CORTEX 157 (2022) 266–273

Available online at www.sciencedirect.com

ScienceDirect

Journal homepage: www.elsevier.com/locate/cortex

Research Report

Individuals with developmental prosopagnosia show independent impairments in face perception, face memory and face matching

Mirta Stantić [a,1], Zoë Pounder [a,*,1], Sarah Bate [b], Tirta Susilo [c], Caroline Catmur [d] and Geoffrey Bird [a]

[a] Department of Experimental Psychology, University of Oxford, UK
[b] Department of Psychology, Bournemouth University, UK
[c] School of Psychology, Victoria University of Wellington, New Zealand
[d] Department of Psychology, King's College London, UK

ARTICLE INFO

Article history:
Received 21 March 2022
Reviewed 20 May 2022
Revised 8 July 2022
Accepted 21 September 2022
Action editor Stefan Schweinberger
Published online 13 October 2022

Keywords:
Face perception
Face matching
Face memory
Developmental prosopagnosia

ABSTRACT

Individuals with developmental prosopagnosia (DP) all exhibit impairments in face memory, but the specificity of these face memory impairments is debated. One problem is that standard behavioural tasks are not able to provide independent measurement of face perception, face memory, and face matching (the decision process required to judge whether two instances of a face are of the same individual or different individuals). The present study utilised a new test of face matching, the Oxford Face Matching Test (OFMT), and a novel analysis strategy to derive these independent indices. Twenty-nine individuals with DP and the same number of matched neurotypical controls completed the OFMT, the Glasgow Face Matching Test, and the Cambridge Face Memory Test. Results revealed individuals with DP exhibit impairments in face perception, face memory *and* face matching. Collectively, these results suggest that face processing impairments in DP are more comprehensive than has previously been suggested.

1. Introduction

Developmental prosopagnosia (DP) is a neurodevelopmental condition characterised by a lifelong inability to recognize faces (e.g., Behrmann & Avidan, 2005; Cook & Biotti, 2016; Susilo & Duchaine, 2013). While DP is, by definition, associated with problems remembering faces, the particular aspect of face processing responsible for these memory problems is unclear. Specifically, there is an ongoing debate as to whether individuals with DP are able to form intact perceptual representations of faces but have difficulty learning/recalling facial identities ('*memory hypothesis*'; Jackson, Counter, & Tree, 2017; Stollhoff, Jost, Elze, & Kennerknecht, 2011), or whether

* *Corresponding author.*
E-mail address: z.pounder@psy.ox.ac.uk (Z. Pounder).
[1] These authors contributed equally.
https://doi.org/10.1016/j.cortex.2022.09.012

individuals with DP have difficulties with both forming perceptual representations of faces and face memory ('*perceptual hypothesis*'; Biotti, Gray, & Cook, 2019; Dalrymple, Garrido, & Duchaine, 2014, see also Towler, Fisher, & Eimer, 2018). This debate is ongoing partly due to the heterogeneity of DP, specifically the possibility that there may be subtypes of DP (e.g., Dalrymple, Corrow, Yonas, & Duchaine, 2012) or variants (e.g., apperceptive or associative) that differ in terms of the nature of their face-processing impairment (Biotti et al., 2019; Corrow, Dalrymple, & Barton, 2016). Nevertheless, little attention has been paid in DP to a third process, face matching, which is also necessary for face recognition. It is worth noting that the term 'Face Matching' is used here in a very specific, and perhaps unusual, way. In contrast to the standard use of the term face matching, which refers to a set of tasks in which participants are required to judge whether two photographs of a face depict the same individual or different individuals, face matching in this context relates to the decision-making process necessary to determine whether two or more face images are of the same individual, or different individuals. Face matching is the psychological process necessary for successful performance on face matching tasks. Face matching is also required, however, when a participant is asked to determine whether a face stimulus matches that stored in memory, such as when deciding whether a photograph of a face matches a recently-learned identity. Although the former task is called a face matching task, and the latter called a face memory or face recognition task, the decision-making process (deciding whether two face instances are of the same individual or different individuals, regardless of whether one of those instances is stored in memory) is the same, and what we call face matching.

