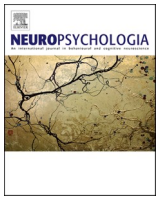
Neuropsychologia 163 (2021) 108067

Contents lists available at ScienceDirect

Neuropsychologia

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



Comparing the sensitivity of face matching assessments to detect face perception impairments

Maruti V. Mishra [a,b], Regan M. Fry [a,b], Elyana Saad [a], Joseph M. Arizpe [c], Yuri-Grace B. Ohashi [d,e], Joseph M. DeGutis [a,b,*]

[a] Department of Psychiatry, Harvard Medical School, Boston, MA, USA
[b] Boston Attention and Learning Laboratory, VA Boston Healthcare, Jamaica Plain Division, 150 S Huntington Ave., Boston, MA, USA
[c] Science Applications International Corporation (SAIC), Fort Sam Houston, TX, USA
[d] Department of Psychology, Harvard University, Cambridge, MA, USA
[e] Harvard Decision Science Laboratory, Harvard Kennedy School, Cambridge, MA, USA

ARTICLE INFO

Keywords:
Face perception
Face matching
Developmental prosopagnosia
Face recognition

ABSTRACT

Numerous neurological, developmental, and psychiatric conditions demonstrate impaired face recognition, which can be socially debilitating. These impairments can be caused by either deficient face *perception* or face *memory* mechanisms. Though there are well-validated, sensitive measures of face memory impairments, it currently remains unclear which assessments best measure face perception impairments. A sensitive, validated face perception measure could help with diagnosing causes of face recognition deficits and be useful in characterizing individual differences in unimpaired populations. Here, we compared the computerized Benton Face Recognition Test (BFRT-*c*) and Cambridge Face Perception Test (CFPT) in their ability to differentiate developmental prosopagnosics (DPs, N = 30) and age-matched controls (N = 30). Participants completed the BFRT-c, CFPT, and two additional face perception assessments: the University of Southern California Face Perception Test (USCFPT) and a novel same/different face matching test (SDFMT). Participants were also evaluated on objective and subjective face recognition tasks including the Cambridge Face Memory Test, famous faces test, and Prosopagnosia Index-20. We performed a logistic regression with the perception tests predicting DP vs. control group membership and used multiple linear regressions to predict continuous objective and subjective face recognition memory. Our results show that the BFRT-*c* performed as well as, if not better than, the CFPT, and that both tests clearly outperformed the USCFPT and SDFMT. Further, exploratory analyses revealed that face lighting-change conditions better predicted DP group membership and face recognition abilities than viewpoint-change conditions. Together, these results support the combined use of the BFRT-*c* and CFPT to best assess face perception impairments.

1. Introduction

Faces convey signals essential for social interactions and are one of the most reliable ways to determine a person's identity. Recognizing a face is a highly specialized, multistage process (Bruce and Young, 1986) involving a network of brain regions (Grill-Spector et al., 2017; Haxby et al., 2000; Haxby and Gobbini, 2011). Though recognition of familiar faces is typically rapid and effortless (Jenkins et al., 2018), this ability can be severely impaired in many neurological, psychiatric, and developmental disorders including prosopagnosia (Albonico and Barton, 2019; Mayer and Rossion, 2007), autism spectrum disorders (Dwyer et al., 2019; Weigelt et al., 2012), schizophrenia (Watson, 2013), Alzheimer's disease (Lavallée et al., 2016), person recognition disorders (Gainotti, 2007), age-related cognitive decline (Boutet et al., 2015), as well as others (Barton et al., 2004; Dimitriou et al., 2015). The causes of face recognition deficits in these disorders vary, and it is crucial to identify measures that can discern the stages of face identification impairments to help with both diagnosis (Benton and Van Allen, 1968; Benton et al., 1994; Duchaine and Nakayama, 2006; Duchaine et al., 2007) and treatment (Bate and Bennetts, 2014; DeGutis et al., 2014a, b).

* Corresponding author. Department of Psychiatry, Harvard Medical school | Boston Attention and Learning Laboratory, VA Boston Healthcare, Jamaica Plain Division, 150 S Huntington Ave., Boston, MA. USA.
E-mail address: degutis@hms.harvard.edu (J.M. DeGutis).

https://doi.org/10.1016/j.neuropsychologia.2021.108067
Received 3 March 2021; Received in revised form 15 October 2021; Accepted 17 October 2021
Available online 19 October 2021
0028-3932/© 2021 Elsevier Ltd. All rights reserved.

When assessing face recognition abilities, it is important to consider there are two stages of processing involved: *face perception* and *face memory* (e.g., De Renzi et al., 1991; Liu et al., 2002; Weigelt et al., 2014). *Face perception* refers to building up/encoding a structural representation of a face (Bruce and Young, 1986). This representation allows one to judge whether simultaneously presented faces belong to the same or different individuals. On the other hand, *face memory* involves the ability to store and retrieve individuated faces from long-term memory. Face memory relies on face perception but also involves processes such as associating semantic and contextual information with a face, storing and retrieving a face and related semantic/contextual information, and, in the case of familiar faces, building up a robust face representation over repeated presentations. One influential developmental study suggests that face perception develops earlier (~age 5) while face memory demonstrates later face-specific development (~age 10, Weigelt et al., 2014). Additionally, patient studies have found dissociations between acquired prosopagnosics with impaired face perception and intact face memory (apperceptive prosopagnosia) and vice versa (associative prosopagnosia, De Renzi et al., 1991; though see Busigny et al., 2014).

