

## Distinct abilities associated with matching same identity faces vs. discriminating different faces: Evidence from individual differences in prosopagnosics and controls

Journal:	<i>Quarterly Journal of Experimental Psychology</i>
Manuscript ID	QJE-STD-21-167.R1
Manuscript Type:	Standard Article
Date Submitted by the Author:	10-Nov-2021
Complete List of Authors:	Berger, Amy; Tufts University Fry, Regan; Harvard Medical School Bobak, Anna; University of Stirling Juliano, Angela; Brigham and Women's Hospital DeGutis, Joseph; Harvard Medical School,
Keywords:	face matching, developmental prosopagnosia, holistic processing, feature processing, eye processing, individual differences

SCHOLARONE™  
Manuscripts

Berger A, Fry R, Bobak A, Juliano A & DeGutis J , EXPRESS: Distinct abilities associated with matching same identity faces vs. discriminating different faces: Evidence from individual differences in prosopagnosics and controls, *Quarterly Journal of Experimental Psychology* (Forthcoming). Copyright © The Authors 2022. Reprinted by permission of SAGE Publications. Reuse is restricted to non-commercial and no derivative uses. DOI: <https://doi.org/10.1177/17470218221076817>

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

Distinct abilities associated with matching same identity faces vs. discriminating  
different faces: Evidence from individual differences in prosopagnosics and controls

Amy Berger<sup>1,3</sup>, Regan Fry<sup>2,3</sup>, Anna K. Bobak<sup>4</sup>, Angela Juliano<sup>3,5</sup>, Joseph DeGutis<sup>2,3</sup>

- 1. Department of Psychology, Tufts University, Medford, MA, USA
- 2. Department of Psychiatry, Harvard Medical School, Boston, MA, USA
- 3. Boston Attention and Learning Laboratory, VA Boston Healthcare System, Boston, MA, USA
- 4. Psychology, University of Stirling, Stirling, UK
- 5. Brigham and Women’s Hospital, Boston, MA

Correspondence concerning this article be addressed to:

Joseph DeGutis  
VA Boston Healthcare System (182JP)  
150 S Huntington Ave., Boston, MA, USA  
E-mail: [degutis@hms.harvard.edu](mailto:degutis@hms.harvard.edu)

Declarations of interest: none

### Abstract

Previous face matching studies provide evidence that matching same identity faces (match trials) and discriminating different face identities (non-match trials) rely on distinct processes. For example, instructional studies geared towards improving face matching in applied settings have often found selective improvements in match or non-match trials only. Additionally, a small study found that developmental prosopagnosics (DPs) have specific deficits in making match but not non-match judgments. In the current study, we sought to replicate this finding in DPs and examine how individual differences across DPs and controls in match vs. non-match performance relate to featural vs. holistic processing abilities. 43 DPs and 27 controls matched face images shown from similar front views or with varied lighting or viewpoint. Participants also performed tasks measuring featural (eyes/mouth) and holistic processing (part-whole task). We found that DPs showed worse overall matching performance than controls and that their relative match vs. non-match deficit depended on image variation condition, indicating that DPs do not consistently show match- or non-match-specific deficits. When examining the association between holistic and featural processing abilities and match vs. non-match trials in the entire group of DPs and controls, we found a very clear dissociation: Match trials significantly correlated with eye processing ability ( $r=.48$ ) but not holistic processing ( $r=.11$ ), whereas non-match trials significantly correlated with holistic processing ( $r=.32$ ) but not eye processing ability ( $r=.03$ ). This suggests that matching same identity faces relies more on eye processing while discriminating different faces relies more on holistic processing.

**Keywords:** *face matching, developmental prosopagnosia, holistic processing, feature processing, eye processing, individual differences*

1

2

3

4

5

6

7

8

9

10

11

12

13

14

15

16

17

18

19

20

21

22

23

24

25

26

27

28

29

30

31

32

33

34

35

36

37

38

39

40

41

42

43

44

45

46

47

48

49

50

51

52

53

54

55

56

57

58

59

60

**1. Introduction**

Comparing two face images is an important task for identity verification routinely performed by passport control officers, police officers, bartenders, and cashiers. Accurately matching faces can be challenging and error-prone, particularly for unfamiliar faces (for a review, see Young & Burton, 2018). Recent studies suggest that matching same identity faces (match trials) and discriminating different face identities (non-match trials) rely on distinct processes (Bindemann & Burton, 2020), though the nature of these processes is poorly understood. Further evidence for this dissociation is from a recent small study that found that developmental prosopagnosics (DPs), individuals with lifelong face recognition deficits (Duchaine & Nakayama, 2006), were impaired on match but not non-match trials (White, 2017). The goals of the current study were to a) further investigate the nature of face matching impairments in a larger sample of DPs under multiple face image variation conditions and b) examine whether individual differences in the ability to match same face identities vs. discriminate different faces rely more or less on feature-based vs. holistic processing.

Studies have consistently found evidence that performance on face match and non-match trials depends on at least partially dissociable mechanisms. First, in contrast to object tasks that typically show that performance on old/match items and new/non-match items is highly correlated (mirror effect, Glazer and Adams, 1985), unfamiliar faces typically show much lower correlations between match and non-match trial accuracy of the same face images (e.g.,  $r=.10$ , Megreya & Burton, 2007). Additionally, face match and non-match trials have been shown to dissociate during instructional intervention studies geared towards enhancing face matching performance in applied settings. For example, Towler, White & Kemp (2017) presented a pair of unfamiliar faces above a list of eleven facial features (e.g., ears, eyes, nose,

etc.), and observers were asked to first rate the similarity of each feature and then make match/non-match judgments. The results demonstrated that, compared to making no ratings, feature similarity ratings improved accuracy on match trials and slightly impaired accuracy on non-match trials. Using a similar paradigm, Megreya and Bindemann (2018) found that, compared to not providing instructions, specifically instructing participants to pay attention to the eyebrows led to improvements in accuracy on match trials but not non-match trials whereas instructions to attend to the ears decreased both match and non-match performance. Contrasting these results, Towler et al. (2021) recently found that instructing individuals to focus specifically on the ears and facial marks (e.g., scars, blemishes) improved accuracy on non-match trials, but not match trials, though these effects were specific to the high-resolution faces condition. Together, these studies provide evidence that match and non-match trials depend on separable mechanisms. However, the results observed have been inconsistent and may be specific to the images used and the mechanisms underlying differential match vs. non-match performance remain to be characterized.

Studying how individual differences in feature and holistic processing abilities associate with match vs. non-match performance could provide important complementary evidence to these studies. Since increased attention to features has often been shown to enhance match trial performance (e.g., Megreya and Bindemann, 2018; though see Towler et al., 2021 for an exception), feature abilities could be particularly associated with match trial performance. It should be noted that the features focused on by facial examiners may be distinct from features commonly attended to in everyday life and feature expertise in the normal population (e.g., Phillips et al., 2018). Using a normal control sample, Abudarham and colleagues (Abudarham et al., 2019; Abudarham et al., 2021) used a reverse-engineering approach and identified specific

features critical for determining face identity in both familiar and unfamiliar faces: lip thickness, eye color, eye shape, eyebrow thickness, and hair. Consistent with these critical features, using the “bubbles” technique, super face recognizers, who have superior face matching abilities (Bobak et al., 2016), have shown to use the eyes/eyebrows and the mouth more for recognition than controls (Tardiff et al., 2018; though they may tend to fixate the bridge of the nose, Bobak et al., 2017). Further, Davis et al. (2021) found that instructing participants to attend to internal features (eyebrows, nose, mouth) resulted in higher match hit rates on a simultaneous matching task, while instructions to attend to external features (ears, chin/jawline, face shape) and no instructions both produced similarly lower hit rates. This suggests that enhanced perceptual sensitivity to features in the eye and mouth regions is important for matching faces and could be particularly associated with improved match trial performance. In addition to specific feature abilities, the ability to perform holistic face processing, the simultaneous integration of features and spacing information into a unified percept (Rossion, 2008), could also be differentially related to match vs. non-match performance. Though some have questioned the degree to which face expertise mechanisms like holistic processing are used when matching unfamiliar faces (e.g., Young and Burton, 2018), better holistic processing abilities, as measured by the part-whole and face inversion effects, have been robustly associated with more accurate matching of unfamiliar faces (Cambridge Face Perception Test, Rezlescu, 2017; though not when using the composite effect to measure holistic processing, see Verhallen, Bosten, et al., 2017). However, no study to date has reported specific associations between holistic processing and match vs. non-match trials.

Examining face matching in individuals with inferior face matching abilities may provide additional insights into the mechanisms underlying match vs. non-match judgments.

Developmental prosopagnosics (DPs) have been shown to be impaired at face matching tasks for both unfamiliar (e.g., Mishra et al., 2021; White et al., 2017; though see Ulrich et al., 2017) and familiar faces (White et al., 2017). DPs have also demonstrated deficits in feature-based processing (DeGutis et al., 2012; Le Grand et al., 2006), especially of internal facial features (e.g., eye region, Fisher, Towler, & Eimer, 2016). They also have been shown to have reduced, though not completely abolished, holistic face processing (DeGutis, Cohan, et al., 2012; DeGutis, Chatterjee, et al., 2012; Avidan, Tanzer, & Behrmann, 2011; Palermo et al., 2011; though see Tsantani et al., 2020). Notably, despite these feature and holistic deficits and anecdotal accounts of DPs relying on extra-facial information such as voice and gait for recognition (Adams et al., 2019), DPs have shown to rely on very similar critical features as controls when performing face matching (Abudarham et al., 2021). In terms of face match vs. non-match trial performance, White et al. (2017) found that a small group of 6 DPs were impaired on match trials but performed similar to controls on non-match trials. This pattern was found when using images of both familiar and unfamiliar faces. This suggests that DPs have a specific deficit in matching same-identity faces. Though the idea that DPs have a match trial-specific deficit but intact non-match performance is provocative, this effect could be driven by simply a non-match response bias in a few DPs. Further, anecdotal reports of DPs suggest that they often confuse two different individuals as being the same person (e.g., Murray et al., 2018), consistent with *non-match* errors. However, these reports could reflect more of a memory phenomenon than a perceptual one. Considering that DP is a heterogeneous disorder (see Corrow, Dalrymple, & Barton, 2016 for a discussion), replication of White et al.'s match trial-specific deficits with a larger DP sample is necessary before strong conclusions can be made.

