*(n= 6)*
Accuracy (%)	90.7 (10.1)	79.2 (14.6)	81.7 (17.7)	76.7 (10.8)
Response Time (s)	6.21 (3.56)	6.32 (2.45)	7.69 (4.51)	8.62 (5.45)

Table 1. Accuracy and response time data for DP and Control participants in the Glasgow Face Matching Test (standard deviations in parenthesis).

Accuracy data were analysed by a two-way mixed factor ANOVA with the between-subject factor of Group (Control, DP) and the within-subject factor of Trial Type (Match, Mismatch). This analysis revealed a marginally significant main effect of Group [F (1,27) = 4.21; p = 0.05, η_p^2 = .135], a non-significant main effect of Trial Type [F (1, 27) = 1.34; p > 0.05, η_p^2 = .047] and a non-significant interaction between factors [F (1, 27) = 0.42; p > 0.05, η_p^2 = .015]. Thus, while DP participants were clearly within the normal range on the GFMT, as defined by normative data accompanying the test (Burton et al. 2010), they were nonetheless less accurate compared to matched control subjects who performed the task under the same testing conditions.

Previous studies have shown that DP participants can perform normally on face matching tasks by spending longer on the task (e.g. Behrmann, Avidan, Marotta & Kimchi, 2005; Humphreys et al. 2007), and so we also analysed response time data. However, response times did not differ between groups (details of this analysis are available in Supplementary Materials).

3. Local Heroes Test

3.1 Familiarity with local heroes

Analyses of performance were conducted separately for unfamiliar and familiar faces. For Australian participants (DP and AU control groups), unfamiliar faces were defined as UK celebrities who were categorized as unfamiliar in the name familiarity task, and familiar faces were Australian celebrities categorized as familiar (and vice-versa for UK participants). Trials showing faces that did not meet these predefined criteria were excluded prior to analysis. Familiarity was measured for each individual by showing names of celebrities at the end of the test and asking participants to respond as to whether the person was familiar or unfamiliar. This procedure was then repeated with images of the celebrities. For each participant, unfamiliar faces were defined as celebrities that were not from their country of residence and that were categorised as unfamiliar in the name familiarity task (Control participants: 36; DP participants 34). Familiar faces were celebrities from their country of residence who were categorized as familiar (Control participants: 37; DP participants 25). Thus, DPs were equivalent with unfamiliar classification but were familiar with fewer famous names, which is typical given that face recognition difficulties are often associated with less interest in mass media. Table S2 shows the average number of celebrity names and faces that were familiar to each group (see Supplementary Materials).

3.2 Accuracy

Accuracy data for the LHT are summarised in Figure 4. We analysed accuracy data on the Local Heroes Test by a three-way ANOVA with between subjects factor of Group (DP, control) and within subjects factors of Familiarity (familiar, unfamiliar) and Trial Type (match, mismatch). Because control participants were familiar with different portions of the LHT (i.e. Australian or UK celebrities), this enabled us to confirm: (i) a general benefit of familiarity to matching accuracy; and (ii) that this benefit was not due to idiosyncratic properties of the image pairs used in each portion of the test. However, because UK and Australian control groups were familiar with different items in the test, we also compared the Australian DP group and the Australian control group directly using planned comparison t-tests. This was used to test for predicted DP impairment, separately for familiar and unfamiliar faces.