Face perception impairments are typically assessed using face matching tasks where the to-be-matched faces are presented simultaneously and, importantly, vary in either viewpoint, lighting, or emotion (Benton and Van Allen, 1968; Duchaine et al., 2007; White et al., 2017). Such variations in face images prevent direct image-based matching, are thought to make judgments rely more on specialized face-specific perceptual mechanisms (e.g., holistic processing, McKone, 2008; Tanaka and Farah, 1993), and better capture impairments in patients with face perception deficits. Face memory is assessed using tasks that involve learning and short-term retention of identities of novel faces, such as the Cambridge Face Memory Test (CFMT: Duchaine and Nakayama, 2006) or long-term recognition of familiar/famous faces (Bobak et al., 2017; Duchaine and Nakayama, 2005). Although many researchers have regarded the CFMT as the gold standard test for assessing deficits in face memory, there is currently no widely agreed-upon measure to reliably characterize face perception impairments. Notably, because perception precedes memory, all visual memory tests including the CFMT depend on both perceptual and memory processes, thus making it difficult to dissociate the independent contributions of the two. For example, studies have reported that factors that impair face matching performance (e.g., lighting, viewpoint changes) similarly impair short-term and long-term memory for faces (Braje et al., 1998; Braje, 2003). The goal of the current study was to identify tests that can best assess face perception in impaired samples.

In the past two decades, numerous tests have been developed to assess face perception (e.g., Duchaine et al., 2007; Burton et al., 2010; Fysh and Bindemann, 2018). The Benton Face Recognition Test[1] (BFRT, Benton and Van Allen, 1968) was one of the first assessments to provide a standardized measure of face perception proficiency to assess deficits. The BFRT requires matching the identity of a front-view target face to three of six faces simultaneously presented, which may vary in lighting or viewpoint, (see Fig. 1). This test has been widely used to assess perceptual deficits in clinical disorders like cortical blindness, lobectomy, visual agnosia, and autism (Benton et al., 1994; Busigny et al., 2009; Demirci and Erdogan, 2016; Duchaine, 2000; Yerys et al., 2018). However, the sensitivity of the original, unspeeded version of the BFRT has been challenged in studies of developmental prosopagnosics (DPs), individuals with severe lifelong impairments in face recognition and group-level face perception deficits (Duchaine and Nakayama, 2004; Duchaine and Weidenfeld, 2003; Nunn et al., 2001). In particular, Duchaine and Nakayama (2004) found that 9 of 11 DPs performed normally on the BFRT (>41 out of 54, z-score > −1). Further, a recent study reported that in 23 DPs, 18 scored within the normal range on the original BFRT (Albonico et al., 2017).

One potentially critical factor with the original version of the BFRT is that it had no time restrictions and did not emphasize response speed. Studies have shown that when given unlimited time to perform face matching, feature-by-feature comparison becomes an available strategy to achieve high accuracy rates (e.g., Towler et al., 2017). Along these lines, studies have reported that prosopagnosics take significantly longer to complete the original BFRT than controls (Albonico et al., 2017; Duchaine, 2000; Duchaine and Nakayama, 2004; Nunn et al., 2001), as they may use a feature-by-feature strategy to achieve accuracy within the normal range. This suggests that speeded tasks may be better at identifying impairments in face perception. In an effort to address these shortcomings of the original BFRT, Rossion and Michel (2018) developed a computerized version of the BFRT (BFRT-*c*) that collects response times and instructs participants to respond, "as quickly and accurately as possible". The normative accuracy of this version (see Dzhelyova et al., 2020) is notably 0.75 *SDs* lower than versions without speeded instructions (e.g., Albonico et al., 2017). This suggests that the BFRT-*c*'s speeded instructions may change participants' strategy, possibly favoring face-specific processes (e.g., holistic processing) rather than more laborious, non-face specific feature matching processes. This may make the BFRT-*c* a substantially more sensitive assessment of face perception impairment than the original version or other currently available assessments.

To date, there have been no studies comparing the validity and sensitivity of this version of the BFRT-*c* or other available face perception tests in clinical samples. DPs are an ideal population to validate face perception tests, as larger samples of DPs consistently show reduced group-level face matching performance compared to controls (e.g., Biotti et al., 2019; Dalrymple et al., 2014; Zhao et al., 2016; though individual DP cases may demonstrate normal face perception) and are prevalent enough to obtain adequate group sample sizes (2.5% of population; Kennerknecht et al., 2006). In addition to differentiating between DPs and controls, a good face perception assessment should also strongly correlate with measures of face memory, since face perception ability is a significant contributor to face memory performance (e.g., Stumps et al., 2020). In the current study, we compared the BFRT-*c* and the widely used Cambridge Face Perception Test (CFPT, Duchaine et al., 2007; Bobak et al., 2017; Bowles et al., 2009; Gonzalez-Perez et al., 2019; Palermo et al., 2017; Rezlescu et al., 2017), in their ability to a) detect face perception impairments and b) predict face memory performance. In the CFPT, participants are required to arrange six front-view morphed faces from most-to-least similar to a target face shown from a ¾-viewpoint. Two important differences between CFPT and BFRT-*c* are that a) speed is emphasized in the BFRT-c whereas in the CFPT there is less emphasis on speed, with 60 s provided to complete each trial, and b) the BFRT-*c* includes lighting and viewpoint changes between faces whereas the CFPT only includes viewpoint change trials. Studies have suggested that prosopagnosics may be particularly worse at matching faces across lighting change trials, even more so than viewpoint change trials (Duchaine and Nakayama, 2004; Rossion and Michel, 2018). Given that BFRT-*c* is speeded and includes lighting change trials, we hypothesized that the BFRT-*c* will predict unique variance from the CFPT in DP vs. control group membership as well as objective and subjective face recognition ability.

In exploratory analyses, we also sought to compare the BFRT-*c* and CFPT to two additional face matching assessments shown to be sensitive to face perception impairments. First, we included the USC Face Perception Test (USCFPT, Margalit et al., 2016; Biederman et al., 2017) that uses computer-generated faces and difficulty levels scaled according to the Gabor-jet model and has shown decreased performance in DPs (Biederman et al., 2017; Margalit et al., 2016; Yue et al., 2012). We also included a novel same/different face matching task (SDFMT) (motivated by White et al., 2017), which demonstrated significant differences

[1] We prefer calling BFRT-*c* a face perception test, to distinguish it from face recognition tests, as the term recognition refers to memory specifically "the ability to identify information as having been encountered before" (APA Dictionary of Psychology).