In the current study, we sought to further clarify the mechanisms of match vs. non-match judgments by a) examining a larger sample of DPs (N=43) and b) performing individual differences analysis with match/non-match performance and measures of feature and holistic processing. Because face matching results may depend on the to-be-matched faces and image variations (Bindemann & Burton, 2020), we examined whether the results were similar across different image changes by showing pairs of faces from either a) front views with similar lighting, b) front views with different lighting, or c) different viewpoints. We had participants view image pairs that disappeared after three seconds, which is approximately the average time to make a match judgment with faces having similar lighting/viewpoint changes (e.g., Rossion & Michel, 2018). This also avoided ceiling effects in accuracy. Finally, to better understand the mechanisms underlying match vs. non-match trials, we administered an assessment of feature discrimination ability (‘Georges’ task, Malcolm et al., 2004) as well as the part-whole task to measure both feature and holistic processing (Tanaka et al., 2004). We then examined individual differences associations between feature and holistic processing and match/non-match conditions in the combined DP and control group, to maximize the range of individual differences across measures (we also analyzed each group separately). Importantly, previous studies have found that, rather than representing two discrete groups, the distribution of DPs’ perceptual performance is shifted towards impairment but highly overlapping with the control distribution (e.g., Biotti and Cook, 2019) and DPs may use similar critical features for face matching (Abudarham et al, 2021), suggesting that examining the combined DP and control group is a valid approach to examining individual differences associations. Based on previous work (Towler et al., 2017; Megreya and Bindemann, 2018), we hypothesized that better feature discrimination abilities, particularly with the eye region that contains many identity-diagnostic

features (Abudarham, Shkiller, & Yovel, 2019) would be associated with better matching trial performance.

## 2. Materials and Methods

### 2.1 Participants

Participants were between the ages of 18 and 70 years old (N=70). Developmental prosopagnosics were recruited from our database of previous DP participants in the Boston area, references from other research labs (Dr. Matthew Peterson, MIT; Professor Brad Duchaine, Dartmouth College, [www.faceblind.org](http://www.faceblind.org)), and individuals who responded to our advertisement on the Massachusetts Bay Transportation Authority subway system. Control subjects were recruited from both the Harvard Decision Science Laboratory in Cambridge, Massachusetts and through flyers distributed in the Boston area.

Developmental prosopagnosics were screened using the 20-Item Prosopagnosia Index (PI-20; Shah, Gaule, Sowden, Bird, & Cook, 2015), a famous faces memory test (FFMT), and the CFMT (Duchaine & Nakayama, 2006b). To qualify as a DP, participants had to report lifelong face recognition deficits (all but three scored  $> 65$  on the PI-20, see Table 1), present with objective face recognition deficits on both the CFMT and FFMT ( $z < -1.5$ , as calculated from the control group in Duchaine et al., 2006b), and have an absence of significant neurological disorders (similar to our recent studies, e.g., Stumps et al., 2020). Typically developing controls did not report any face recognition deficits and all scored 45 or above on the CFMT. All participants had normal or corrected-to-normal vision and had to have scored within the normal range on the Leuven Perceptual Organization Screening Test (L-POST; Torfs, Vancleef, Lafosse, Wagemans, & de-Wit, 2014) to rule out other causes of poor face recognition. It should be noted that since the PI-20 measures the severity of face recognition difficulties and

controls had to report no lifelong difficulties in face recognition, controls did not receive the PI-  
20. Informed consent was obtained for all participants prior to data collection according to the  
Declaration of Helsinki. Participants were compensated for their time at a rate of \$10 per hour.  
The study was approved by the VA Boston Healthcare System and Harvard Medical School  
Institutional Review Boards, and all study tasks were completed at either the VA Boston  
Healthcare System in Jamaica Plain or the Harvard Decision Science Lab.

**2.2 Same/Different Face Matching Task**

In this task, participants were presented with two face images on the screen side-by-side  
(4.5 x 6 cm each) for 3 seconds and had to indicate by pressing 1 or 0 whether the faces were the  
same or different identities, respectively. There was a one second inter-trial interval. We used  
neutral expression face images from the multi-PIE database (Gross et al., 2010) that we  
converted to grayscale and cropped to remove external features such as hair or clothing.  
Individual foil faces were carefully selected to be matched to each individual target face based  
on gender, age, ethnicity, and distinctive features (e.g., thin eyebrows, dark eyes). In particular,  
foil faces had very similar verbal descriptions to target faces in terms of their top three most  
distinctive features. There were seven different trial types: 1) same identity from front view (face  
images were taken on different days), 2) different identity from front view, 3) same identity with  
lighting change (fully lit vs. lit from the side), 4) different identity with lighting change, 5) same  
identity with viewpoint change (front view vs. 3/4 view), 6) different identity with viewpoint  
change, 7) identical images with two different crops. There were 30 trials per trial type which  
were randomly intermixed for a total of 210 trials. The same identity and same day but cropped  
differently trials were included as an effort check. Participants were ~99% accurate on these  
trials and they were not included in the subsequent analyses.

### 2.3 Part-whole Task

We used the version of the part-whole task from Tanaka et al (2004), used with permission of Jim Tanaka, University of Victoria. Target faces were created using the outline of one Caucasian male face. By inserting a combination of six different pairs of eyes, noses, and mouths, six unique target faces were created. For whole trials, foil faces were created by switching one of the three facial features (eyes, nose, or mouth) with that of a different target face. For part trials, foil stimuli were an isolated facial feature (eyes, nose, or mouth) from another target face. Each trial began with a central fixation display presented for 500 ms. Next, one of the six target faces was centrally presented for 1,000 ms, and subjects attempted to encode this face. Next, a scrambled face mask was displayed for 500 ms. During the subsequent test period, participants were presented with a pair of probe images side by side, either whole faces (whole trials) or isolated features (part trials). One of these images matched the target, and the other image was a foil. Stimuli remained on the screen until participants indicated with a button press which probe stimulus matched the target face (subjects responded 1 for left image, 2 for right image). For whole trials, subjects chose between the whole target face and a whole foil face, which was the same as the target face except that one of the features (eyes, nose, or mouth) was replaced with a foil feature. For part trials, subjects chose between a face part from the target face (eyes, nose, or mouth) and the same facial feature from a foil face. On a given trial, subjects were given no indication on which feature they would be tested, nor did they know whether isolated features or whole faces would be shown during the test period. There were 72 trials (36 parts trials and 36 whole trials), 24 for each feature type. We calculated the holistic advantage by first regressing the part trial ‘control condition’ from the whole trial ‘condition of interest’ (using

the regression equation in the control sample, e.g., see DeGutis et al., 2013), and then applying this equation to calculate residuals for DPs and controls.

**2.4 Georges Task**

The Georges task has been used in previous studies as a measure of the ability to discriminate feature identities and configurations (e.g., Malcolm et al., 2004; Pancaroglu et al., 2016). In the Georges face task, participants are presented with frontal views of three same-identity unfamiliar faces in a triangular arrangement for two seconds, with the lower two faces slightly offset horizontally. Two faces were identical and one had a single manipulation. These faces were manipulated in six ways. These six ways involved three categories of change: internal feature position, feature size, and external contour. Each category of change had one manipulation in the upper face and one in the lower face. To manipulate feature position, either the eyes were edited to be closer together or the mouth was edited to be higher on the face. To manipulate feature shape, the width of both eyes was increased or the vertical width of the mouth was increased. To manipulate external contour, the hairline was elevated or the chin was narrowed. Participants indicated which of the three faces differed from the other two using the left, right, and up arrow keys. There were 108 trials, with six different face identities.

**2.5 Analyses**

**2.5.1 Sample size justification.** Our sample size was guided by previous studies comparing DPs and controls (White et al., 2017) as well as individual differences studies (Richler et al., 2011; DeGutis et al., 2013; Rezlescu et al., 2017). White et al. (2017) found significant differences in face matching ability between controls and DPs with a sample of only 21 controls and 6 DPs. However, considering that DPs have been shown to be quite perceptually heterogeneous (Corrow et al., 2016), we wanted to include additional DPs in the current study to

ensure that our findings would be robust and replicable. Unfortunately, our sample of controls was smaller than the DP sample due to not being allowed to run in-lab participants during the COVID-19 pandemic. To test for individual differences associations, we used a combined sample similar in size to studies that have found significant individual differences between feature and holistic processing and face recognition (N=38, Richler et al., 2011; N=43, DeGutis et al., 2013) as well as face matching ability (N=80 in-lab, Rezlescu et al., 2017). Including both DPs and controls in our sample provided a more expanded range of face matching performance than simply examining an unimpaired population, which likely improved our ability to detect individual differences associations.

**2.5.2 Analysis Plan.** To confirm that DPs showed overall deficits in face matching, we first compared DPs and controls in their overall face matching ability collapsed across match and non-match trials using accuracy and d-prime. To examine if there were response bias differences (e.g., DPs responding “non-match” on the majority of trials), we also compared criterion c between the groups. Next, to test if DPs showed a selective deficit on match trials, we compared DPs and controls in their overall match and non-match trial accuracy. We then proceeded to test whether match/non-match differences between DPs and controls varied by condition by performing a repeated measures ANOVA on group x match/non-match x condition (lighting change, viewpoint change, and front view conditions).