The main effect of Group was significant [F (1, 27) = 24.6; p < 0.05; η_p^2 = .477], reflecting lower overall accuracy in DP participants (M = 79.7%; SD = 6.81%) compared to controls (M = 87.8%; SD = 8.14%). The main effect of Familiarity was also significant [F (2, 26) = 37.4; p < 0.05; η_p^2 = .581]. However, contrary to our prediction, the interaction between Familiarity and Group was non-significant [F (1, 29) = 1.10; p > 0.05; η_p^2 = .039], reflecting a general benefit of familiarity for all groups (Familiar: M = 89.9%, SD = 7.59%; Unfamiliar: M = 81.1%, SD = 8.60%).
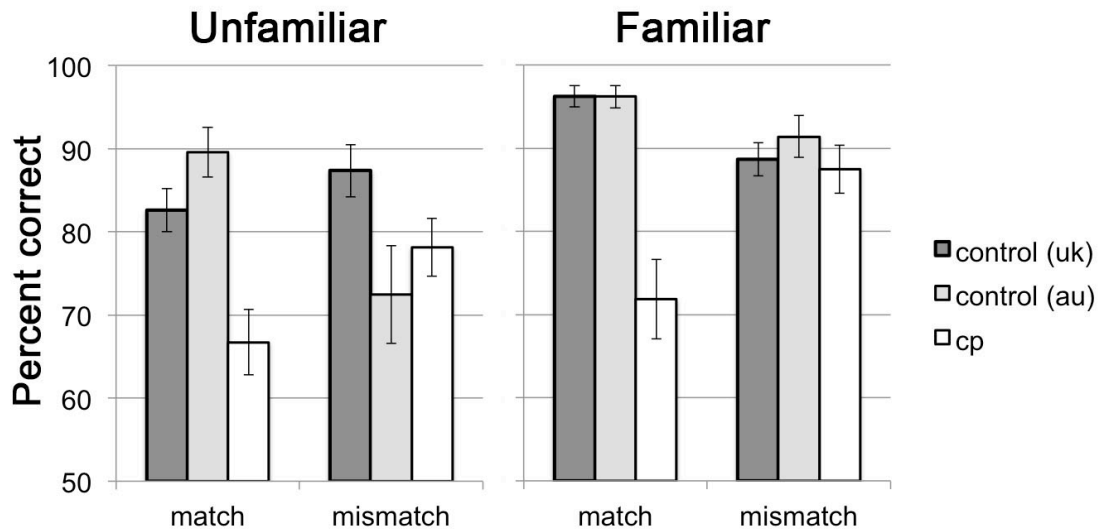
Figure 4. Mean accuracy on familiar and unfamiliar portions of the LHT, separately for match and mismatch trials. Error bars denote standard error.

Analysis also revealed a significant interaction between Trial Type and Group [F (2, 29) = 7.39; p < 0.05; η_p^2 = .215]. Visual inspection of Figure 4 suggested that this interaction was driven by impairment in DP performance for match trials only. Consistent with this interpretation, planned comparisons revealed impairment in DP relative to the Australian control group in match trials but not mismatch trials, both for unfamiliar [match: t (16) = 4.34, p < 0.05, Cohen's d = 2.10; mismatch: t (16) = 0.64, p > 0.05, Cohen's d = 0.35] and familiar face matching [match: t (16) = 5.49, p < 0.05, Cohen's d = 2.29; mismatch: t (16) = 0.90, p > 0.05, Cohen's d = 0.45]. Thus, analysis of accuracy data revealed poorer face matching ability in DP participants, for

both familiar and unfamiliar faces. Interestingly, DP's impairment in the LHT was carried entirely by performance differences in match trials[2].

GENERAL DISCUSSION

We aimed to clarify the nature of perceptual impairment in DP participants with proven deficits in face recognition. Previous studies with similar aims have used perceptual matching tasks that either did not test face identification directly (CFPT, Duchaine et al. 2007), or were constructed using highly constrained photographic capture settings (e.g. BFRT: Benton et al. 1994). To address this we tested DP participants using challenging face identification tasks that do not require a response based on memory. These tasks involved matching identity of photographs captured on the same day in controlled studio conditions (GFMT) and also matching identity across images captured in unconstrained environmental conditions that included natural day-to-day variations in a person's appearance (LHT).