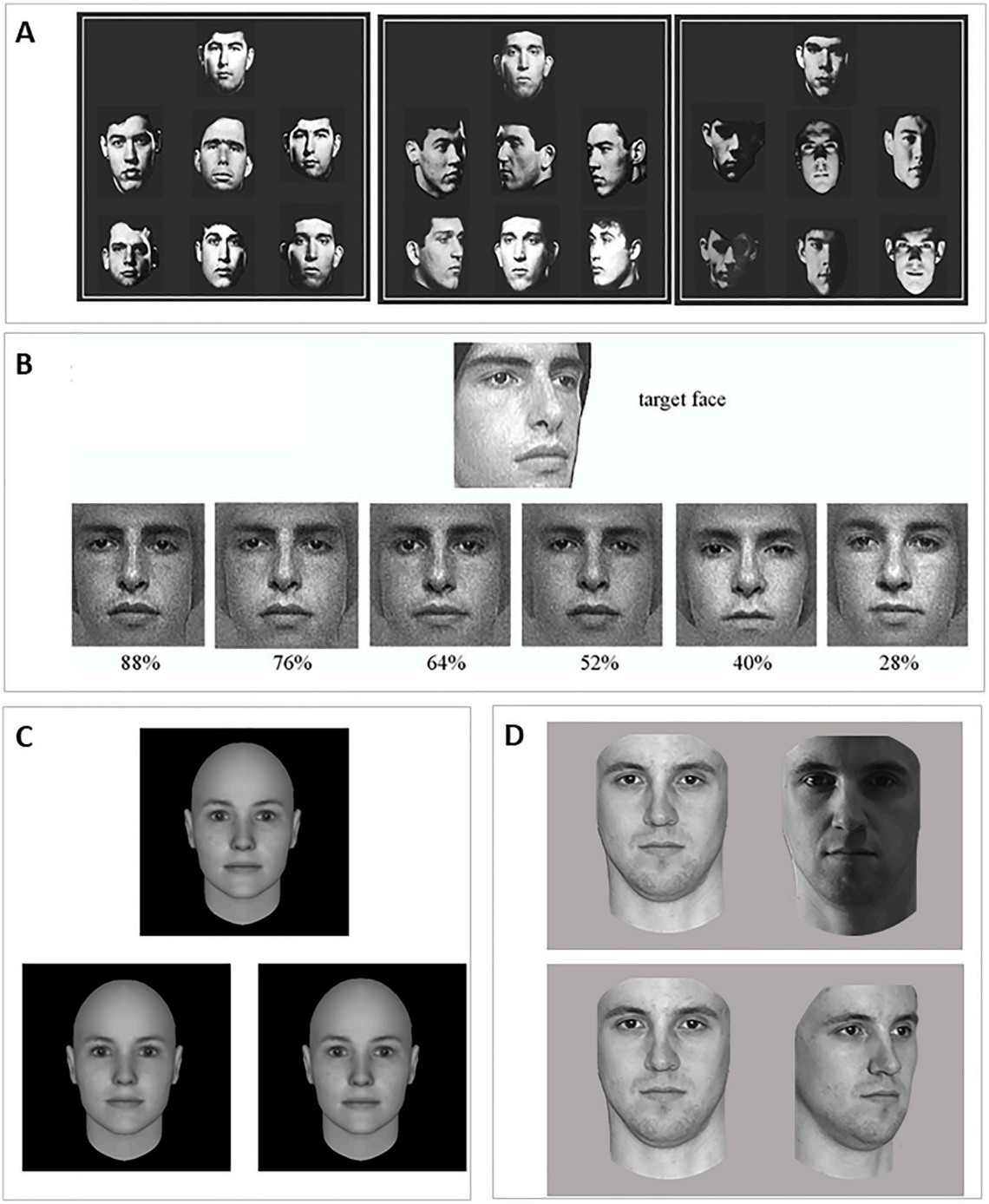


Fig. 1. ***Representative images from the four face matching tasks used in this study.*** A) In the BFRT-*c*, the top image is the target face and the bottom two rows of images are faces from which the target face must be selected. The subpanel shows different trial types, where one out of six front-view faces is the target face (left), or three out of six faces are the same identity as the target face in the view-change trials (middle) or lighting-change trials (right). B) In the CFPT, a 3/4th view target face must be best matched with the six-identity morphed-front view facing faces in descending order of similarity by moving the faces using a mouse. The morph percentages shown below represent the objective similarity to the target face. C) In the USCFPT, the top target face must be matched to either of the two bottom test faces D) The SDFMT displays two faces simultaneously to judge as being same or different, with trials having lighting-change (top) and viewpoint variations (bottom) in faces. In this subpanel, for both the trial types (top and bottom) the responses would be 'same'.

between DPs and controls. Finally, in the tests with lighting and viewpoint change trials (BFRT-c and SDFMT), we also sought to directly compare these trial types in their ability to differentiate DPs and controls and predict objective and subjective face recognition.

2. Materials and methods

2.1. Participants

The study included 60 adults between the ages of 18 and 70 years old. Controls and DPs were recruited from different sources. DPs were

recruited from: a) Our database of Boston DPs who previously participated in laboratory studies, b) referrals from Dr. Matthew Peterson at MIT, who recently completed a Boston-area DP study (Peterson et al., 2019), c) referrals from Dr. Brad Duchaine's website, www.faceblind.org, and d) responses to our advertisement posted on public transportation (Massachusetts Bay Transportation Authority subway system – "T"). Control subjects were recruited from the greater Boston community primarily through flyers and through the Harvard Decision Science Lab in Cambridge, MA. All participants had normal or corrected-to-normal vision, provided informed consent before participating in the study in accordance with the declaration of Helsinki and were compensated for their time. The study was approved by the VA Boston Healthcare System and Harvard Medical School Institutional Review Board Committee, and all study tasks were completed at the Boston VA Medical Center or the Harvard Decision Science Lab.