To examine how individual differences in match vs. non-match trial performance relate to holistic processing as well as eyes/mouth processing, we calculated holistic processing from the part-whole task and focused on feature composites. We created composite measures for two reasons. First, the individual feature conditions were made up of a limited number of trials and had modest reliability (see supplementary materials) and composite measures were employed to

improve reliability, which is particularly important for individual differences analyses (e.g., DeGutis et al., 2013). Second, composites provide a summary measure of participants’ overall sensitivity to changes to a particular feature. Composites were created for eyes by averaging all part-whole eyes trials (whole and part) with Georges task eyes trials (eye size/width) and for mouths by averaging all part-whole mouth trials (whole and part) and Georges task mouth (mouth size/vertical position). Because knowing the reliabilities of measures is particularly important for individual differences research (i.e., knowing the reliability gives an indication of the maximum correlation between two measures), we calculated reliabilities of all measures using Cronbach’s alpha and Guttman’s Lambda 2 (similar to DeGutis et al., 2013; Rezlescu et al., 2017). For eyes and mouth composite reliability, we used the Wang and Stanley (1970) composite reliability formula. We then ran correlations between holistic processing and eye/mouth processing and match and non-match trials. We also ran exploratory correlations between the other Georges (forehead, chin) and part-whole features. Because each of these features were only represented by a single condition (e.g., only changing the size or the position) and were therefore less reliable, we did not focus our main analyses on these features. We tested for differences between correlations for match and non-match trials by using Fisher’s r-to-z transformation and computing z-tests for dependent correlations. Finally, we ran exploratory multiple regressions using eye processing and holistic processing to predict overall face matching performance.

3. Results

3.1 Demographics and Diagnostic Test Performance

Our sample consisted of 43 DPs (33 females) and 27 controls (19 females) with a similar mean age of 38.07 (SD=13.91) and 40.03 (SD=11.69), respectively (see Table 1). According to

the DSM-5 criteria of cognitive impairment ( $< -1$  SD for mild,  $< -2$  SD for major), based on diagnostic face recognition measures our DP sample included 6 mild DPs and 37 major DPs. We conducted DP vs. control group analyses excluding these 6 mild DPs and the results were very similar. Unsurprisingly, the entire DP group performed substantially worse than controls on the CFMT (DP  $M=39.9$  vs. control  $M=59.8$ ) and to a lesser extent on the same/different face matching task (see Table 1). Note that the control CFMT Mean and SD was similar to the original Duchaine et al. study (2006b) and there were only three controls with scores of 70 or above, suggesting that controls with very high face recognition abilities were not driving the results.

--- Insert Table 1 about here ---

### 3.2 Comparing Face Matching Performance between DPs and Controls

#### 3.2.1 Overall face matching performance collapsed across match/non-match trials.

Before separately analyzing match and non-match trials, we first sought to confirm that DPs were impaired in their overall performance on the face matching task and determine if this varied by condition. Starting with accuracy, we ran a 2 (DP/control)  $\times$  3 (lighting change/viewpoint change/front view) ANOVA. We found a significant main effect of group ( $F(1,68)=9.40$ ,  $p=.003$ ,  $\eta^2=.12$ ), with DPs performing worse than controls, and condition ( $F(2,136)=46.21$ ,  $p<.001$ ,  $\eta^2=.41$ ), with participants performing better on front view than viewpoint and lighting change trials. However, we did not find a significant group  $\times$  condition interaction ( $F(2,136)=1.08$ ,  $p=.342$ ). A very similar pattern was found when examining d-prime, with significant main effects of group ( $F(1,68)=17.16$ ,  $p<.001$ ,  $\eta^2=.20$ ) and condition ( $F(2,136)=39.55$ ,  $p<.001$ ,  $\eta^2=.37$ ), but no significant interaction ( $F(2,136)=0.99$ ,  $p=.374$ ). Finally, we examined criterion c to

determine if there was a significant difference in response bias between DPs and controls across conditions. We found no significant main effect of group ( $F(1,68)=1.34, p=.252$ ), but we did find a significant main effect of condition ( $F(2,136)=132.92, p<.001, \eta^2=.66$ ), with lighting change trials showing a slight “non-match” response bias ( $M=-.08, SD=.53$ ) while viewpoint change ( $M=.73, SD=.45$ ) and front-view conditions ( $M=.28, SD=.40$ ) having clearer “match” response biases. We did not find a group x condition interaction ( $F(2,136)=1.99, p=.141$ ).

**3.2.2 Match vs. non-match trial accuracy.** We next sought to examine one of the main goals of this study, to test whether, similar to White et al. (2017), DPs’ face matching deficits were greater for match than non-match trials. Since White et al. (2017) included various lighting and viewpoint changes in their face stimuli, we performed a group x match/non-match trial ANOVA collapsed across conditions. We found significant main effects of group ( $F(1,68)=9.95, p=.002, \eta^2=.13$ ), with DPs performing worse than controls, and match/non-match ( $F(1,68)=34.18, p<.001, \eta^2=.34$ ), with worse performance on non-match trials. However, critically, we did not find a significant group x match/non-match interaction ( $F(1,68)=0.52, p=.472$ ).

--- Insert Figure 1 about here ---

One possible reason we did not replicate White et al. (2017) is that the match/non-match trial difference between DPs and controls varied as a function of the nature of the image changes in our lighting change, viewpoint change, and front view conditions. To examine this possibility, we conducted a 3-way ANOVA with group (DP/control) x trial type (match/non-match trials) x condition (lighting, viewpoint, front view) and indeed found a significant 3-way interaction between group, trial type, and condition ( $F(2,68)=4.28, p=.016, \eta^2=.06$ ). As can be seen in

Figure 1, DPs had numerically reduced performance on all conditions, but only showed significantly worse performance on the lighting change match trials. When performing DP/control x match/non-match ANOVAs for each separate condition, we found that only lighting change trended towards a significant interaction ( $F(1,68)=2.85$ ,  $p=.096$ ), with DPs having greater performance decrements for match than non-match trials. Viewpoint change and front view conditions failed to demonstrate any group x match/non-match interaction (viewpoint:  $F(1,68)=0.21$ ,  $p=.646$ ; front view:  $F(1,68)=0.17$ ,  $p=.683$ ). Together, our results demonstrate that across conditions DPs did not consistently have greater deficits on match trials compared to non-match trials, but rather that DPs' relative match vs. non-match trial deficits depended on image variation conditions.

**3.2.3 Match vs. non-match trial reaction time.** We next analyzed reaction times to see if there was a match/non-match x group interaction (see Table 2). Two control participants were missing reaction times due to a computer error. Similar to previous studies (e.g., Ozbek and Bindemann, 2011), only correct trials were included in RT calculations. We conducted a 3-way ANOVA with group (DP/control) x trial type (match/non-match trials) x condition (lighting, viewpoint, front view). We found a main effect of group, with DPs slower than controls ( $F(1,66)=14.090$ ,  $p<.001$ ,  $\eta^2=.22$ ), a main effect of match/non-match ( $F(1,66)=38.45$ ,  $p<.001$ ,  $\eta^2=.33$ ), with non-match slower than match trials, and a main effect of condition ( $F(2,66)=19.66$ ,  $p<.001$ ,  $\eta^2=.34$ ), with lighting change trials slower than front view and view change trials. We did not observe any significant interactions with group (all  $p's>.17$ ).

--- Insert Table 2 about here ---

**3.3 Individual differences in feature/holistic processing associated with match/non-match accuracy**

**3.3.1 Reliability of measures.** The second goal of the study was to measure associations between individual differences in feature and holistic processing abilities and match/non-match trials. Before quantifying these associations, we first wanted to determine whether our measures had adequate reliability. As can be seen in Supplementary Table 1, with regards to the face matching task, we found that the reliability was acceptable (.75 or greater) for overall accuracy (Cronbach’s alpha=.94/Guttman’s lambda 2=.94), match trials (Cronbach’s alpha=.89/Guttman’s lambda 2=.90), and non-match trials (Cronbach’s alpha=.92/Guttman’s lambda 2=.93). The eye composite also had adequate reliability (Cronbach’s alpha=.71/Guttman’s lambda 2=.75), whereas the mouth composite showed modest reliability (Cronbach’s alpha=.61/Guttman’s lambda 2=.65). Finally, part-whole holistic processing residuals showed modest reliability (Cronbach’s alpha=.61/Guttman’s lambda 2=.65), consistent with previous part-whole task studies (e.g., DeGutis et al., 2013; Rezlescu et al., 2017). However, this is quite good for a derived measure (i.e., residuals or subtraction scores), which typically have reduced reliability compared to measures of a single process.

Using these measures, we calculated the theoretical upper bound of the match/non-match correlations (similar to Verhallen et al., 2017). Computed as the geometric mean of the reliabilities of the measures being correlated (Schmidt & Hunter, 1996), the upper bound is the correlation that would be expected between these measures, once measurement error is taken into account, if their true correlation was 1.0. Using these reliabilities, we calculated the upper bound correlations for match trials with eye processing composite (Cronbach’s alpha: .79, Guttman’s lambda 2: .82) and holistic processing (Cronbach’s alpha: .74, Guttman’s lambda 2: .76). We

also calculated upper bound correlations for non-match trials with eye processing composite (Cronbach's alpha: .81, Guttman's lambda 2: .84) and holistic processing (Cronbach's alpha: .75, Guttman's lambda 2: .78).