Consistent with previous work (e.g. Dalrymple et al., 2014) the impairment in face perception, as measured by the CFPT, varied considerably across DP individuals. Some DP participants performed like controls on the task and others performed outside the normal range. This pattern of results reinforces the idea that DP is primarily a disorder of memory mechanisms, and that perceptual encoding of face images is often unimpaired in individual cases of DP. However, it is also important to know whether the ability to *identify* faces in the absence of memory constraints is impaired in DP. Contrary to our prediction, results show that accuracy on the GFMT – – a standard test of this ability — was far less variable than CFPT scores, with five of six DPs scoring within one standard deviation of mean performance on this test. Moreover, at the group level, performance of DP participants did not differ

---

[2] Because this pattern is suggestive of a difference in response bias between DP and control participants, we conducted additional analysis of signal detection measures. This analysis shows both reduced sensitivity (d') and more conservative Criterion scores in the DP group, who show a tendency to respond "different". Details of this analysis are available in Supplementary Materials.

significantly from normative performance, although their accuracy was slightly reduced compared to control participants in this study.

Given DPs very poor face identification abilities, the fact that this group achieved typical levels of accuracy on the GFMT suggests that normal performance on this task can be achieved by using cognitive processing strategies that are distinct from those supporting face memory. Indeed, this has been proposed in previous studies to account for the fact that: i) individual differences in familiar face identification does not predict performance in unfamiliar face matching tasks (Megreya & Burton, 2006), and ii) experts in unfamiliar face matching use qualitatively different processes to non-experts on this task (White et al. 2015). The strongest version of this account proposes that matching photographs of unfamiliar faces does not rely on mechanisms specific to face processing at all, but on processes of comparison that are common across stimulus classes (Megreya & Burton, 2006).

While GFMT scores are largely consistent with this proposal, performance data from the LHT show impairment in participants' ability to match identity of face images – for both familiar and unfamiliar faces. A major difference between the GFMT and the LHT is that the latter is created using images that vary substantially with respect to changeable aspects of facial appearance such as lighting, expression and head angle. It is possible that this difference can account for the much larger impairment in this task. This interpretation is also consistent with the pattern of errors observed in this task – whereby the observed impairment was specific to 'match' trials. That is, for both familiar and unfamiliar faces, DP participants made more errors than control participants when the two images showed the same person, but were not impaired relative to controls when images were of different people. In short, DP's did not have difficulty in telling faces apart, but in telling them *together*.

These group differences in match trial accuracy may also be interpreted as changes in bias, with DP participants showing a bias towards responding 'different' while typical participants have a bias towards responding 'same' (for other evidence of a general tendency towards 'same' responses in typical participants see for example: Ozbek & Bindemann, 2011; Alenezi & Bindemann, 2013; Bindemann, Fysh & Johnston, 2015). Signal detection analyses are consistent with this account (see Supplementary

Materials), showing more conservative criterion scores in DP relative to control participants (i.e. DPs were less likely to endorse matches). Interestingly, recent work has shown that a shift in response bias towards more liberal criterion can be induced in unfamiliar face matching tasks, by administering the hormone oxytocin to typical participants (Bate et al., 2015). Conversely, a shift in criterion towards a conservative bias can be induced by inhalation of carbon dioxide, which evokes acute anxiety (Attwood et al., 2013). In this context, it is interesting that oxytocin inhalation has recently been shown to improve DP participants' accuracy in a simultaneous face matching task in which participants had to select a target face from an array of images that always contained the target image (Bate et al., 2014). Future work that examines the underlying causes of criterion shifting in face matching tasks (cf. Menon, White & Kemp, 2015) and the close association between DP impairments and match trial accuracy, may shed light on brain mechanisms supporting face identification.

Also contrary to our predictions was the equivalent familiarity-based enhancement in face matching performance shown by DP and control participants[3]. One possible explanation for this finding is that DP participants used a feature-based comparison strategy in both unfamiliar and familiar face matching tasks, and that this provided an additional route to identification in the case of familiar face matching (where distinctive features were cues to identity). In support of this, previous studies have shown that DP participants can achieve normal levels of accuracy on face memory tests by memorising local features, such as distinctive hairlines and eyebrows (Duchaine & Nakayama, 2004; Duchaine & Weidenfeld, 2003; Stollhoff, Jost, Elze & Kennerknecht, 2010). Importantly, these studies show that DP participants achieve comparable levels of performance by spending longer inspecting the images (e.g. Duchaine, 2000; Nunn, Postma & Pearson, 2001), indicating a more entailed serial processing of facial features (Stollhoff et al., 2010; see also Behrmann et al., 2005).