All participants underwent a pre-visit phone screening to ensure they did not meet any of the following exclusionary criteria: a history of a significant neurological disorder, lifetime moderate to severe traumatic brain injury (TBI) or mild TBI in the last 6 months, musculoskeletal or sensory impairments that would interfere with performing computer tasks, lack of English proficiency, current psychiatric disorders, diagnosed social cognitive disorders such as autism, or current dependence on alcohol or other substances.

2.2. Qualifying as a DP or control participant

To qualify as DP, similar to our previous studies (e.g., Fry et al., 2020) we required individuals to a) report a lifelong history of face recognition difficulties (e.g., not resulting from an event such as a brain injury), b) score >65 on the Prosopagnosia Index questionnaire (PI20; Shah et al., 2015), a self-report measure of prosopagnosia symptoms; c) score lower than 44 out of 72 on Cambridge Face Memory Test (CFMT) and d) score <70% on a Famous Faces test (FFMT). To rule out other causes of poor face recognition, participants had to score normally on a visual acuity/contrast sensitivity test (The Functional Acuity Contrast Test; Ferris et al., 1982), mid-level vision tests (The Leuven Perceptual Organization Screening Test; Torfs et al., 2014), and the Autism Spectrum Quotient questionnaire (<33; Baron-Cohen et al., 2001).

Controls had similar screening criteria to DPs except they could not complain of lifelong face recognition difficulties, score higher than 64 on the PI20, and/or exceed the cutoffs on both FFMT and CFMT. Based on these criteria, 30 DPs and 30 controls were included in our study. The two groups did not differ significantly in either age or gender.

2.3. Procedure: Face Perception Tests

The experiments were designed either in PsychoPy v1.85.4 or Java (for CFPT, CFMT) and run on a laptop (34.5 × 19.5 cm display, 1920 × 1080 pixels, 60 Hz). Participants were seated 60 cm from the computer screen and instructed to indicate their responses using either a keyboard or a computer mouse, based on task demands. The study had four different face matching tests (Fig. 1). Written and spoken instructions were provided. The order of the tests was fixed for all participants: a) USCFPT b) BFRT-*c* c) SDFMT and then d) CFPT. This was done to reduce order-related individual differences effect (Ruiz et al., 2019) and to detect training-related changes more sensitively for the DPs who went on to perform cognitive training. The detailed description of each face perception test is as follows:

2.3.1. Computerized Benton facial recognition test (BFRT-c)

The BFRT-*c* was adapted from Rossion and Michel (2018), and only differs from the 1968 task in that the original instructions are changed to emphasize speed along with accuracy. A series of grayscale photographs of unfamiliar faces (3 × 3.5 cm) are presented with external information cropped out and very little visible hair present (Fig. 1A). In all the trials, a target face is presented at the top of the screen and six faces are simultaneously presented in two rows at the bottom of the screen. Participants are instructed as follows: "You will see a face at the top of the screen that you will have to match to one of the six faces presented below. Click on the matching face. Try to respond as quickly and accurately as possible." The stimuli are displayed until the participant completes their response choices. The test is divided into two parts: in the first part (six trials, front view faces), participants must select one face per trial and in the second part (sixteen trials, eight lighting change and eight head rotation/viewpoint change faces), they must select three faces per trial. There were a total of 54 possible points and data for both accuracy and total task response time was collected. Note that RTs were the total time to complete *all* BFRT trials (i.e., accumulated time to complete entire task) and not just the correct trials as it was not possible to calculate meaningful correct trial RTs from the three response choice trials (Rossion and Michel, 2018).

2.3.2. Cambridge Face Perception Test (CFPT)

The CFPT (Duchaine et al., 2007) is a computerized face sorting task (Fig. 1B), wherein participants arrange the six-upright front-view faces (3 × 4 cm) according to the similarity with the 3/4th profile view of the target face (left-being most similar, right-most dissimilar). The trials are restricted to 1 min per trial. The six faces are generated by morphing the varying proportion of the identity of the target face with six new individual faces. Eight sorts were created, each with upright and inverted face trials that were intermixed in the block. Following Rezlescu et al. (2012) and Rezlescu et al. (2017), we calculated the correct score as (1 % total errors). Chance level is 35.6%.

2.3.3. USC face perception test (USCFPT)

The USCFPT is a face-matching task that uses synthetic grayscale computer-generated faces. The face stimuli were generated using FaceGen software and were generously provided by Irving Biederman (Yue et al., 2012). One core face was used to create eight levels for two dimensions namely, distance between eyes and mouth and height of cheekbones, yielding 64 faces, conceptualized using the Gabor Euclidean distance between the stimuli. The faces are presented on white backgrounds and all faces are devoid of external cues such as hair or clothing (Fig. 1C). Each trial displayed a single target face (3.2 × 4.2 cm) at the top of a screen and two test faces below it for 5 s. The participants were instructed as follows: "In this experiment we are interested in your ability to discriminate between similar images. During each trial, you will identify which of the bottom two images is an exact match of the top image, using the number keys. There will always be an exact match and the location of the target (left or right) will be random. The trials will vary in difficulty, with faces which look very different and those that look very similar to one another. Respond as fast and as accurately as you can". Responses were recorded even after the faces went away. There were total of 96 trials with accuracy as the measure of interest.

2.3.4. Same/different face matching task (SDFMT)

Neutral expression face images from the multi-PIE database (Gross et al., 2010) were 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). Foil faces had a very similar verbal description to target faces. In this task, participants were presented with two face images side-by-side on the screen (4.5 × 6 cm each) for 3 s and had to press 1 or 0 to indicate whether the faces were the same (50% of trials) or different identities (50% trials), respectively. The responses were collected even after 3 s. There was a 1 s inter-trial interval. 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) same identity and same day but cropped differently. 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 showed ~99% accuracy on these trials and they were not included in the subsequent analyses.