Problematically, existing tests are largely unable to isolate these distinct face processes. For example, whether individuals with DP are impaired at face memory is normally tested using the Cambridge Face Memory Test (CFMT; Duchaine & Nakayama, 2006), a task in which participants have to identify a pre-learned facial identity among distractors. Although face memory is required in order to perform accurately on this task, face perception and face matching are also required. Therefore, it is unclear whether poor performance on the CFMT reflects impaired face memory, face perception or face matching, or any combination of these processes. Similarly, face perception is often assessed using face matching tasks such as the Glasgow Face Matching Test (GFMT; Burton, White, & McNeill, 2010; see also Kent Facial Matching Test; Fysh & Bindemann, 2018), in which two facial images are presented and participants are required to judge whether the facial images are of the same person or different people. Matching tasks require intact face perception and face matching, and so impaired performance on matching tasks in DP may reflect either one, or both, of impaired face perception and face matching. When face matching tasks are not used to assess face perception, the Cambridge Face Perception Task (CFPT; Duchaine, Germine & Nakayama, 2007) is often used. This test requires participants to order a set of test face stimuli in terms of their similarity to a target face, where those test stimuli are morphed faces containing varying proportions of the target face and a foil face. The contribution of face matching to performance on the CFPT is therefore hard to determine, as increasingly similar faces are achieved by the test face becoming increasingly the same face as the target face, due to the test face containing objectively more of the target face in the morph. Conversely, decreasingly similarity is achieved by the test face containing objectively less of the target face. Thus, responses of a participant basing their judgement on objective face perceptual similarity only, and a participant basing their responses on the outcome of a face matching decision process, would be perfectly correlated. Ideally, perceptual similarity would be unconfounded from facial identity, so that facial images of the same person are sometimes less similar (due to ageing, weight gain etc.), than facial images of different people. What has been lacking thus far, therefore, is a means to obtain independent measures of face perception, face matching, and face memory, in order to determine the degree to which these are affected in DP.

The Oxford Face Matching Test (OFMT; Stantić et al., 2021) is a novel test designed so that it can be used to assess individual differences in face processing abilities in clinical and non-clinical populations in a non-biased manner (Stantić et al., 2021; Stantić et al., under review). Of relevance to the current study is that independent measures of face perception and face matching can be obtained from the OFMT, and when these scores are used to partition variance in CFMT scores, an independent measure of face memory can also be derived. This approach has previously been used to show that, in neurotypical individuals, face perception contributes to performance on face matching tests, and that face perception and face matching make independent contributions to CFMT performance. In addition, when this approach was used with volunteers with autism,[2] results suggested that autism was associated with deficits in face perception and face memory, but not face matching (Stantić et al., under review).

Accordingly, the current study uses the OFMT and CFMT to derive independent measures of face perception, face matching, and face memory in a group of adults with DP and a matched neurotypical control group, such that the nature of face processing impairment(s) in DP can be identified.

2. Methods

We report how we determined our sample size, all data exclusions, all inclusion/exclusion criteria, whether inclusion/exclusion criteria were established prior to data analysis, all manipulations, and all measures in the study.

2.1. Participants

Twenty-nine developmental prosopagnosics were recruited to participate in the study (9 males; M_{age} = 42.31, SD = 11.31). Participants were selected as DPs from author databases of individuals with DP. These participants met the criteria for impaired performance (defined as 2 SDs below the neurotypical mean score) on at least two of three face processing

[2] To respect the wishes of autistic individuals and report the study in line with scientific parlance, we use language preferred by clinical professionals (e.g., 'individuals with autism'), as well as the term 'autistic', a term endorsed by many individuals with ASD (see Kenny et al., 2016).

measures (CFMT, CFPT, and the Famous Face Test (Bobak, Parris, Gregory, Bennetts, & Bate, 2017; Duchaine & Nakayama, 2006; Duchaine et al., 2007). No DP participants were excluded from the current study for failing the attention check trials on the OFMT (see below). On the PI-20 (see below), participants with DP scored a mean of 83.72 (SD = 5.80). An age- and gender-matched sample of 31 neurotypical participants were recruited via Prolific. co and the authors' database. Two participants were excluded for failing to pass attention checks on the OFMT, providing a final sample of 29 neurotypical participants (11 males; $M_{age} = 42.41$, $SD = 9.81$). On the PI-20, neurotypical controls scored a mean of 45.69 (SD = 9.97). The DP and neurotypical groups did not differ significantly in terms of age [$t(56) = .04$, $p = .97$, $d = .01$] or gender [$X^2(1) = .31$, $p = .58$, $w = .07$], but, as expected, the DP group self-reported more problems with face recognition on the PI-20 than the neurotypical group ($U = 0$, $p = < .001$, $r = .86$). All participants reported having normal or corrected-to-normal vision. Ethical approval was obtained from the Central University Research Ethics Committee, University of Oxford. The procedures used in this study adhere to the tenets of the Declaration of Helsinki.

2.2. Procedure

In a randomized order, participants completed the PI-20 and three measures of face processing ability: two matching paradigms, the OFMT (Stantić et al., 2021) and the GFMT (Burton et al., 2010), as well as a face memory paradigm, the CFMT (Duchaine & Nakayama, 2006). Participants for whom the CFMT scores were available from previous testing (specifically, DPs who completed the CFMT as part of their prosopagnosia screening) did not complete the test again and instead the existing score was used to avoid practice effects. All tasks were undertaken using the online behavioural platform Gorilla (www.gorilla.sc).