In the present study, longer response times were also observed in the LHT for DP participants (see Supplementary Materials), and so it appears likely that a similarly

---

[3] This was surprising because DP participants were impaired in forming memory representations of faces, and also because they appear to have less familiarity with celebrity names and so may have had less exposure to the people in general (see Supplementary Materials).

entailed strategy produced the benefit of familiarity observed in the LHT. As such, our results provide some support to previous studies showing that people with DP can use qualitatively different strategies to circumvent more typical routes to identification. Promisingly, some studies have attempted to harness feature-based strategies to develop compensatory training that can enable people with prosopagnosia to identify faces in their daily lives. Although few studies have explored the effectiveness of such training, some have shown gains in identification accuracy by directing attention to diagnostic features (both in Developmental Prosopagnosia: Schmalzl, Palermo & Coltheart, 2008, and in Acquired Prosopagnosia: Powell, Letson, Davidoff, Valentine & Greenwood, 2008).

In parallel to this work, recent studies have also examined the abilities of people with specialist training and expertise in unfamiliar facial identification tasks. Interestingly, 'forensic facial examination' experts – who provide identification evidence in court by comparing photographs of unfamiliar faces – are trained to use feature comparison strategies. Results of a recent study suggest that these forensic examiners adopt a slower and more feature-based strategy than untrained novices, and that this approach confers an additive benefit to face identification accuracy (White et al. 2015). Therefore, future research that aims to develop understanding of the benefits of feature-based processing strategies can improve accuracy of face identification not only in people with DP, but may also benefit people across the broader population that are required to identify unfamiliar faces in their daily work.

In summary, our results show that DP participants were relatively unimpaired on a standard test of face matching ability, suggesting that normal levels of accuracy on the GFMT can be attained independently of deficits in core face recognition ability. This is consistent with accounts of DP proposing a basis in storage and retrieval deficits, and also with the proposal that unfamiliar face matching is less reliant on abstractive levels of representation than familiar face recognition. However, we observed a pronounced deficit in matching faces in the LHT that was specific to match trials, suggesting that people with DP have difficulty in matching identity across natural day-to-day variations in a persons appearance. Future work should aim to establish the causes of this perceptual deficit.

REFERENCES

Alenezi, H. M., & Bindemann, M. (2013). The effect of feedback on face-matching accuracy. *Applied Cognitive Psychology*, *27*(6), 735-753.

Alenezi, H. M., Bindemann, M., Fysh, M. C., & Johnston, R. A. (2015). Face matching in a long task: enforced rest and desk-switching cannot maintain identification accuracy. *PeerJ*, *3*, e1184.

Attwood, A. S., Penton-Voak, I. S., Burton, A. M., & Munafò, M. R. (2013). Acute anxiety impairs accuracy in identifying photographed faces. *Psychological Science*, *24*(8), 1591-1594.

Barton, J. J. (2008). Structure and function in acquired prosopagnosia: lessons from a series of 10 patients with brain damage. *Journal of Neuropsychology*, *2*(1), 197-225.

Bate, S., Cook, S. J., Duchaine, B., Tree, J. J., Burns, E. J., & Hodgson, T. L. (2014). Intranasal inhalation of oxytocin improves face processing in developmental prosopagnosia. *Cortex*, *50*, 55-63.

Bate, S., Bennetts, R., Parris, B. A., Bindemann, M., Udale, R., & Bussunt, A. (2015). Oxytocin increases bias, but not accuracy, in face recognition line-ups. *Social Cognitive and Affective Neuroscience, 10*(7), 1010-1014.