2.4. Face recognition tests

In addition to administering face perception tasks, we also included well-validated measures of objective and subjective face recognition. We used the total score from the original Cambridge Face Memory Test (Duchaine and Nakayama, 2006). We also assessed famous face memory test (FFMT)/recognition by using a set of 20 very famous celebrities from testmybrain.org (see Mishra et al., 2019). We used the percent correct out of the total number of people that participants reported being familiar with. Finally, to assess self-reported face recognition difficulties, we used the total score from the Prosopagnosia Index 20 (Shah et al., 2015).

2.5. Statistical analysis approach

All the statistical analyses were conducted on the *z* scores, calculated using the mean (M) and standard deviation (SD) values of this study's control group. Two-tailed independent sample *t*-tests and an $\alpha = 0.5$ were used when needed. Zero-order correlations between the face perception and face recognition tests (CFMT, FFMT) were conducted to establish that the perceptual tasks were all reasonably reliable at detecting prosopagnosia in our sample. Additionally within perceptual tests correlation was also conducted to evaluate the relationship across these tests, separately, in DPs and control sample. Within DP and control groups, the inter-item reliability for each test was assessed using Cronbach's alpha (α) and Guttmann's lambda-2 (λ_2).

Given that BFRT and CFPT have been standard assessments in face perception studies, we first sought to investigate if the performance in BFRT-*c* better explains DP diagnosis and performance in CFMT and FFMT, as standard measures of face memory assessment. To further investigate the role of USCFPT and SDFMT, the regressions were also run with all four tests together. We performed a binary logistic regression to assess which of the above tests predicted unique variance in DP diagnosis or the likelihood that the participants will be categorized as a DP. We used multiple linear regression to predict how these tests relate to face memory performance measures (CFMT and FFMT) and PI20 scores.

We also performed a receiver operating characteristic (ROC) curve analysis (Hajian-Tilaki, 2013) to directly compare among the four face perception tests in terms of sensitivity (proportion of DPs correctly identified), specificity (proportion of controls correctly identified), and area under the curve (AUC; probability that a random DP will score lower than a random control). It should be emphasized that the bimodal distribution of facial recognition ability along with the equal proportion of those with and those without prosopagnosia in our subject sample do not represent the distribution of facial recognition ability for the general population. For this reason, the ROC metrics are to be interpreted only with respect to the relative discrimination performance of the four face perception tests. The metrics from our sample are not meant to predict the absolute discrimination performance of these tests within the general population or within any given subpopulation (e.g., those presenting to a clinic or researcher). If the same ROC metrics were to be derived

Table 1
DPs' raw data and DP and Control group mean scores: Demographics, Face Recognition and Face Perception Performance.

Sub. No.	Gender	Age	PI20 scores	FFMT	CFMT Scores	BFRT-c (/54)	CFPT	SDFMT	USCFPT
1	F	22	88	0.27	34	39	0.71	0.70	0.67
2	F	29	88	0.35	37	42	0.51	0.71	0.84
3	F	34	75	0.33	39	37	0.65	0.81	0.84
4	M	61	89	0.29	38	36	0.54	0.65	0.69
5	X*	36	93	0.54	35	35	0.75	0.68	0.64
6	M	33	80	0.53	36	33	0.47	0.71	0.51
7	M	27	80	0.19	38	47	0.78	0.83	0.84
8	F	46	75	0.39	34	39	0.56	0.74	0.65
9	F	53	86	0.40	35	39	0.72	0.77	0.74
10	F	26	80	0.47	42	42	0.76	0.79	0.72
11	F	35	81	0.45	43	47	0.69	0.78	0.85
12	F	30	69	0.42	41	42	0.63	0.76	0.74
13	F	32	58	0.55	40	38	0.58	0.73	0.78
14	F	27	86	0.29	44	39	0.72	0.80	0.84
15	F	63	63	0.20	37	38	0.28	0.73	0.72
16	F	31	89	0.10	37	42	0.71	0.73	0.71
17	F	55	96	0.47	33	34	0.57	0.69	0.85
18	F	39	78	0.47	33	42	0.64	0.80	0.72
19	F	28	80	0.27	42	47	0.60	0.82	0.84
20	M	37	91	0.35	33	40	0.65	0.74	0.67
21	F	28	80	0.47	39	42	0.68	0.84	0.80
22	F	64	85	0.25	43	43	0.60	0.79	0.86
23	F	52	87	0.12	38	41	0.40	0.68	0.49
24	F	25	88	0.26	44	42	0.68	0.79	0.80
25	M	50	82	0.20	44	39	0.64	0.76	0.82
26	F	33	89	0.35	39	39	0.68	0.73	0.69
27	F	39	71	0.50	41	36	0.79	0.74	0.80
28	F	23	92	0.37	32	35	0.60	0.76	0.67
29	F	70	83	0.13	36	39	0.68	0.73	0.66
30	F	27	76	0.08	42	38	0.51	0.72	0.83
DP group (n = 30)	24F	38.5 ± 13.69	81.93 ± 8.74	0.34 ± 0.14	38.3 ± 3.72	39.73 ± 3.61	0.63 ± 0.11	0.75 ± 0.05	0.74 ± 0.10
Control (n = 30)	18F	38.83 ± 10.18	35.86 ± 7.95	0.77 ± 0.17	59.4 ± 7.86	45.33 ± 4.25	0.74 ± 0.10	0.80 ± 0.07	0.79 ± 0.091

Note. Accuracy for FFMT: Famous Faces Test, SDFMT: Same/different face matching test and USCFPT: University of Southern California face perception test; Raw scores for BFRT-c: Computerized Benton face recognition test and CFMT: Cambridge Face Memory Test; CFPT: Cambridge face perception test report upright deviation scores. *This participant's gender identity is nonbinary. Summary data for DPs and controls are represented by the mean ± standard deviation.

in a random sample of the general population, the absolute performance would differ from that of our sample (see Arizpe et al., 2019).