**3.3.2 Performance on Part-whole and Georges tasks.** Before examining individual differences associations, we sought to validate that the part-whole and Georges tasks performed similarly to previous studies (e.g., DeGutis et al., 2013; Rezlescu et al., 2017; Barton et al., 2008). For the part-whole task, we focused our analysis on the whole trials after residualizing out the part trial accuracy as our measure of holistic advantage, as this better reflects the proposed mechanism behind the part-whole task and has been shown to be more reliable than subtraction scores (DeGutis et al., 2013). We found that across all participants, there was significantly greater accuracy on the whole trials compared to part trials ( $t(69)=6.37, p<.001$ ). When comparing the holistic advantage residual scores between groups, we found that controls had a significantly greater holistic advantage than DPs ( $t(68)=3.13, p=.003$ ).

When analyzing feature change accuracy on the Georges task, we replicated previous work showing that prosopagnosics perform worse than controls at discriminating internal features (Barton et al., 2008). We performed a 2 (group) x 6 (condition) repeated-measures ANOVA and found a main effect of feature ( $F(5,335)=10.38, p<0.001$ ), with more accurate performance across groups on the forehead size and poorer performance on the mouth size (see supplementary Figure 2). We also found an interaction between feature accuracy and group ( $F(5,335)=5.10, p<.001$ ), where DPs performed worse than controls at the eye size, eye spacing, and mouth spacing trials, while performing *better* than controls on chin trials (see supplementary Figure 2).

In assessing internal facial feature processing abilities, to create the most reliable

measures possible we focused on the average of the eyes and mouth conditions from the Georges task (combining both size and spacing trials) and eyes and mouth trials from the part-whole task. We did not assess nose trials from the part-whole task due to floor effects. When comparing DPs and controls on the eyes and mouth composite average, we found DPs performed worse than controls on the eyes (DP  $M=.65$ ,  $SD=.11$ ; Control  $M=.75$ ,  $SD=.12$ ) and similarly on the mouth (DP  $M=.63$ ,  $SD=.08$ ; Control  $M=.65$ ,  $SD=.09$ ). Performing a 2 x 2 ANOVA, we found a main effect of group, with DPs performing significantly worse than controls ( $F(1,68)=9.68$ ,  $p=.003$ ), a main effect of feature, with eye accuracy exceeding mouth accuracy ( $F(1,68)=14.42$ ,  $p<.001$ ), and a significant group x feature interaction, with DPs showing greater impairments on eye trials relative to the mouth trials ( $F(1,68)=5.12$ ,  $p=.027$ ).

**3.3.3 Predicting match and non-match performance from feature and holistic processing.** We began by correlating all the separate feature conditions with match/non-match trials separately for DPs, controls, as well as in the collapsed group (see Figure 2). Notably, with regards to the eye region, forehead, and holistic processing, DPs and controls demonstrated a very similar pattern of correlations with match/non-match trials, with DPs showing slightly *stronger* associations. This similarity between DPs and controls is consistent with recent findings showing that DPs and controls use very similar features to match facial identity (Abudarham et al., 2021). However, control participants' mouth and chin trials correlated with match trials whereas DPs' mouth and chin trial correlations were less consistent and were larger with non-match trials. Because of the inconsistency between DPs and controls in mouth associations, we chose to focus our individual differences analyses of the combined DP/control group on the eye composite and holistic processing. Further, we found that the eye feature conditions were

strongly correlated within DP, control, and combined DP/control groups whereas the mouth feature conditions were less correlated with each other (see supplementary Tables 5-10).

--- Insert Figure 2 about here ---

We next separately examined match/non-match correlations in the combined DP and control group. As can be seen in Figure 3, for match trials we found a significant positive correlation between eye accuracy and match trial accuracy ( $r(69) = .482, p < .001$ ) but did not find a significant correlation between holistic processing accuracy and match trials ( $r(69) = .108, p = .373$ ). Notably, these correlation coefficients were significantly different from each other ( $z = 3.456, p < .001$ ). For non-match trials, we found a complementary pattern, observing a significant positive correlation between holistic processing accuracy and non-match trial accuracy ( $r(69) = .320, p = .007$ ), but no significant correlation between eye accuracy and non-match trials ( $r(69) = .027, p = .822$ ). Again, these correlations were significantly different from each other ( $z = -2.564, p = .010$ ). The observed correlation between eye processing/match trials and holistic processing/non-match trials are impressive when compared to their upper bound correlations (the maximum correlation one could observe considering measurement reliability) of .82 and .78, respectively. This suggests that, after correcting for measurement error, the correlation between eye processing and match trials is  $r = .588$  and correlation between holistic processing and non-match trials is  $r = .410$ .

We next sought to examine the consistency of these relationships across conditions. We separately examined front view, viewpoint change, and lighting change conditions. We found a very similar pattern across conditions (see Figure 4) suggesting that these relationships are not specific to certain image variations.

--- Insert Figure 3 about here ---

--- Insert Figure 4 about here ---

**3.3.4 Predicting overall face matching performance from eye processing and holistic processing.** Because eye processing and holistic processing selectively predicted match and non-match performance, respectively, we sought to determine if these mechanisms uniquely predicted *overall* face matching performance and how much combined variance they predicted. In the first regression model predicting overall matching accuracy, we found that eye processing ( $t=2.33$ ,  $p=.023$ ) and holistic processing ( $t=2.26$ ,  $p=.027$ ) each significantly predicted unique variance and combined predicted 25% of the variance in matching accuracy ( $F(2,67)=11.19$ ,  $p<.001$ ). Similarly, when predicting overall d-prime, we found that eye processing predicted unique variance ( $t=2.39$ ,  $p=.02$ ) and holistic processing trended toward predicting unique variance ( $t=1.86$ ,  $p=.067$ ) and combined predicted 22% of the variance in d-prime ( $F(2,67)=9.66$ ,  $p<.001$ ).

**3. Discussion**

In the current study, we sought to better understand the mechanisms involved in match vs. non-match judgments by studying individuals with face matching deficits and by examining individual differences associations between match/non-match trials and measures of feature and holistic processing. We recruited a large sample of 43 DPs and found that they had significantly worse overall face matching accuracy compared to controls. When examining match vs. non-match trial performance, we found that DPs’ relative match vs. non-match deficit depended on image variation condition, indicating that DPs do not consistently show match or non-match specific deficits. When examining individual differences associations predicting match and non-match performance across the entire sample of DPs and controls, we found a clear dissociation.

1  
2  
3 In particular, match trials significantly correlated with eye processing ability but not holistic  
4 processing whereas non-match trials significantly correlated with holistic processing but not eye  
5 processing ability. Together, this suggests that matching same-identity faces relies more on the  
6 ability to detect differences in the eye region while discriminating different-identity faces relies  
7 more on holistic processing abilities. These findings have important implications for  
8 understanding face matching deficits in prosopagnosia as well as the mechanisms underlying  
9 face match vs. non-match trial performance.

10  
11 We found that DPs performed worse at face matching compared to controls, but we did  
12 not find that their impairment was specific to match trials, as previously reported in a small DP  
13 study (N=6) by White et al. (2017). Interestingly, we found that DPs' relative match vs. non-  
14 match performance decrement varied as a function of the image matching condition, with no  
15 difference between match and non-match trials for the front view matching condition, a  
16 numerically larger deficit for non-match trials for the viewpoint change condition, and a trend  
17 towards a significantly larger deficit for match trials for the lighting change condition. This  
18 suggests that, as a group, DPs do not consistently show deficits specific to either match or non-  
19 match trials. This aligns with self-reports from DPs suggesting that they struggle with both  
20 matching identities and distinguishing identities (Murray et al., 2018). Even though the results  
21 suggest that DPs generally do not have match or non-match specific deficits, our individual  
22 differences findings suggest that some DPs may be more likely to have deficits with match trials  
23 whereas others may be more likely to have deficits with non-match trials, depending on the  
24 individual underlying impairment. Specifically, our results suggest that DPs with worse eye  
25 processing abilities are more likely to have reduced performance on match trials whereas DPs  
26 with worse holistic processing abilities (and possibly worse mouth processing abilities) are more

likely to have reduced performance on non-match trials. It would be informative in future studies to test whether DP interventions shown to improve holistic processing (e.g., DeGutis et al, 2014) or eye processing (e.g., oxytocin, Bate et al, 2014) would differentially enhance match vs. non-match trial performance.

Beyond providing insights into DP, the current results provide compelling evidence that successfully matching same identity faces is associated with the ability to detect feature changes in the eye region but is not significantly related to holistic processing ability. This is consistent with studies showing that the eye region provides important identity diagnostic information (Royer et al., 2018) and that humans have high perceptual sensitivity to detect differences in the eye region between different identities for both familiar and unfamiliar faces (Abudarham, Shkiller, & Yovel, 2018). The current results extend these studies by showing that individual differences in eye region discrimination ability are specifically related to the ability to match images of the same identity.