2.3. The 20-item prosopagnosia index (PI-20; Shah, Gaule, Sowden, Bird, & Cook, 2015)

The PI-20 is a self-report measure of face recognition ability and comprises 20 items whereby participants are asked to rate on a Likert scale (1 = strongly disagree to 5 = strongly agree) their face recognition difficulties in everyday life (e.g., '*I have always had a bad memory for faces*'). Five questions are reverse scored, and the total calculated by summing the scores from all items. The PI-20 can be used in conjunction with the CFMT to identify individuals with DP (Gray, Bird, & Cook, 2017); note that the PI-20 was not used to identify DPs in this study). The maximum possible score is 100. The PI-20 is publicly available from: %https://royalsocietypublishing.org/action/downloadSupplement?doi=10.1098 2Frsos.140343&file=rsos140343supp1.pdf.

2.4. Oxford Face Matching Test (OFMT; Stantić et al., 2021)

The OFMT (Figure 1A) is a novel face matching task that contains 200 trials (100 match (same) and 100 mismatch (different) face pairs). As an attention check, the OFMT contains an additional 12 trials that are designed to be answered correctly even by individuals with severe face processing impairments. Participants were excluded from all analyses if

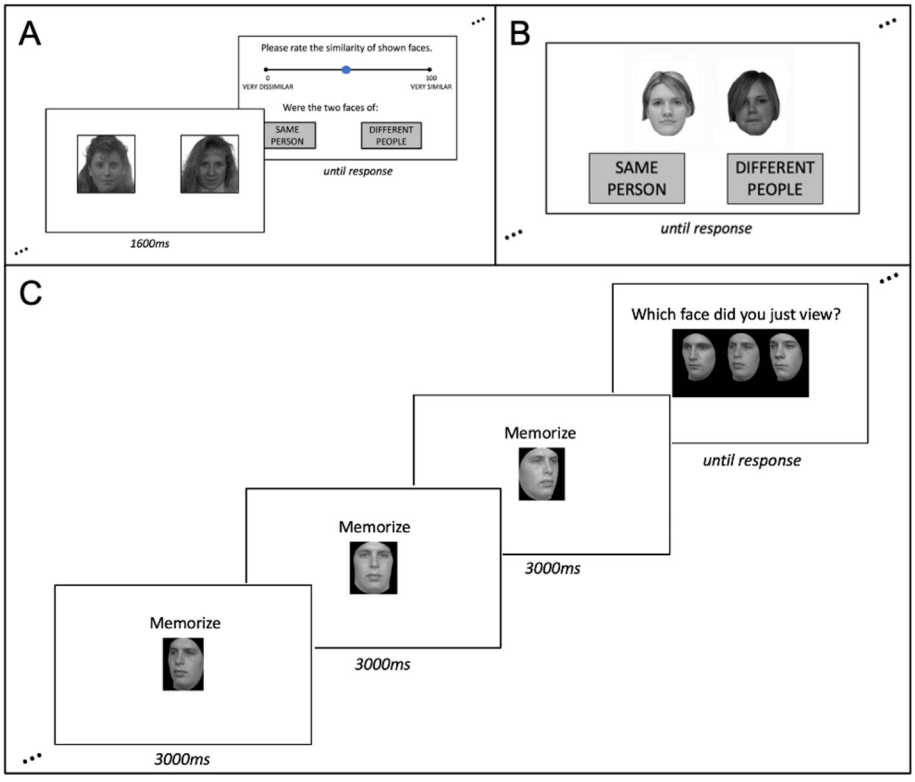

Fig. 1 – A sample trial of three face processing tasks: A – the Oxford Face Matching Test (OFMT); B – the Glasgow Face Matching Test, GFMT, and C – the Cambridge Face Memory Test, CFMT.

they answered two or more of these trials incorrectly. Participants are presented with a face pair for 1600 ms and asked to determine whether faces are of the same person or different people. The maximum possible matching score is 200. In addition, participants provide a perceptual similarity judgment for each pair of faces on a scale from 0 to 100 (from very dissimilar to very similar). The OFMT is deliberately constructed such that faces in match and mismatch trials contain overlapping similarity distributions – two images of the same person can be perceptually markedly different, while images of two different individuals can be perceptually very similar. Thus, perceptual similarity can be dissociated from the outcome of a face matching process. The OFMT is available to researchers on the Gorilla Open Materials repository (https://gorilla.sc/openmaterials/134286) for non-commercial use upon request.

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

The GFMT (Figure 1B) is an established face matching task that contains 40 trials (20 match and 20 mismatch). Participants are presented with face pairs and can view them for an unlimited amount of time before making a decision about whether the faces are the same or different. The maximum possible score is 40. We lack legal permission to publicly archive the code and materials for the GFMT, however, interested readers can contact the owners of the task in the cited references.