2.6. Sample size justification

Our sample size was guided by previous studies comparing face perception between DPs and controls (White et al., 2017) and studies of individual differences in face recognition (DeGutis et al., 2013; Richler et al., 2011). White et al. (2017) found significant DP and control differences in face matching when using a sample of 21 in-lab controls and 6 DPs. We wanted to include additional DPs in the current study to improve our sensitivity to detect differences between face perception tests.

3. Results

3.1. Demographics and diagnostic test performance

Participants included 30 DPs (24 female) and 30 controls (18 female) with a mean age of 38.5 years ($SD = 13.69$) and 38.8 years ($SD = 10.18$), respectively. There was no significant difference between the two groups in age ($p = .92$), but there was a trend toward a higher proportion of females in the DP group ($p = .09$). As expected, the DP group performed significantly worse (all $p < .001$) than controls on objective and subjective diagnostic measures of face recognition (CFMT, FFMT, and PI20; Table 1). We also verified that our control group performed similarly to previous normative samples (CFMT: control $M/SD = 59.4/7.9$, normative sample $M/SD = 57.9/7.9$, Duchaine and Nakayama, 2006; PI20: control $M/SD = 35.86/7.95$, normative sample $M/SD = 38.90/10.88$, Shah et al., 2015; CFPT: control $M/SD = 0.74/0.10$, normative sample $M/SD = 0.72/0.10$, Rezlescu et al., 2017; BFRT-*c*: control $M/SD = 45.33/4.25$, normative sample $M/SD = 44.81/3.44$, Rossion and Michel, 2018; USCFPT: control $M/SD = 0.79/0.09$, normative sample $M/SD = 0.83/0.08$; Irving Biederman, personal communication).

Table 2
Internal reliability of the perceptual tests for DP and control groups.

Reliability	BFRT-c		CFPT		SDFMT		USCFPT	
	α	λ_2	α	λ_2	α	λ_2	α	λ_2
DP	0.53	0.60	0.79	0.82	0.71	0.76	0.82	0.85
Control	0.74	0.77	0.74	0.76	0.88	0.89	0.81	0.84
Normative data	0.61	–	–	0.68	–	–	–	–

Note: Normative data for BFRT-c and CFPT were obtained from Rossion and Michel (2018) (Guttman's lambda; Rezlescu et al., 2017). No normative data were available for USCFPT and SDFMT. α = Cronbach's alpha, λ_2 = Guttman's lambda 2.

3.2. Reliabilities of, and correlations between, perceptual measures

To help with interpretation of correlation and group differences and provide a sense of whether these tests are tapping into similar constructs, we first measured the reliability of each face perception test and how strongly they inter-correlated using *Cronbach's* α and *Guttman's* λ_2 (Table 2). DPs showed reduced internal consistency than controls on the BFRT-*c* and SDFMT while the reliabilities were similar for the other two tests. The values were also comparable to previously established normative samples. Results of Pearson correlations between the four face perception tests combined for DP and control group as well as independently for the two groups showed medium to large positive correlations ($0.58 \leq r\ (58) \geq 0.37$, $p < .001$), suggesting that these tests measure similar perceptual abilities (Fig. 2).

3.3. The ability of face matching tests to detect face perception deficits in DPs and predict face recognition ability

We next examined each test's individual ability to detect face perception performance differences in DPs vs. controls and in their correlations with face recognition ability. Table 1 displays the raw scores for DPs and mean values of DPs and control groups across the four perceptual tests. Overall, there were significant differences between the groups across different perceptual tests, with medium to large effect

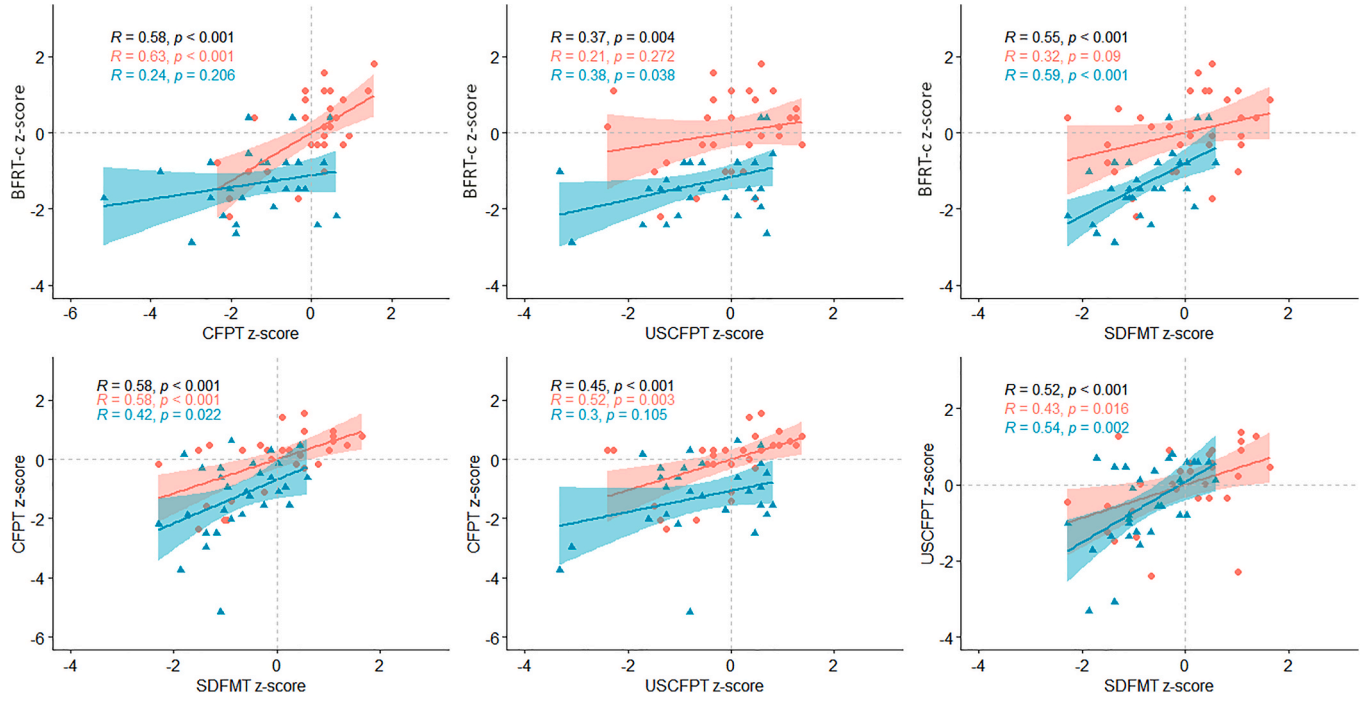


Fig. 2. ***Correlation between Face Perception Tests.*** Within test correlations for DPs (blue) and Controls (red) group across the four perceptual tests. The values in black represent overall group coefficient value. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