Behrmann, M., Avidan, G., Marotta, J. J., & Kimchi, R. (2005). Detailed exploration of face-related processing in congenital prosopagnosia: 1. Behavioral findings. *Journal of Cognitive Neuroscience*, *17*(7), 1130-1149.

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

Bowles, D. C., McKone, E., Dawel, A., Duchaine, B., Palermo, R., Schmalzl, L., ... & Yovel, G. (2009). Diagnosing prosopagnosia: Effects of ageing, sex, and participant–stimulus ethnic match on the Cambridge Face Memory Test and Cambridge Face Perception Test. *Cognitive Neuropsychology, 26*(5), 423-455.

Bruce, V. (1994). Stability from variation: the case of face recognition the MD Vernon memorial lecture. *The Quarterly Journal of Experimental Psychology*, *47*(1), 5-28.

Bruce, V., Henderson, Z., Greenwood, K., Hancock, P. J. B., Burton, A. M., & Miller, P. (1999). Verification of face identities from images captured on video. *Journal of Experimental Psychology: Applied, 5*(4), 339-360.

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

Burton, A. M. (2013). Why has research in face recognition progressed so slowly? The importance of variability. *The Quarterly Journal of Experimental Psychology*, *66*(8), 1467-1485.

Burton, A. M., White, D., & McNeill, A. (2010). The Glasgow face matching test. *Behavior Research Methods, 42*(1), 286-291.

Burton, A. M., Schweinberger, S. R., Jenkins, R., & Kaufmann, J. M. (2015). Arguments against a configural processing account of familiar face recognition. *Perspectives on Psychological Science*, *10*(4), 482-496.

Chatterjee, G., & Nakayama, K. (2012). Normal facial age and gender perception in developmental prosopagnosia. *Cognitive Neuropsychology*, *29*(5-6), 482-502.

Clutterbuck, R., & Johnston, R. A. (2004). Matching as an index of face familiarity. *Visual Cognition*, *11*(7), 857-869.

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

Dalrymple, K. A., Oruc, I., Duchaine, B., Pancaroglu, R., Fox, C. J., Iaria, G., ... & Barton, J. J. (2011). The anatomic basis of the right face-selective N170 IN acquired prosopagnosia: a combined ERP/fMRI study. *Neuropsychologia*, *49*(9), 2553-2563.

Dalrymple, K. A., & Palermo, R. (2016). Guidelines for studying developmental prosopagnosia in adults and children. *Wiley Interdisciplinary Reviews: Cognitive Science, 7,* 73–87.

DeGutis, J., Cohan, S., & Nakayama, K. (2014). Holistic face training enhances face processing in developmental prosopagnosia. *Brain, 137* (6), 1781–1798.

Dennett, H. W., McKone, E., Tavashmi, R., Hall, A., Pidcock, M., Edwards, M., & Duchaine, B. (2012). 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.

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

Duchaine B. C. (2000). Developmental prosopagnosia with normal configural processing. *NeuroReport*, *11*, 79–83.

Duchaine, B. C., 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. C., & Nakayama, K. (2004). Developmental prosopagnosia and the Benton Facial Recognition test. *Neurology*, *62*(7), 1219-1220.

Duchaine, B. C., & Nakayama, K. (2006). 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. C., & Weidenfeld, A. (2003). An evaluation of two commonly used tests of unfamiliar face recognition. *Neuropsychologia*, *41*(6), 713-720.

Huis in't Veld, E., Van den Stock, J., & de Gelder, B. (2012). Configuration perception and face memory, and face context effects in developmental prosopagnosia. *Cognitive Neuropsychology*, *29*(5-6), 464-481.

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.

Jenkins, R., & Burton, A.M. (2011). Stable face representations. *Philosophical Transactions of the Royal Society B, 366*, 1671-1683.

Jenkins, R., White, D., Van Montfort, X., & Burton, A. M. (2011). Variability in photos of the same face. *Cognition, 121*(3), 313-323.

Megreya, A. M., & Bindemann, M. (2015). Developmental improvement and age-related decline in unfamiliar face matching. *Perception*, *44*, 5-22.