2.6. *Cambridge face memory task (CFMT; Duchaine & Nakayama, 2006)*

The CFMT (Figure 1C) is an established face memory task that contains 72 trials in three stages of increasing difficulty. Participants initially learn six faces and are afterwards tested on three-alternative-forced-choice trials with two distractors and one image of a previously learnt identity. There are 18 trials with no changes to viewpoint or lighting, 30 trials with changes to viewpoint and lighting, and 24 trials with changes to viewpoint and lighting as well as the addition of visual noise. The maximum possible score is 72. We lack legal permission to publicly archive the code and materials for the CFMT, however, interested readers can contact the owners of the task in the cited references.

2.7. *Analysis strategy*

Independent measures of face perception, face matching and face memory are required to address the aims of the study. To derive a measure of face *perception*, participants' ratings of the similarity of face pairs on the OFMT were compared to similarity ratings derived from the average of three leading facial recognition algorithms (AWS Rekognition (https://aws.amazon.com/rekognition/), FaceSoft (retrieved from http://facesoft.io/) and Azure Face Recognition (https://azure.microsoft.com/en-us/services/cognitive-services/face/), see Stantić et al., 2021). Each of these algorithms provide a similarity index from 0 to 100, from which a mean index of similarity can be calculated. For each participant, an average absolute deviation from algorithmically-provided similarity was calculated. This value represents the difference between a participant's similarity score and the average similarity score provided by the algorithms (i.e., the higher the deviation score, the greater the difference between the average similarity value provided by the algorithms and the value provided by the participant). Participant similarity ratings were compared with algorithmic similarity ratings for two reasons: the first is that such algorithms regularly outperform human observers (Phillips & O'Toole, 2014; Phillips et al., 2018) implying that their similarity ratings are valid, and the second is that the use of algorithms (rather than large groups of human raters) to determine similarity avoids a systemic bias towards whichever group rates the stimuli.

An index of face *matching* independent of face perception was derived by regressing OFMT accuracy scores on average algorithmic deviations across participants. The residuals from this regression constitute face matching scores as they represent the variance in OFMT accuracy (i.e., in same/different judgements) that cannot be explained by perceptual face similarity judgements. These face matching scores (the residuals from the above regression) are thus statistically independent from face perceptual similarity judgements due to the way they are derived. Face matching scores can also be derived from GFMT scores, if GFMT scores are regressed on average algorithmic deviations obtained from the OFMT. This double coding of face matching ability provides a data robustness check – allowing the replicability of results to be ascertained when using matching scores from the GFMT after controlling for OFMT perceptual similarity judgements. Note that this robustness check stems from the independence of the two judgments made on each trial of the OFMT. The binary same/different decision about two faces, akin to the one made on each trial of the GFMT, constitutes an independent data point from the judgment of perceptual similarity of the two presented faces. The latter can therefore be used as an index of participants' perceptual ability independent of their face matching ability. In the current study, we use these residuals (i.e., the measure of face matching independent of face perception accuracy) in univariate analyses, e.g., in post-hoc tests. In multivariate analyses, entering OFMT or GFMT test scores and algorithmic deviations into the same analysis achieves the same aim (the variance explained by face matching performance can be identified independently of that accounted for by face perception—i.e., when face perception ability is held constant).

Finally, face *memory* scores that are independent of both face perception and face matching can be obtained by regressing CFMT accuracy scores on average algorithmic deviations from the OFMT and face matching test scores from either the OFMT or GFMT.

In the current study, we determined whether DP impacts face perception, face matching (independent of face perception) and face memory (independent of face perception and face matching) using a series of regression analyses in which group (DP vs Control), predictors (e.g., face perception), and their interaction, were used to predict test scores (e.g., CFMT test scores). Including the interaction term in regression models allows for the relationship between, e.g., face perception (i.e., algorithmic deviation) and CFMT scores, to

vary across groups. Data was analysed using SPSS Statistics Version 28. For between group comparisons (DP vs Control), participant demographics and task performance were analysed with independent *t*-tests or the non-parametric equivalent, the Mann Whitney test, when normality assumptions were violated. All statistical analyses were performed with a significance level of $p < .05$, and all p values are two-tailed. All data are available at https://osf.io/vzsqd and no part of the study procedures or analysis was formally preregistered (though these are the same as used in Stantić et al., under review).

3. Results

3.1. Group comparisons – standard test scores

3.1.1. OFMT: matching performance

Nine participants with DP (31%) performed two standard deviations below the Control group's mean score, and an additional 13 participants with DP (45%) performed one standard deviation below the Control group mean. One DP had an OFMT score that was greater than the neurotypical mean score (but less than one standard deviation above the control mean). At the group level, an independent *t*-test showed a significant difference in OFMT accuracy [$t(56) = 5.90$, $p < .001$, $d = 1.55$] with DPs less accurate ($M = 66\%$, $SD = 6\%$, range: 55%–78%) than control participants ($M = 74\%$, $SD = 5\%$, range: 65%–84%).