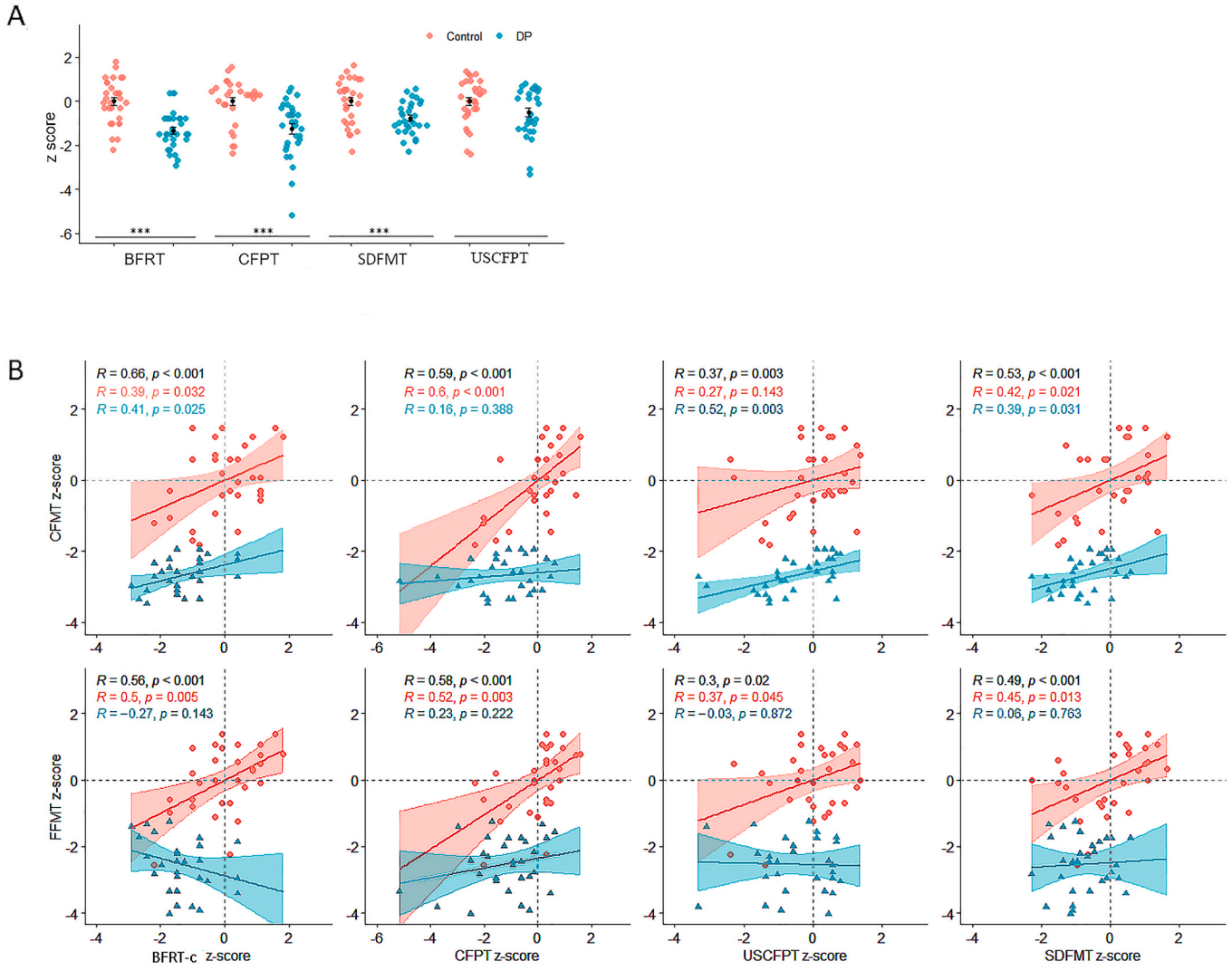


Fig. 3. ***Performance in Four Face Perception tests*** A) z scores of individual participants' data- Difference between DP and Control. The black dot represents the mean and the error bars represent standard error of mean (SEM). The width for each group represents the frequency of the repeated datapoint while the height represents the variability across the group. B) Zero order correlations -for CFMT and FFMT across four different face matching tests used in the study. *Note.* *p < .05, **p < .01, ***p < .001, 95% confidence interval in brackets.

sizes (Table 1, Fig. 3A). The BFRT-*c* (t (58) = 5.5, $p < .001$, Cohen's d = 1.42) most robustly differentiated between DPs and controls, followed by the CFPT (t (58) = 4.19, $p < .001$, Cohen's d = 1.08), and SDFMT (t (58) = 3.34, p = .001, Cohen's d = 0.86). The USCFPT scores only trended towards showing a DP/control difference (t (58) = 1.92, p = .06, Cohen's d = 0.5).

Next, we examined the associations between face perception tests and objective measures of face recognition ability for DP and control groups independently as well as collapsed across groups (Fig. 3B). We found that the BFRT-*c* and CFPT had significantly higher correlations with face recognition memory scores, followed by SDFMT and then USCFPT scores.