Megreya, A. M., & Burton, A. M. (2006). Unfamiliar faces are not faces: Evidence from a matching task. *Memory and Cognition, 34*(4), 865-876.

Megreya, A. M., & Burton, A. M. (2007). Hits and false positives in face matching: A familiarity-based dissociation. *Perception and Psychophysics, 69*(7), 1175-1184.

Menon, N., White, D. & Kemp, R. I. (2015). Identity-level representations affect unfamiliar face matching performance in sequential but not simultaneous tasks. *Quarterly Journal of Experimental Psychology, 68* (9), 1777-1793.

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

O'Toole, A. J., An, X., Dunlop, J., Natu, V., & Phillips, P. J. (2012). Comparing face recognition algorithms to humans on challenging tasks. *ACM Transactions on Applied Perception (TAP), 9*(4), 16.

Özbek, M., & Bindemann, M. (2011). Exploring the time course of face matching: Temporal constraints impair unfamiliar face identification under temporally unconstrained viewing. *Vision Research, 51*(19), 2145-2155.

Palermo, R., Rivolta, D., Wilson, C.E., & Jeffery, L. (2011). Adaptive face space coding in congenital prosopagnosia: Typical figural aftereffects but abnormal identity aftereffects. *Neuropsychologia., 49*(14), 3801-3812.

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.

Powell, J., Letson, S., Davidoff, J., Valentine, T., & Greenwood, R. (2008). Enhancement of face recognition learning in patients with brain injury using three cognitive training procedures. *Neuropsychological rehabilitation, 18*(2), 182-203.

Rivolta, D., Palermo, R., & Schmalzl, L. (2013). What is overt and what is covert in congenital prosopagnosia? *Neuropsychology Review, 23*(2), 111-116.

Rivolta, D., Palermo, R., Schmalzl, L., & Coltheart, M. (2012). Covert face recognition in congenital prosopagnosia: a group study. *Cortex, 48*(3), 344-352.

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

Stollhoff, R., Jost, J., Elze, T., & Kennerknecht, I. (2010). The early time course of compensatory face processing in congenital prosopagnosia. *PLoS One, 5*(7), e11482.

Susilo, T., & Duchaine, B. (2013). Advances in developmental prosopagnosia research. *Current Opinion in Neurobiology, 23*(3), 423-429.

Tippett, L. J., Miller, L. A., & Farah, M. J. (2000). Prosopamnesia: A selective
impairment in face learning. *Cognitive Neuropsychology*, *17*(1-3), 241-255.

White, D., Kemp, R. I., Jenkins, R., Matheson, M. & Burton, A. M. (2014a). Passport
Officers' errors in face matching. *PLoS One* 9(8): e103510.
doi:10.1371/journal.pone.0103510

White, D., Kemp, R. I., Jenkins, R., & Burton, A. M. (2014b). Feedback training for
facial image comparison. *Psychonomic Bulletin and Review, 21*(1), 100-106.

White, D., Burton, A. M., Jenkins, R., & Kemp, R. I. (2014c). Redesigning photo-ID
to improve unfamiliar face matching performance. *Journal of Experimental
Psychology: Applied*, *20*(2), 166.

White, D., Dunn, J. D., Schmid, A. C. & Kemp, R. I. (2015a). Error Rates in Users of
Automatic Face Recognition Software. PLoS ONE 10(10): e0139827.

White, D., Phillips, P. J., Hahn, C. A., Hill, M., & O'Toole, A. J. (2015b). Perceptual
expertise in forensic facial image comparison. *Proceedings of the Royal
Society of London B: Biological Sciences*, *282,* 1814-1822.

Young, A. W., Newcombe, F., Haan, E. H. D., Small, M., & Hay, D. C. (1993). Face
perception after brain injury. *Brain*, *116*(4), 941-959.