Our findings showing a specific relationship between eye processing and match trials seem potentially at odds with instructional intervention face matching studies. For example, Megreya and Bindemann (2018) found that specifically instructing participants to pay attention to the eyes *did not* lead to improvements in matching same identity faces, though instructions to attend to eyebrows did. Further, Towler (2021) found that instructions to pay attention to ears and markings improved non-match trials but not match trial performance. It could be that instructions to attend more to the eyes may not improve matching because most individuals may already be at ceiling in their attention to the eyes and that instructional interventions involving slightly less attended features (e.g., ears or eyebrows) may be more effective in improving overall face matching performance. For example, Ozbek and Bindemann (2011) showed that,

1 during a face matching task, the average first fixation across participants was to the eye region.  
2  
3 That being said, recent studies have shown that individuals reliably vary in their face preferential  
4  
5 face fixation location (PFL, Peterson et al., 2013), the area where one generally fixates on a face,  
6  
7 with some individuals having PFLs closer to the eye region and others closer towards the mouth.  
8  
9 It would be interesting to examine how these PFLs relate to features used during match and non-  
10  
11 match performance and if individuals with different PFLs respond more or less to instructional  
12  
13 interventions.  
14  
15  
16  
17  
18

19 The non-match trial results showed a complementary pattern to the match trial results,  
20  
21 demonstrating a strong association with holistic processing abilities but little-to-no association  
22  
23 with eye processing ability. This is consistent with a previous study that found greater holistic  
24  
25 processing abilities, as measured by both the part-whole effect (similar to the current study) and  
26  
27 face inversion effect, were associated with better overall face matching ability (Rezlescu et al.,  
28  
29 2017). However, the face matching measure used in that study did not allow for separate analysis  
30  
31 of match and non-match trials. This suggests that non-match trials may depend on analysis  
32  
33 across the whole face in contrast to match trials, which may depend more on a feature-by-feature  
34  
35 analysis and an accumulation of feature similarities with less regard for the overall facial gestalt.  
36  
37 It may be that holistic processing provides enhanced sensitivity to the spatial configuration of  
38  
39 features (Mckone & Yovel, 2009), which could be particularly useful for efficiently determining  
40  
41 whether two face images are different identities when particular features are similar (e.g., in the  
42  
43 current study, target and foil faces had generally similar verbal feature descriptions). Holistic  
44  
45 processing abilities may be less useful for determining an exact match between images of the  
46  
47 same identity because these judgments may rely more on matching the intricacies of diagnostic  
48  
49 features.  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

There are several reasons why the current study is the first study, to our knowledge, to find a significant association between holistic processing abilities and non-match accuracy. First, previous studies have typically allowed participants unlimited viewing time during face matching tasks (e.g., Bindemann & Burton, 2020), which may foster a more feature-based approach over holistic processing. Indeed, when given unlimited time, unfamiliar face matching performance has often correlated highly with object performance (e.g., Burton, White, & McNeill, 2010), suggesting a more parts-based approach rather than a holistic approach. Notably, Rezlescu et al.'s (2017) face matching/holistic association was found using a matching task with time pressure. Relatedly, Ozbek and Bindemann (2011) found that match trial performance reached peak accuracy with a shorter viewing duration than non-match trials. With shorter viewing duration, comparing non-match faces feature-by-feature across all the features is challenging and holistic processing may allow for more features and their configuration to be efficiently concurrently compared. Another possible reason why holistic processing/non-match associations have not been previously observed could be because studies used the composite effect to measure holistic processing, which has failed to show significant associations with overall face matching accuracy (e.g., Konar et al., 2010; Rezlescu et al., 2017; VanHellen, Bosten, 2017). Notably, Rezlescu et al., 2017 found significant face matching associations between the face inversion and part-whole effects but not the composite effect. This could either be because the particular version of the composite effect was less reliable than the part-whole and inversion effects (e.g., Rezlescu et al., 2017; Richler et al., 2014) or rather it measures different aspects of holistic processing. For example, the composite effect operationalizes holistic processing as “a failure of selective attention” whereas the part-whole effect measures holistic processing as “the whole is greater than the sum of its parts” (Richler, 2012).

Perhaps unsurprisingly, eye processing and holistic processing each contributed unique variance and together predicted 25% of the overall face matching accuracy variance. The standardized betas of eye processing and holistic processing in the regression model were very similar, suggesting that these mechanisms are equally important to face matching accuracy. Notably, the current results suggest that eye processing and holistic processing are not competing mechanisms to one another in terms of their contribution to face matching, but rather seem to represent important complementary processes that underlie successful face matching judgments for the different trial types. The importance of both eye processing and holistic processing is consistent with studies of the N170 event-related potential, a face-selective marker of perceptual encoding (Bentin et al., 1996). The N170 has been shown to be both sensitive to processing eyes alone (Bentin et al., 1996; Itier et al., 2007; Nemrodov et al., 2014) as well as holistic face processing (Jacques & Rossion, 2010; Nemrodov et al., 2014). Similarly, fMRI has revealed that coordination between the face-selective occipital face area, which responds more to facial features (Henriksson et al., 2015), and the fusiform face area, which has been implicated in holistic processing (Zhang, 2012; Andrews et al., 2010), is important to face perception. Indeed, damage to either area or their connections can result in profound face perception deficits and prosopagnosia (Barton et al., 2008; Steeves et al., 2006).

Though the current findings are compelling, they have limitations that should be addressed in future studies. First, our individual differences sample included a large proportion of DPs, which increased the range of performance and likely improved our sensitivity to detect significant associations. Notably, rather than being a distinctly separate distribution from controls, we found that DPs' perceptual performance across tasks greatly overlapped with controls, albeit slightly shifted towards impairment (as others have reported, Biotti and Cook,

2019). Still, it would be important to replicate the current findings in a normally distributed control sample to better measure the strength of associations in the general population. Additionally, it would also be important for future studies to measure feature processing beyond the eye region to understand if the eye processing/match trial relationship represents a special case or a more general feature phenomenon. Also, we chose to present faces in our matching task for three seconds and it is unclear if the current findings generalize to situations where people have less than three seconds to view an image (e.g., a rushed bartender) or situations where people have more than three seconds to view an image (e.g., forensic examiners). Using a similar task with both shorter and longer presentation times would be useful to understand how eye processing and holistic processing change in their contribution to match and non-match trials with viewing time. An additional limitation is that some of our trials included cropped images, which excluded the ears. This was done to prevent individuals from solely matching faces based on external features. However, this study should be replicated with uncropped images and more external feature abilities should be assessed in order to examine whether external feature abilities, such as with the ears, improves face matching as recent studies have suggested (e.g., Towler et al., 2021). Finally, we found individual difference associations that were similar across lighting changes, viewpoint changes, and front-view conditions for unfamiliar faces, yet it is unclear whether this would extend to familiar faces or other image translations and face variations, such as across emotions. This would be important to investigate in future research.

4. Conclusion

Overall, our findings provide novel evidence for distinct mechanisms involved in match vs. non-match trial judgments. Though we did not find that DPs have specific deficits in matching images of the same person, our results clearly show that eye region discrimination

ability is related to better match performance whereas enhanced holistic processing ability is related to better non-match performance. Future research would be useful to determine whether these findings generalize to more real-world face matching circumstances (e.g., passport control officers) and to further examine the complementary nature of eye processing and holistic processing in face matching tasks and perceptual abilities in general.

### Acknowledgements

We want to thank the developmental prosopagnosics and control participants for completing our challenging battery of tasks. This study was funded by R01 from the National Eye Institute grant #R01EY026057 awarded to JD. AKB is funded by the Leverhulme Trust Early Career Fellowship; grant number ECF-2019-416.

### Data Accessibility

Upon acceptance of the manuscript, all data will made be publicly available on Dryad in accordance with NIH policy and with the approval of the VA Boston Healthcare System and Harvard Medical School IRBs.

### Supplementary Material

The Supplementary Material is available at: [qjep.sagepub.com](http://qjep.sagepub.com)

References

Abudarham, N., Shkiller, L., & Yovel, G. (2019). Critical features for face recognition. *Cognition*, 182, 73-83.

Abudarham, N., Bate, S., Duchaine, B., & Yovel, G. (2021). Developmental prosopagnosics and super recognizers rely on the same facial features used by individuals with normal face recognition abilities for face identification. *Neuropsychologia*, 160, 107963.

Adams, A., Hills, P. J., Bennetts, R. J., & Bate, S. (2019). Coping strategies for developmental prosopagnosia. *Neuropsychological Rehabilitation*, 1–20.  
doi:10.1080/09602011.2019.1623824

Andrews, T. J., Davies-Thompson, J., Kingstone, A., & Young, A. W. (2010). Internal and external features of the face are represented holistically in face-selective regions of visual cortex. *Journal of Neuroscience*, 30(9), 3544-3552.

Avidan, G., Tanzer, M., & Behrmann, M. (2011). Impaired holistic processing in congenital prosopagnosia. *Neuropsychologia*, 49(9), 2541-2552.  
doi:10.1016/j.neuropsychologia.2011.05.002

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.

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.

Bentin, S., Allison, T., Puce, A., Perez, E., & McCarthy, G. (1996). Electrophysiological studies of face perception in humans. *Journal of cognitive neuroscience*, 8(6), 551-565.

- Bindemann, M., Burton, M. A. (2020). Steps Towards a Cognitive Theory of Unfamiliar Face Matching. In 1162974601 873005493 M. Bindemann (Author), *FORENSIC FACE MATCHING* (pp. 38-61). Place of publication not identified, NY: OXFORD UNIV Press
- Biotti, F., Gray, K. L., & Cook, R. (2019). Is developmental prosopagnosia best characterised as an apperceptive or mnemonic condition?. *Neuropsychologia*, 124, 285-298.
- Burton, A. M., White, D., & McNeill, A. (2010). The Glasgow Face Matching Test. *Behavior Research Methods*, 42(1), 286-291.
- Corrow, S. L., Dalrymple, K. A., & Barton, J. J. (2016). Prosopagnosia: current perspectives. *Eye and brain*, 8, 165.
- Davis JP, Dray C, Petrov N, Belanova E. Low prevalence match and mismatch detection in simultaneous face matching: Influence of face recognition ability and feature focus guidance. *Atten Percept Psychophys*. 2021 Oct;83(7):2937-2954. doi: 10.3758/s13414-021-02348-4. Epub 2021 Aug 13. PMID: 34386883.
- DeGutis, J., Cohen, S., Mercado, R., Wilmer, J., & Nakayama, K. (2012) Holistic processing of the mouth but not the eyes in developmental prosopagnosia, *Cognitive Neuropsychology*, 29:5-6, 419-446, DOI: [10.1080/02643294.2012.754745](https://doi.org/10.1080/02643294.2012.754745)
- DeGutis, J., Chatterjee, G., Mercado, R. J., & Nakayama, K. (2012). Face gender recognition in developmental prosopagnosia: Evidence for holistic processing and use of configural information. *Visual Cognition*, 20(10), 1242-1253.
- DeGutis, J., Wilmer, J., Mercado, R. J., & Cohan, S. (2013). Using regression to measure holistic face processing reveals a strong link with face recognition ability. *Cognition*, 126(1), 87-100. <https://doi.org/10.1016/j.cognition.2012.09.004>

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

Duchaine, B. C., & Nakayama, K. (2006a). Developmental prosopagnosia: a window to content-specific face processing. *Current opinion in neurobiology*, 16(2), 166-173.