Signal Detection Theory (Green & Swets, 1966) was used to characterise performance on the matching component of the OFMT, providing a measure of sensitivity (d') and bias (criterion). An independent *t*-test showed a significant difference in d' [$t(56) = 5.40$, $p < .001$, $d = 1.42$] and criterion [$t(56) = 3.21$, $p = .002$, $d = .84$] values between the DP and Control groups, with the DP group showing less sensitivity and an increased bias (DP d': $M = .91$, $SD = .31$; criterion: $M = -.36$, $SD = .33$; Control d': $M = 1.38$, $SD = .35$; criterion: $M = -.06$, $SD = .38$). Thus, individuals with DP exhibited less sensitivity to signals relevant to face matching, and needed more evidence that the faces were the same before they made this response.

3.1.2. OFMT: algorithmic deviation

Nine participants with DP (31%) performed two standard deviations above (indicating worse performance) the Control mean deviation score, while an additional 11 participants with DP (38%) performed one standard deviation above the Control mean. Eight participants with DP (28%) had a deviation score that was greater than the neurotypical deviation mean (but less than one standard deviation). At group level, a Mann–Whitney test revealed a significant difference in deviation scores ($U = 105$, $p = < .001$, $r = .64$) with the DP group having worse deviation scores, i.e., higher deviation from algorithmic judgements (median = 26.69, range: 22.81–48.28), than the Control group (median = 23.57, range: 16.66–27.38).

3.1.3. GFMT

Nine participants with DP (31%) performed two standard deviations below the Control mean score, while an additional 11 DPs (38%) performed one standard deviation below the Control mean. Five participants with DP (17%) had a GFMT score that was greater than the neurotypical mean score (but less than one standard deviation above the control mean). At group level, an independent *t*-test showed a significant difference in GFMT accuracy [$t(56) = 5.14$, $p < .001$, $d = 1.35$], with DPs less accurate ($M = 26.59$, $SD = 4.78$, range: 19–35) than Control participants ($M = 32.83$, $SD = 4.46$, range: 23–39).

As with the OFMT, Signal Detection Theory (Green & Swets, 1966) was used to characterise performance on the matching component of the GFMT. An independent *t*-test showed a significant difference in d' [$t(56) = 4.69$, $p < .001$, $d = 1.23$] with individuals with DP less sensitive to face matching signals ($M = .98$, $SD = .83$) than Control participants ($M = 2.10$, $SD = .98$). A Mann–Whitney test revealed a significant difference in criterion scores ($U = 254$, $p = .01$, $r = .34$; DP: median: −.14, range: −1.46 to 2.49; Control: median: 0, range: −.49 to 1.24). Thus, as on the OFMT, individuals were less sensitive to signals suggesting the faces were the same, and needed more evidence to make this response, on the GFMT.

3.1.4. CFMT

Thirteen participants with DP (45%) performed 2 SDs below the neurotypical mean score, with the remaining members of the DP group (55%) performing between one and two standard deviations below the Control mean. As expected (given it was used to select participants), a Mann–Whitney test revealed a significant difference in CFMT performance between groups ($U = 40.50$, $p = < .001$, $r = .78$), with DPs less accurate (median = 36, range: 23–43) than the Control group (median = 56, range: 33–70).

3.2. Face matching, controlling for face perception

Group (DP vs Control), Deviation scores, and their interaction were entered into two regressions, the first predicting OFMT matching accuracy and the second GFMT matching accuracy. For the OFMT analysis, Deviation scores were a significant predictor ($\beta = -.70$, $t = -6.58$, $p = <.001$) of OFMT matching accuracy, suggesting that face perception abilities are related to OFMT performance. Group was also a significant predictor of OFMT matching accuracy ($\beta = -.24$, $t = -2.46$, $p = .02$), indicating that face matching was worse in the DP group even after accounting for face perception ability. The interaction between Group and Deviation scores was not a significant predictor ($\beta = .09$, $t = 1$, $p = .32$), suggesting that the relationship between face perception and OFMT matching performance did not vary as a function of group.

The same pattern of significance was observed in the analysis predicting GFMT matching accuracy – both Deviation scores ($\beta = -.53$, $t = -3.95$, $p = <.001$) and Group ($\beta = -.28$, $t = -2.27$, $p = .03$) were significant predictors of GFMT matching accuracy, while the interaction between Group and Deviation was not ($\beta = .08$, $t = .72$, $p = .48$).

3.3. Face memory, controlling for face perception and face matching

Group (DP vs control), Deviation scores, Face Matching scores (OFMT, and separately GFMT matching accuracy) and the interactions between Deviation scores and Group, and Face Matching and Group, were entered into regressions predicting

CFMT scores. For the OFMT analysis, results showed significant independent contributions of face perception (as measured by Deviation scores; $\beta = -.28$, $t = -2.08$, $p = .04$) and face matching ($\beta = .32$, $t = 2.52$, $p = .02$). Group was a significant predictor ($\beta = -.39$, $t = -4.20$, $p = <.001$), indicating that face memory was worse in DP even after controlling for face perception and face matching. The two interactions were not significant predictors, indicating the relationships between face memory and face perception, and face memory and face matching, did not vary as a function of group (Group x Deviation scores: $\beta = .16$, $t = 1.38$, $p = .17$; Group × Face Matching: $\beta = -.10$, $t = -1.03$, $p = .31$).