Supplementary Materials for "Face matching in developmental prosopagnosia"

David White*, A. Mike Burton, Davide Rivolta, Shahd Al-Janabi, Romina Palermo

*Corresponding author: david.white@unsw.edu.au

Raw scores for diagnostic tests

	MFFT		CFMT		CFPT		CCMT	
Sex/ Age	Percent Correct	Z-score	No. Correct	Z-score	No. Errors	Z-score	No. Correct	Z-score
F43	35.3	-2.51	37	-2.16	68	-2.95	41	-1.64
F49	6.3	-4.00	39	-1.81	52	-1.41	45	-0.72
F33	23.1	-3.47	38	-2.09	66	-2.86	63	-0.23
M59	15.0	-3.14	28	-2.83	70	-1.93	35	1.94
F34	40.0	-2.46	39	-1.98	40	-0.64	59	0.13
M59	30.8	-2.14	35	-1.93	38	0.69	67	2.01

Table S1. Percent correct on diagnostic memory tests for DP participants. Raw scores for the CFMT and CCMT are out of a possible 72. CFPT scores are expressed as errors with a chance-level upper bound of 93.

Face and name familiarity (LHT)

	Name Familiarity		Face Familiarity	
	Familiar	*Unfamiliar*	*Familiar*	*Unfamiliar*
UK	38.0 (1.2)	37.0 (2.3)	32.7 (7.1)	37.2 (2.3)
AU	35.4 (3.7)	36.3 (3.8)	36.3 (3.8)	30.2(12.9)
DP	30.2 (8.7)	36.4 (7.7)	23.9(11.2)	35.6 (5.9)

Table S2: Summary familiarity data for unfamiliar and familiar portions of the Local Hero Test (standard deviation in parenthesis).

Response time analysis (GFMT and LHT)

Mean response times for items in the Glasgow Face Matching Test (GFMT; see Table 1) revealed a non-significant main effect of Group (F < 1). The main effect of Trial Type was significant [F (1,27) = 2.72; p < 0.05, η_p^2 = .081], reflective of slower response times in mismatch (M = 7.88 seconds; SD = 4.63) compared to match trials (M = 6.23 seconds;

SD = 3.32). The interaction between factors was non-significant (F < 1). Thus, DP participants spent an equivalent amount of time performing the GFMT as controls.

Mean response time data for the Local Heroes Test are shown in Figure S1. These data were analysed to test whether DP performance in the LHT was supported by lengthier processing of face stimuli, using a three-way ANOVA with a between subjects factor of Group (DP, AU control, UK control) and within subjects factors of Familiarity (familiar, unfamiliar) and Trial Type (match, mismatch). The main effect of Group was non-significant [$F_{(1, 27)} = 3.24$; $p > 0.05$; $\eta_p^2 = .107$]. Non-significant main effects of Familiarity [$F_{(1, 27)} = 3.24$; $p > 0.05$; $\eta_p^2 = .107$] and Trial Type (F < 1) were also observed.
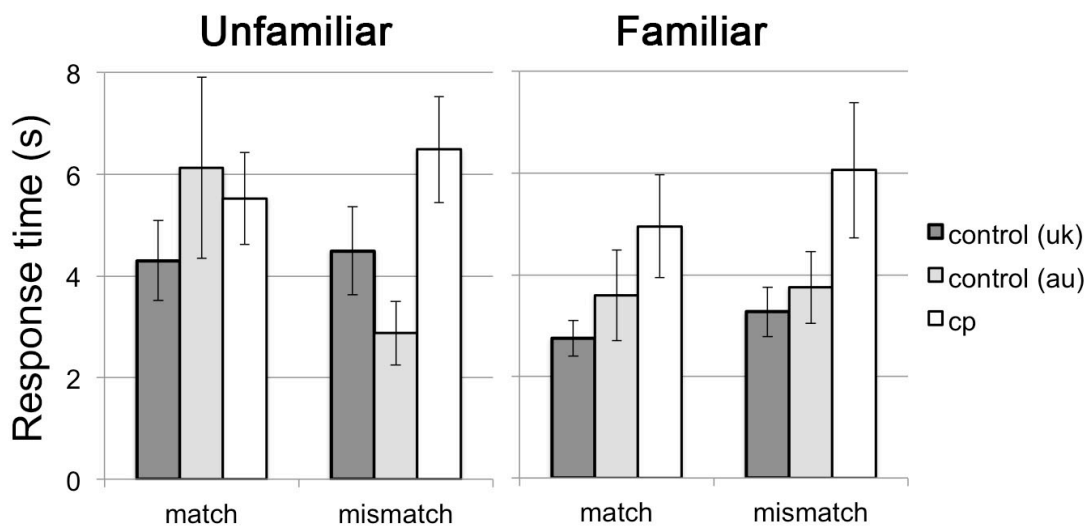