Duchaine, B., & Nakayama, K. (2006b). 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., Yovel, G., & Nakayama, K. (2007). No global processing deficit in the Navon task in 14 developmental prosopagnosics. *Social cognitive and affective neuroscience*, 2(2), 104-113.

Farah, M. J., Wilson, K. D., Drain, M., & Tanaka, J. N. (1998). What is "special" about face perception? *Psychological Review*, 105(3), 482-498. doi:10.1037/0033-295x.105.3.482

Fisher, K., Towler, J., & Eimer, M. (2016). Reduced sensitivity to contrast signals from the eye region in developmental prosopagnosia. *Cortex*, 81, 64-78. doi:10.1016/j.cortex.2016.04.005

Garrido, L., Duchaine, B., & Nakayama, K. (2008). Face detection in normal and prosopagnosic individuals. *Journal of Neuropsychology*, 2(1), 119-140.

Glanzer, M., Adams, J. K., Iverson, G. J., & Kim, K. (1993). The regularities of recognition memory. *Psychological Review*, 100(3), 546-567. https://doi.org/10.1037/0033-295X.100.3.546

Gross, R., Matthews, I., Cohn, J., Kanade, T., & Baker, S. (2010). Multi-pie. *Image and vision computing*, 28(5), 807-813.

- Henriksson, L., Mur, M., & Kriegeskorte, N. (2015). Faciotopy—A face-feature map with face-like topology in the human occipital face area. *Cortex*, 72, 156-167.
- Itier, R. J., Alain, C., Sedore, K., & McIntosh, A. R. (2007). Early face processing specificity: It's in the eyes!. *Journal of cognitive neuroscience*, 19(11), 1815-1826.
- Jacques, C., & Rossion, B. (2010). Misaligning face halves increases and delays the N170 specifically for upright faces: Implications for the nature of early face representations. *Brain Research*, 1318, 96-109.
- Konar, Yaroslav & Bennett, Patrick & Sekuler, Allison. (2010). Holistic Processing Is Not Correlated With Face-Identification Accuracy. *Psychological science*. 21. 38-43.  
10.1177/0956797609356508.
- Malcolm, G. L., Leung, C., & Barton, J. J. (2004). Regional variation in the inversion effect for faces: differential effects for feature shape, feature configuration, and external contour. *Perception*, 33(10), 1221–1231. <https://doi.org/10.1068/p5372>
- Mckone, E., & Yovel, G. (2009). Why does picture-plane inversion sometimes dissociate perception of features and spacing in faces, and sometimes not? Toward a new theory of holistic processing. *Psychonomic Bulletin & Review*, 16(5), 778-797.  
doi:10.3758/pbr.16.5.778
- Megreya, A.M., Burton, A.M. (2006). Unfamiliar faces are not faces: Evidence from a matching task. *Memory & Cognition* 34, 865–876. <https://doi.org/10.3758/BF03193433>
- Megreya, A. M., & Burton, A. M. (2007). Hits and false positives in face matching: A familiarity-based dissociation. *Perception & Psychophysics*, 69(7), 1175-1184.  
doi:10.3758/bf03193954

Megreya, A. M. (2018). Feature-by-feature comparison and holistic processing in unfamiliar face matching. *PeerJ*, 6. doi:10.7717/peerj.4437

Megreya, A. M., & Bindemann, M. (2018). Feature instructions improve face-matching accuracy. *PloS one*, 13(3), e0193455. <https://doi.org/10.1371/journal.pone.0193455>

Mishra, M. V., Fry, R. M., Saad, E., Arizpe, J. M., Ohashi, Y. G. B., & DeGutis, J. M. (2021). Comparing the sensitivity of face matching assessments to detect face perception impairments. *Neuropsychologia*, 108067.

Murray, E., Hills, P. J., Bennetts, R. J., & Bate, S. (2018). Identifying hallmark symptoms of developmental prosopagnosia for non-experts. *Scientific Reports*, 8(1), 1-12.

Nemrodov, D., Anderson, T., Preston, F. F., & Itier, R. J. (2014). Early sensitivity for eyes within faces: A new neuronal account of holistic and featural processing. *NeuroImage*, 97, 81-94.

Ozbek, 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. <https://doi.org/10.1016/j.visres.2011.08.009>

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. doi:10.1016/j.neuropsychologia.2011.02.021

Phillips, P. J., Yates, A. N., Hu, Y., Hahn, C. A., Noyes, E., Jackson, K., ... & O'Toole, A. J. (2018). Face recognition accuracy of forensic examiners, superrecognizers, and face recognition algorithms. *Proceedings of the National Academy of Sciences*, 115(24), 6171-6176.

- Ramon, M., Bobak, A. K., & White, D. (2019). Super-recognizers: From the lab to the world and back again. *British Journal of Psychology*, 110(3), 461-479.
- Rezlescu, C., Susilo, T., Wilmer, J. B., & Caramazza, A. (2017). The inversion, part-whole, and composite effects reflect distinct perceptual mechanisms with varied relationships to face recognition. *Journal of Experimental Psychology: Human Perception and Performance*, 43(12), 1961–1973. <https://doi.org/10.1037/xhp0000400>
- Richler, J. J., Cheung, O. S., & Gauthier, I. (2011). Holistic processing predicts face recognition. *Psychological science*, 22(4), 464-471.
- Richler, J. J., Palmeri, T. J., & Gauthier, I. (2012). Meanings, Mechanisms, and Measures of Holistic Processing. *Frontiers in Psychology*, 3. doi:10.3389/fpsyg.2012.00553
- Richler, J. J., Floyd, R. J., & Gauthier, I. (2014). The Vanderbilt Holistic Face Processing Test: A short and reliable measure of holistic face processing. *Journal of Vision*, 14(11), 10-10.
- Rossion, B., & Boremanse, A. (2008). Nonlinear relationship between holistic processing of individual faces and picture-plane rotation: evidence from the face composite illusion. *Journal of vision*, 8(4), 1–13. <https://doi.org/10.1167/8.4.3>
- Rossion, B., & Michel, C. (2018). Normative accuracy and response time data for the computerized Benton Facial Recognition Test (BFRT-c). *Behavior Research Methods*, 50(6), 2442-2460. doi:10.3758/s13428-018-1023-x
- Rossion, B. (2018). Humans are visual experts at unfamiliar face recognition. *Trends in cognitive sciences*, 22(6), 471-472
- Royer, J., Blais, C., Charbonneau, I., Déry, K., Tardif, J., Duchaine, B., . . . Fiset, D. (2018). Greater reliance on the eye region predicts better face recognition ability. *Cognition*, 181, 12-20. doi:10.1016/j.cognition.2018.08.004

Russell, R., Duchaine, B., & Nakayama, K. (2009). Super-recognizers: people with extraordinary face recognition ability. *Psychonomic bulletin & review*, 16(2), 252–257.  
<https://doi.org/10.3758/PBR.16.2.252>

Sadr, J., Jarudi, I., & Sinha, P. (2003). The Role of Eyebrows in Face Recognition. *Perception*, 32(3), 285-293. doi:10.1068/p5027

Schmidt, F. L., & Hunter, J. E. (1996). Measurement error in psychological research: Lessons from 26 research scenarios. *Psychological Methods*, 1(2), 199.

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

Steeves, J. K., Culham, J. C., Duchaine, B. C., Pratesi, C. C., Valyear, K. F., Schindler, I., ... & Goodale, M. A. (2006). The fusiform face area is not sufficient for face recognition: evidence from a patient with dense prosopagnosia and no occipital face area. *Neuropsychologia*, 44(4), 594-609.

Stumps, A., Saad, E., Rothlein, D., Verfaellie, M., & DeGutis, J. (2020). Characterizing developmental prosopagnosia beyond face perception: Impaired recollection but intact familiarity recognition. *cortex*, 130, 64-77.

Tanaka, J. W., Kiefer, M., & Bukach, C. (2004). A holistic account of the own-race effect in face recognition: Evidence from a cross-cultural study. *Cognition*, 93(1), B1-B9.  
[doi:https://doi.org/10.1016/j.cognition.2003.09.011](https://doi.org/10.1016/j.cognition.2003.09.011)

Tanaka, J., & Xu, B. (2018). Holistic gist: the speed of holistic face processing. *Journal of Vision*, 18(10), 166-166.