Similarly, for the GFMT analysis, results showed significant independent contributions of face perception (as measured by Deviation scores; $\beta = -.36$, $t = -3.38$, $p = .001$) and face matching ($\beta = .26$, $t = 2.70$, $p = .009$). Group was also a significant predictor ($\beta = -.40$, $t = -4.46$, $p = <.001$). The interactions between Group and Deviation scores and Group and Face Matching scores were not significant predictors (Group × Deviation scores: $\beta = .16$, $t = 1.72$, $p = .09$; Group × Face Matching: $\beta = -.16$, $t = -2.00$, $p = .050$). Given that the Group × Face Matching interaction approached significance, exploratory analyses assessed the correlation between GFMT residuals (which represent GFMT face matching independent of face perception, see Analysis Strategy section) and CFMT scores in each group. These revealed a significant positive correlation in the Control group ($r_s = .46$, $p = .01$) and a weaker, non-significant correlation ($r_s = .02$, $p = .91$) in the DP group, suggesting that the contribution of face matching to face memory performance, independent of face perception, was stronger for the control compared to the DP group.

4. Discussion

This study examined the performance of individuals with DP on several face processing tasks, with the aim of deriving independent measures of face matching, face perception and face memory. Results showed that, as expected when standard test scores were compared, the group of individuals with DP performed worse on the three face processing tasks (OFMT, GFMT and CFMT) compared to the matched control group. It should be noted that previous studies have not always found that prosopagnosic individuals perform poorly on the GFMT (e.g. Fysh & Ramon, 2022; White, Rivolta, Burton, Al-Janabi, & Palermo, 2017), which may be due to issues with the sensitivity of the GFMT itself (White, Guilbert, Varela, Jenkins, & Burton, 2022), as it tends to be the case that studies with larger sample sizes are more likely to find an impairment in DP. More interestingly, analysis of OFMT facial similarity judgements revealed that individuals with DP were worse at judging the perceptual similarity of two faces compared to the matched control group. Furthermore, even after controlling for face perception, individuals with DP exhibited worse face matching performance (i.e., they were worse at deciding whether two face images were from the same person, or different people). This result was robust: it was observed when matching was assessed using both the OFMT and the GFMT. This novel finding indicates that individuals with DP have difficulties with both face perception *and* face matching. Interestingly, this pattern is different from that seen in autistic individuals (Stantić et al., under review), whereby autistic individuals exhibit difficulties in face perception but not in face matching. Although it is usually claimed that there are high rates of prosopagnosia in the autistic population (e.g., Cook, Shah, Gaule, Brewer, & Bird, 2015; Wilson, Palermo, Schmalzl, & Brock, 2010), these data indicate that there may be subtle differences in face processing between autistic individuals and non-autistic individuals with DP.

Comparison of the results of autistic and prosopagnosic individuals reinforces the distinction between the psychological processes of face perception and face matching. Furthermore, it suggests that although face perception is likely necessary for accurate face matching it is not sufficient. In addition to being able to form accurate perceptual representations of faces from memory or from a pictorial representation, one must have an accurate model of how, and how much, faces are allowed to vary before deciding they belong to different people. Results of the Signal Detection Theory analysis of the OFMT and GFMT matching task may be informative as to this point. In addition to a lower *d* prime, participants with DP exhibited a more extreme bias towards 'different' responses. That is, they needed more evidence that the faces were the same before they responded that they were the same. This is consistent with personal reports from prosopagnosic individuals who report that they fail to recognise individuals – i.e., they fail to recognise instances of faces (whether stored in memory or available for visual inspection) are of the same facial identity, and therefore faces are more likely to be judged as different. It is also consistent with claims that individuals with DP show impaired performance on matching tasks specifically for trials in which the two faces depict the same person, compared to trials in which different individuals are depicted (White et al., 2017).

Finally, individuals with DP exhibited impaired face memory even after accounting for their difficulties with face perception and face matching, indicating problems with all three facial identity processes tested. Despite the clear pattern of impairments seen in those with DP, it should be acknowledged that there were two main limitations of the current study which raise questions about the degree of generalisability that can be assumed from the current results. The first is the lack of any control stimulus class, meaning that it is not clear whether the perceptual, matching and memory impairments seen for faces would extend to other stimulus classes. This problem is a general one for the field, and reflects the fact that appropriate control stimuli are difficult to identify (e.g., Fry, Wilmer, Xie, Verfaellie, & DeGutis, 2020; Susilo et al., 2010). The second limitation is that prosopagnosic individuals were selected on the basis that they showed impaired performance on two out of three tests, one of which was the Cambridge Face Perception Task. As such, the sample of DP individuals identified may have been biased towards those that have perceptual difficulties in addition to problems with face memory. This possibility will only be able to be tested with further testing of the DP population with varying recruitment criteria, though it is worth noting that a fairly large degree of variation in test scores was observed within the group of individuals with DP. Unfortunately, the sample size was too low to allow formal testing for the presence of sub-groups (Dalmaijer, Nord, & Astle, 2022) but if future work

adopts the same testing and analysis procedure then samples could be combined to enable this approach.