Figure S1. Mean response times on familiar and unfamiliar portions of the LHT, separately for match and mismatch trials. Error bars denote standard error.

Main effects were qualified by a significant interaction between Group and Trial Type [$F_{(1, 27)} = 4.61$; $p < 0.05$; $\eta_p^2 = .146$]. Analysis of Simple Main Effects revealed that this interaction was driven by slower response times by DP participants in mismatch trials (M = 6.93; SE = 1.08) compared to control participants (M = 5.78; SE = 1.41) [$F_{(1, 27)} = 7.68$; $p < 0.05$], but no difference between response times for match trials (DP: M = 3.58; SE = 0.55; Control: M = 4.22; SD = .072; F < 1).

As with accuracy data, the interaction between group and familiarity was non-significant [$F (2, 27) = 1.33$; $p > 0.05$; $\eta_p^2 = .047$]. The three-way interaction between factors was also non-significant [$F (2, 27) = 1.34$; $p > 0.05$; $\eta_p^2 = .047$].

Signal detection analysis (GFMT and LHT)

In both GFMT and Local Heroes tests, DPs were impaired on match, but not mismatch, trials. This result is consistent with a difference in response bias in DP participants. Therefore we analysed sensitivity (d') and criterion (C) for both the GFMT and the LHT. Summary data for the signal detection analysis are shown in Table S3. Analysis of sensitivity data (d-prime) for the GFMT revealed significantly higher sensitivity in control participants ($M = 2.50$; $SD = 0.80$) compared to DP participants [$t (31) = 2.47$, $p < 0.05$, Cohen's d = 1.30]. The difference between Criterion (C) scores for DP and control groups ($M = -0.15$; $SD = 0.45$) was non-significant [$t (31) = 0.35$, $p < 0.05$, Cohen's d = 0.16]

	Sensitivity (d-prime)			Response bias (C)		
	GFMT	LHT (familiar)	LHT (unfamiliar)	GFMT	LHT (familiar)	LHT (unfamiliar)
UK	2.76 (0.78)	2.86 (0.68)	2.33 (0.59)	-.20 (.42)	-.21 (.45)	.20 (.46)
AU	2.27 (0.77)	3.41 (0.68)	2.23 (0.74)	-.10 (.50)	-.16 (.37)	-.37 (.48)
DP	1.69 (0.37)	1.89 (0.53)	1.28 (0.38)	-.08 (.44)	.30 (.40)	.18 (.31)

Table S3. Signal detection measures for DP and Control participants in the face matching tests (standard deviations in parenthesis).

For sensitivity data on the LHT, main effects of group [$F(1, 27) = 19.2$; $p < 0.05$, $\eta_p^2 = .416$], and familiarity [$F(1, 29) = 33.9$; $p < 0.05$, $\eta_p^2 = .556$] were significant. interaction between factors was non-significant [$F(2, 29) = 1.43$; $p > 0.05$, $\eta_p^2 = .050$]. For criterion data, the main effect of Group was significant [$F(1, 27) = 6.63$; $p < 0.05$, $\eta_p^2 = .197$], reflective of a more conservative response bias in the DP group (i.e. less likely to respond 'same'). The main effect of Familiarity ($F < 1$) and the interaction [$F(1, 27) = 1.29$; $p < 0.05$, $\eta_p^2 = .046$] were non-significant.