- Taubert, J., Apthorp, D., Aagten-Murphy, D., & Alais, D. (2011). The role of holistic processing in face perception: evidence from the face inversion effect. *Vision research*, 51(11), 1273–1278. <https://doi.org/10.1016/j.visres.2011.04.002>
- Torfs, K., Vancleef, K., Lafosse, C., Wagemans, J., & de-Wit, L. (2014). The Leuven Perceptual Organization Screening Test (L-POST), an online test to assess mid-level visual perception. *Behavior Research Methods*, 46(2), 472–487.
- Towler, Alice & White, David & Kemp, Richard. (2017). Evaluating the Feature Comparison Strategy for Forensic Face Identification. *Journal of Experimental Psychology Applied*. 23. 47-58. 10.1037/xap0000108.
- Towler, A., Keshwa, M., Ton, B., Kemp, R. I., & White, D. (2021). Diagnostic feature training improves face matching accuracy. *Journal of experimental psychology. Learning, memory, and cognition*, 10.1037/xlm0000972. Advance online publication. <https://doi.org/10.1037/xlm0000972>
- Ulrich, P. I., Wilkinson, D. T., Ferguson, H. J., Smith, L. J., Bindemann, M., Johnston, R. A., & Schmalzl, L. (2017). Perceptual and memorial contributions to developmental prosopagnosia. *Quarterly journal of experimental psychology (2006)*, 70(2), 298–315. <https://doi.org/10.1080/17470218.2016.1177101>
- Verhallen, R. J., Bosten, J. M., Goodbourn, P. T., Lawrance-Owen, A. J., Bargary, G., & Mollon, J. (2017). General and specific factors in the processing of faces. *Vision Research*, 141, 217–227. doi:10.1016/j.visres.2016.12.014
- Vinette, Céline & Gosselin, Frédéric & Schyns, Philippe. (2004). Spatio-temporal dynamics of face recognition in a flash: It's in the eyes. *Cognitive Science*. 28. 289–301. 10.1016/j.cogsci.2004.01.002.

Wang, M and Stanley, J (1970) ‘Differential weighting: A review of methods and empirical studies’, *Review of Educational Research*, 40, pp. 663-705.

White, David & Kemp, Richard & Jenkins, Rob & Matheson, Michael & Burton, A. (2014). Passport Officers' Errors in Face Matching. *PloS one*. 9. e103510. 10.1371/journal.pone.0103510.

White, D., Rivolta, D., Burton, A. M., Al-Janabi, S., & Palermo, R. (2017). Face Matching Impairment in Developmental Prosopagnosia. *Quarterly Journal of Experimental Psychology*, 70(2), 287–297. <https://doi.org/10.1080/17470218.2016.1173076>

Young, A. W. & Burton, A. M. (2018). Are We Face Experts? *Trends in Cognitive Sciences*, 22(2), 100-110. doi:10.1016/j.tics.2017.11.007

Zhang, J., Li, X., Song, Y., & Liu, J. (2012). The fusiform face area is engaged in holistic, not parts-based, representation of faces. *PloS one*, 7(7), e40390.

Zhao, M., Bülthoff, H. H., & Bülthoff, I. (2016). Beyond faces and expertise: Facelike holistic processing of nonface objects in the absence of expertise. *Psychological science*, 27(2), 213-222.

### Figure Captions

Figure 1. Match trial and non-match trial accuracy across front view, viewpoint change, and lighting change conditions in DPs and controls. Error bars represent the standard error of the mean.  $**p < .01$  for uncorrected DP/control differences for a specific condition.

Figure 2. Match trial and non-match trial accuracy correlated with the different feature conditions and holistic processing for controls ( $N=27$ ), developmental prosopagnosics (DPs,  $N=43$ ), and the combined group of controls and DPs ( $N=70$ ). The two shaded sections represent facial features in the upper (left) and lower (right) halves of the face.  $*p < .05$  and  $**p < .01$  uncorrected Pearson correlation.

Figure 3. Average match trial and non-match trial accuracy correlated with either composite eye accuracy or holistic advantage from the part-whole task. The black line is the trendline for the overall group whereas the blue dotted line is the trendline for controls and orange dotted line is for DPs.  $**p < .01$  uncorrected Pearson correlation for the entire group.

Figure 4. Match trial and non-match trial accuracy correlated with either composite eye accuracy or holistic advantage from the part-whole task for overall average, front view condition, viewpoint change condition, and lighting change condition.  $*p < .05$  and  $**p < .01$  uncorrected Pearson correlation for the entire group.

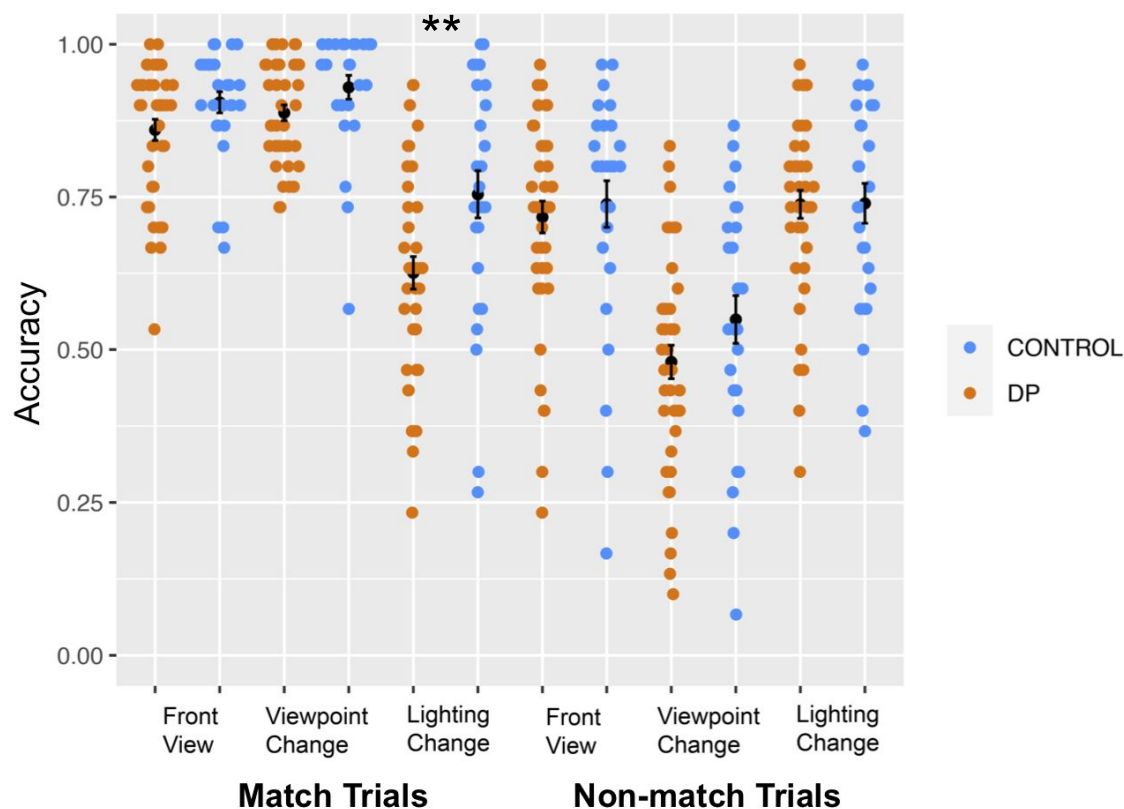


Figure 1. Match trial and non-match trial accuracy across front view, viewpoint change, and lighting change conditions in DPs and controls. Error bars represent the standard error of the mean. \*\* $p < .01$  for uncorrected DP/control differences for a specific condition.

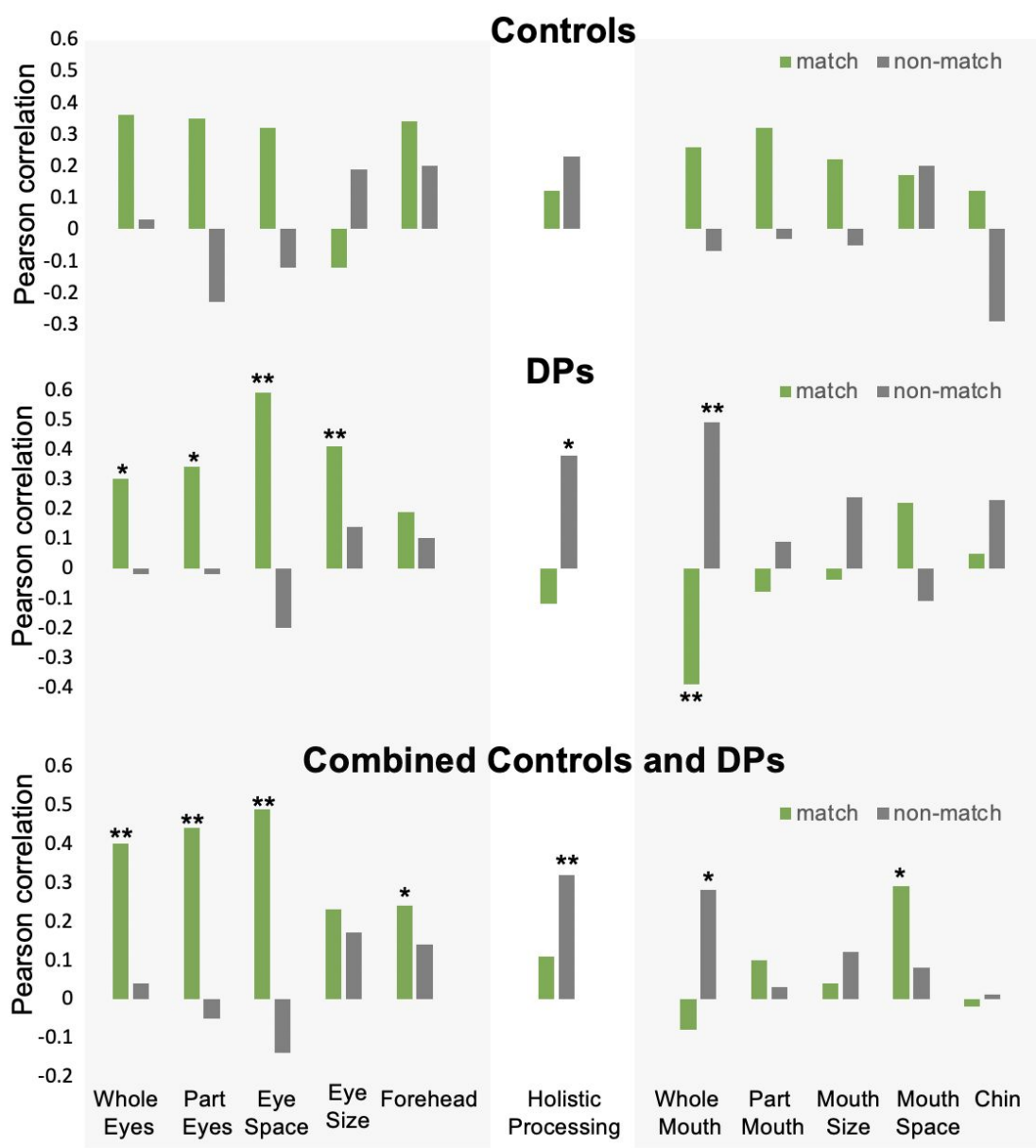


Figure 2. Match trial and non-match trial accuracy correlated with the different feature conditions and holistic processing for controls (N=27), developmental prosopagnosics (DPs, N=43), and the combined group of controls and DPs (N=70). The two shaded sections represent facial features in the upper (left) and lower (right) halves of the face. \* $p < .05$  and \*\* $p < .01$  uncorrected Pearson correlation.