Collectively, these results suggest that individuals with DP exhibit impaired face perception, face memory *and* face matching. Thus, we suggest that neither the 'memory' nor the 'perceptual' hypothesis is a sufficiently comprehensive account of face processing difficulties in DP, and that decision-making processes involved in matching perceptual stimuli with stored face representations (e.g., Bruce & Young, 1986), and how they may be impaired, are also of importance in understanding DP.

CRediT authorship contribution statement

M. Stantić: Study conceptualisation, Methodology, Investigation, Project administration, Software, Data curation, Writing – review & editing; **Z. Pounder:** Formal analysis, Writing – original draft, Writing – review & editing, Visualisation; **S. Bate:** Methodology, Resources; Writing – review & editing; **T. Susilo:** Methodology, Resources; Writing – review & editing; **C. Catmur:** Study conceptualisation, Methodology, Writing – review & editing; **G. Bird:** Study conceptualisation, Methodology, Funding acquisition, Supervision, Writing – review & editing.

Open practices

The study in this article earned an Open Data – Protected Access badge for transparent practices. Materials and data for the study are available at osf.io/vzsqd.

Data availability

All research data are available at https://osf.io/vzsqd

Acknowledgements

We would like to greatly thank all our participants who took the time to undertake this study. MS is funded by an ESRC DTP studentship and a Wilfrid Knapp Science Scholarship. TS was supported by the Royal Society of New Zealand Marsden Fund 16-VUW-175. This publication was made possible through the support of a grant from the John Templeton Foundation. The opinions expressed in this publication are those of the author(s) and do not necessarily reflect the views of the John Templeton Foundation.

REFERENCES

Behrmann, M., & Avidan, G. (2005). Congenital prosopagnosia: Face-blind from birth. Trends. *Cognitive Science, 9*(4), 180–187. https://doi.org/10.1016/j.tics.2005.02.011

Biotti, F., Gray, K. L. H., & Cook, R. (2019). Is developmental prosopagnosia best characterised as an apperceptive or mnemonic condition? *Neuropsychologia, 124*, 285–298. https://doi.org/10.1016/j.neuropsychologia.2018.11.014

Bobak, A. K., Parris, B. A., Gregory, N. J., Bennetts, R. J., & Bate, S. (2017). Eye-movement strategies in developmental prosopagnosia and "Super" face recognition. *The Quarterly Journal of Experimental Psychology: QJEP, 70*(2), 201–217. https://doi.org/10.1080/17470218.2016.1161059

Bruce, V., & Young, A. (1986). Understanding face recognition. *British Journal of Psychology, 77*(3), 305–327. https://doi.org/10.1111/j.2044-8295.1986.tb02199.x

Burton, A. M., White, D., & McNeill, A. (2010). The Glasgow face Matching test. *Behavior Research Methods, 42*(1), 286–291. https://doi.org/10.3758/BRM.42.1.286

Cook, R., & Biotti, F. (2016). Developmental prosopagnosia. *Current Biology: CB, 26*(8), R312–R313. https://doi.org/10.1016/j.cub.2016.01.008

Cook, R., Shah, P., Gaule, A., Brewer, R., & Bird, G. (2015). Autism and developmental prosopagnosia: A cross-disorder study. *Journal of Vision, 15*(12), 1211. https://doi.org/10.1167/15.12.1211

Corrow, S. L., Dalrymple, K. A., & Barton, J. J. (2016). Prosopagnosia: Current perspectives. *Eye Brain, 8*, 165–175. https://doi.org/10.2147/eb.S92838

Dalmaijer, E. S., Nord, C. L., & Astle, D. E. (2022). Statistical power for cluster analysis. *BMC Bioinformatics, 23*(1). https://doi.org/10.1186/s12859-022-04675-1

Dalrymple, K. A., Corrow, S., Yonas, A., & Duchaine, B. (2012). Developmental prosopagnosia in childhood. *Cognitive Neuropsychology, 29*(5–6), 393–418. https://doi.org/10.1080/02643294.2012.722547

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. https://doi.org/10.1016/j.dcn.2014.07.003

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. https://doi.org/10.1080/02643290701380491

Duchaine, B., & 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. https://doi.org/10.1016/j.neuropsychologia.2005.07.001