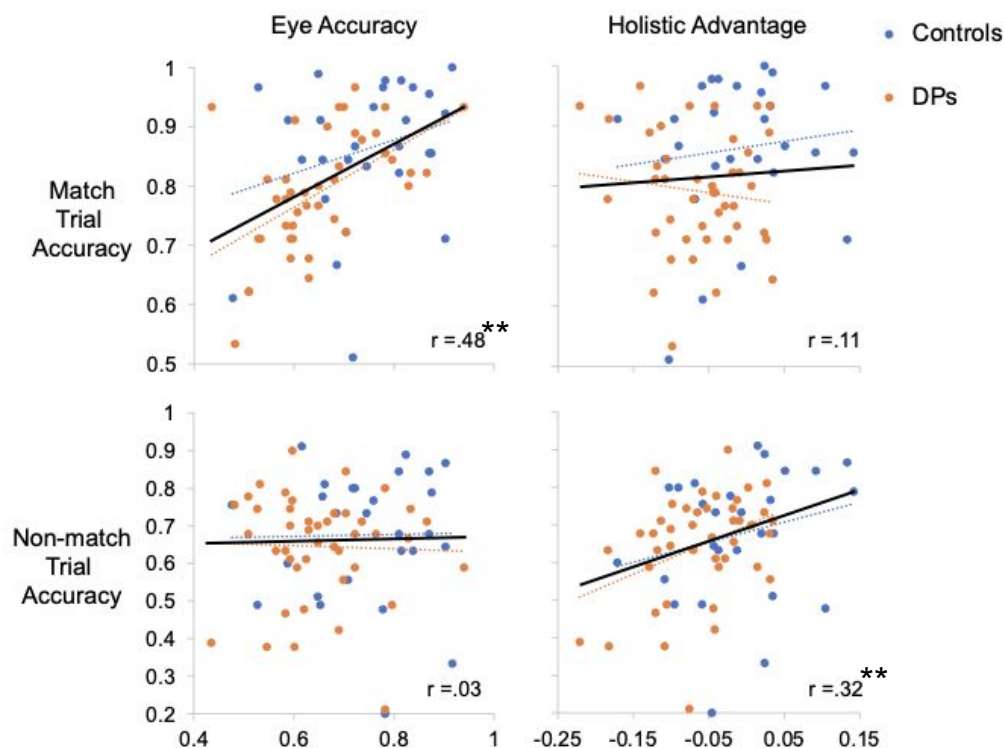


Figure 3. Average match trial and non-match trial accuracy correlated with either composite eye accuracy or holistic advantage from the part-whole task. The black line is the trendline for the overall group whereas the blue dotted line is the trendline for controls and orange dotted line is for DPs.  $**p < .01$  uncorrected Pearson correlation for the entire group.

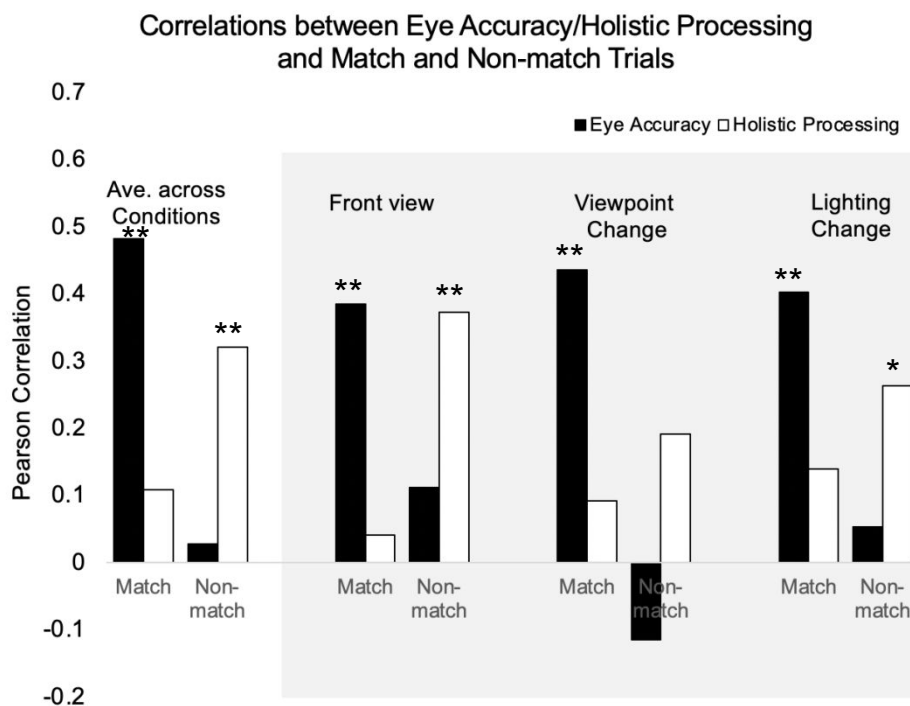


Figure 4. Match trial and non-match trial accuracy correlated with either composite eye accuracy or holistic advantage from the part-whole task for overall average, front view condition, viewpoint change condition, and lighting change condition. \* $p < .05$  and \*\* $p < .01$  uncorrected Pearson correlation for the entire group.

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

**Table 1.** DP Demographics, Face Recognition, and Face Matching Performance compared to controls.

DP Participant #	Age	Gender	CFMT	PI-20	Face Matching Match Trial Accuracy	Face Matching Non-Match Trial Accuracy
1	22	F	34	88	0.76	0.59
2	29	F	37	88	0.84	0.49
3	61	M	38	89	0.81	0.38
4	23	F	32	92	0.71	0.74
5	70	F	36	83	0.62	0.78
6	27	F	42	76	0.77	0.61
7	50	M	44	82	0.83	0.63
8	41	M	35	90	0.78	0.63
9	23	F	46	93	0.93	0.21
10	25	F	44	80	0.93	0.42
11	32	M	45	61	0.64	0.71
12	38	F	44	87	0.78	0.63
13	52	F	43	92	0.62	0.68
14	64	F	39	82	0.71	0.81
15	52	F	38	87	0.53	0.76
16	51	F	45	92	0.91	0.38
17	36	F*	35	93	0.79	0.48
18	33	M	36	80	0.93	0.39
19	46	F	34	75	0.71	0.74
20	32	F	40	58	0.74	0.64
21	63	F	37	63	0.68	0.69
22	55	F	33	96	0.81	0.47
23	37	M	33	91	0.77	0.66
24	53	F	35	86	0.89	0.59
25	26	F	42	80	0.82	0.71
26	31	F	37	89	0.68	0.70
27	39	F	33	78	0.81	0.73
28	28	F	42	80	0.88	0.71
29	31	F	47	87	0.90	0.71
30	35	F	43	81	0.93	0.56
31	30	F	41	69	0.80	0.67
32	36	F	47	77	0.72	0.73
33	27	F	44	86	0.72	0.84
34	34	F	39	75	0.89	0.68
35	39	F	41	71	0.79	0.61
36	57	M	44	87	0.82	0.74
37	20	M	40	72	0.73	0.77
38	27	M	38	80	0.71	0.90
39	28	F	39	80	0.97	0.68
40	64	F	43	85	0.73	0.79
41	25	F	44	88	0.93	0.59
42	20	F	48	78	0.80	0.70
43	25	F	39	81	0.86	0.80
DP Average	38.1 ± 13.9	34:9	39.9 ± 4.4	82.0 ± 8.8	0.79 ± 0.10	0.64 ± 0.14
Control Average	40.0 ± 11.7	19:11	59.8 ± 7.8		0.86 ± 0.12	0.68 ± 0.17

*Note.* Mean ± Standard Deviation, CFMT=Cambridge Face Memory Test, PI-20=prosopagnosia index 20. \*Participant's gender identity is nonbinary.

Table 2. Match trial and non-match trial mean reaction times (standard deviation) across front view, viewpoint change, and lighting change conditions in control and DP groups. \*\* $p < .01$  for uncorrected DP/control differences for a specific condition.

	Match Trials			Non-match Trials		
	Front view	Viewpoint change	Lighting change	Front view	Viewpoint change	Lighting change
<b>Controls</b>	1200(301)**	1363(584)**	1601(415)**	1600(418)**	1747(440)**	1628(446)**
<b>DPs</b>	1540(415)	1564(361)	1942(447)	1984(500)	2043(476)	1989(442)