Fry, R., Wilmer, J., Xie, I., Verfaellie, M., & DeGutis, J. (2020). Evidence for normal novel object recognition abilities in developmental prosopagnosia. *Royal Soc Open Sci, 7*(9), Article 200988. https://doi.org/10.1098/rsos.200988

Fysh, M. C., & Bindemann, M. (2018). The Kent face Matching test. *British Journal of Psychology, 109*(2), 219–231. https://doi.org/10.1111/bjop.12260

Fysh, M. C., & Ramon, M. (2022). Accurate but inefficient: Standard face identity matching tests fail to identify prosopagnosia. *Neuropsychologia, 165*, Article 108119. https://doi.org/10.1016/j.neuropsychologia.2021.108119

Gray, K. L. H., Bird, G., & Cook, R. (2017). Robust associations between the 20-item prosopagnosia index and the Cambridge Face Memory Test in the general population. *Royaloyal Society Open Science, 4*(3), Article 160923. https://doi.org/10.1098/rsos.160923

Green, D. M., & Swets, J. A. (1966). *Signal detection theory and psychophysics*. Oxford, England: John Wiley.

Jackson, M. C., Counter, P., & Tree, J. J. (2017). Face working memory deficits in developmental prosopagnosia: Tests of encoding limits and updating processes. *Neuropsychologia, 106*, 60–70. https://doi.org/10.1016/j.neuropsychologia.2017.09.003

Kenny, L., Hattersley, C., Molins, B., Buckley, C., Povey, C., & Pellicano, E. (2016). Which terms should be used to describe autism? Perspectives from the UK autism community. *Autism: the International Journal of Research and Practice, 20*(4), 442–462. https://doi.org/10.1177/1362361315588200

Phillips, P. J., & O'Toole, A. J. (2014). Comparison of human and computer performance across face recognition experiments. *Image and Vision Computing, 32*(1), 74–85. https://doi.org/10.1016/j.imavis.2013.12.002

Phillips, P. J., Yates, A. N., Hu, Y., Hahn, C. A., Noyes, E., Jackson, K., et al. (2018). Face recognition accuracy of forensic examiners, superrecognizers, and face recognition algorithms. *Proceedings of the National Academy of Sciences, 115*(24), 6171–6176. https://doi.org/10.1073/pnas.1721355115

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. *Royaloyal Society Open Science, 2*(6), Article 140343. https://doi.org/10.1098/rsos.140343

Stantic, M., Brewer, R., Duchaine, B., Banissy, M. J., Bate, S., Susilo, T., et al. (2021). The Oxford face Matching test: A non-biased test of the full range of individual differences in face perception. *Behavior Research Methods, 54*(1), 158–173.

Stantić, M., Brown, K., Ichijo, E., Catmur, C., Pounder, Z., & Bird, G., & Stantić. (Under review). Independent measurement of face perception, face matching, and face memory reveals impairments in face perception and memory, but not matching, in autism. Submitted for publication.

Stollhoff, R., Jost, J., Elze, T., & Kennerknecht, I. (2011). Deficits in long-term recognition memory reveal dissociated subtypes in congenital prosopagnosia. *Plos One, 6*(1), Article e15702. https://doi.org/10.1371/journal.pone.0015702

Susilo, T., & Duchaine, B. (2013). Advances in developmental prosopagnosia research. *Current Opinion in Neurobiology, 23*(3), 423–429. https://doi.org/10.1016/j.conb.2012.12.011

Susilo, T., McKone, E., Dennett, H., Darke, H., Palermo, R., Hall, A., et al. (2010). Face recognition impairments despite normal holistic processing and face space coding: Evidence from a case of developmental prosopagnosia. *Cognitive Neuropsychology, 27*(8), 636–664. https://doi.org/10.1080/02643294.2011.613372

Towler, J., Fisher, K., & Eimer, M. (2018). Holistic face perception is impaired in developmental prosopagnosia. *Cortex; a Journal Devoted To the Study of the Nervous System and Behavior, 108*, 112–126. https://doi.org/10.1016/j.cortex.2018.07.019

White, D., Guilbert, D., Varela, V. P. L., Jenkins, R., & Burton, A. M. (2022). GFMT2: A psychometric measure of face matching ability. *Behavior Research Methods, 54*(1), 252–260. https://doi.org/10.3758/s13428-021-01638-x

White, D., Rivolta, D., Burton, A. M., Al-Janabi, S., & Palermo, R. (2017). Face matching impairment in developmental prosopagnosia. *The Quarterly Journal of Experimental Psychology: QJEP, 70*(2), 287–297. https://doi.org/10.1080/17470218.2016.1173076

Wilson, C. E., Palermo, R., Schmalzl, L., & Brock, J. (2010). Specificity of impaired facial identity recognition in children with suspected developmental prosopagnosia. *Cognitive Neuropsychology, 27*(1), 30–45. https://doi.org/10.1080/02643294.2010.490207