With respect to the BFRT-*c*, surprisingly, we did not find that DPs (M = 243.7 s, SD = 79.78 s) had significantly slower response times (RTs, time to complete the entire task) than the control group (M = 213.67 s, SD = 75.29 s, p = .14). Notably, our older control group (M_{age} = 38 yrs) had longer RTs than a recent BFRT-*c* study with a younger control group (e.g., Rossion and Michel, 2018; M = 180.85 s; SD = 59.87 s, M_{age} = 22 yrs), suggesting that age may be a factor in response times. To further address if RTs had a significant association with DP vs. Control group membership (given that there is an average of 30 s difference between the means of the two samples), we correlated the accuracy and RTs for the groups.

As can be seen in Fig. 4, we found weak, non-significant correlations between RT and accuracy for BFRT-c for all trials (controls: r = −.21; DPs: r = −0.01; All participants: r = −0.20, [0.97 < all p's > 0.1]) and similar non-significant accuracy/RT correlations when examining the last 48 trials of the BFRT-c (controls: r = 0.16; DPs: r = 0.09; All participants: r = 0.02, [0.9 < all p's > 0.4]). We also examined combination RT-accuracy measures. Given the weak correlation between RT and accuracy, using the more widely used Inverse Efficiency Score (IES) to integrate RTs and accuracy is less valid (see Bruyer and Brysbaert, 2011), so we also calculated the Balanced Integrated Score (BIS; Liesefeld and Janczyk, 2019; Liesefeld et al., 2015) using subtraction between the z transformed values of RT and Accuracy. Here, we calculated RT and accuracy z scores for all subjects using the Mean and SD of the control group. In the current study, neither the raw RTs (Odds ratio = 0.995), the IES (Odds ratio = 0.99; p = .15), nor BIS (Odds ratio = 2.4; p < .001) explained unique variance on their own in logistic and linear regressions. Finally, in our current analysis, just one DP (Mean = 366 s) was 2 SD slower than the control group mean (Mean = 213 s, SD = 75 s). Together, these results suggest that BFRT-*c* RTs were not useful in diagnosing DPs and did not enhance the performance of BFRT-*c* accuracy.

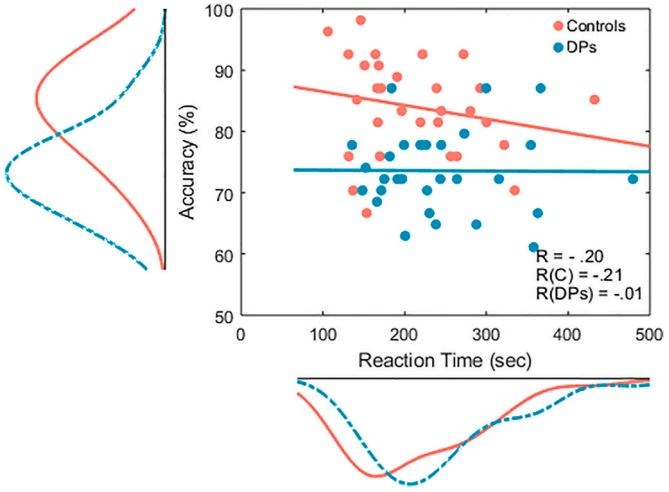


Fig. 4. ***RT vs. Accuracy correlation*****:** The figure shows the zero-order correlation for all participants between RT and Accuracy of BFRT-c all trials. R(C) = Control Group, R (DPs) = developmental prosopagnosia group.

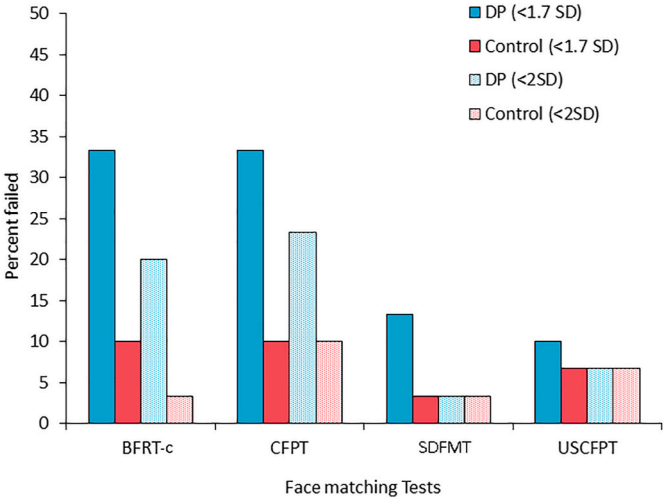


Fig. 5. ***Participants with Perceptual Deficits in Each Test***. The figure shows percent of people who failed each of the four perceptual tests, for both the groups at < 1.7 SD and <2 SD values.

Complementing these analyses, we compared each test's sensitivity to perceptual impairment at the individual participant level (<2SD, < 1.7 SD, Geskin and Behrmann, 2018). As can be seen in Fig. 5, we found that the BFRT-*c* and CFPT were very similar in their ability to detect face perception impairments, while the USCFPT and SDFMT were considerably less able to detect impaired performance.

To provide a more fine-grained understanding of how well each face perception measure predicted DP vs. control group membership, we next performed ROC analyses (Fig. 6) and measured the sensitivity, specificity, Youden Index (J = sensitivity + specificity −1), and area under the curve (AUC). The sensitivity and specificity values reported here are for the optimal (maximum) point on the curve. Here, we used the four face perception tests and plotted the ROC curve to calculate the AUC for each test (Fig. 6, AUC ≥0.9 is excellent, ≥ 0.8 is very good and ≥0.6 is fair). The analysis showed that both BFRT-*c* (AUC = 0.83, p < .001, sensitivity = 0.900, specificity = 0.733, J = 0.633) and CFPT (AUC = 0.80, p < .001, sensitivity = 0.800, specificity = 0.767, J = 0.567) were best at discriminating DP from the control group while SDFMT test results were moderate (AUC = 0.73, p = .002, sensitivity = 0.867, specificity = 0.567, J = 0.433) and the USCFPT (AUC = 0.64, p = .059, sensitivity = 0.533, specificity = 0.800, J = 0.333) had the weakest performance.

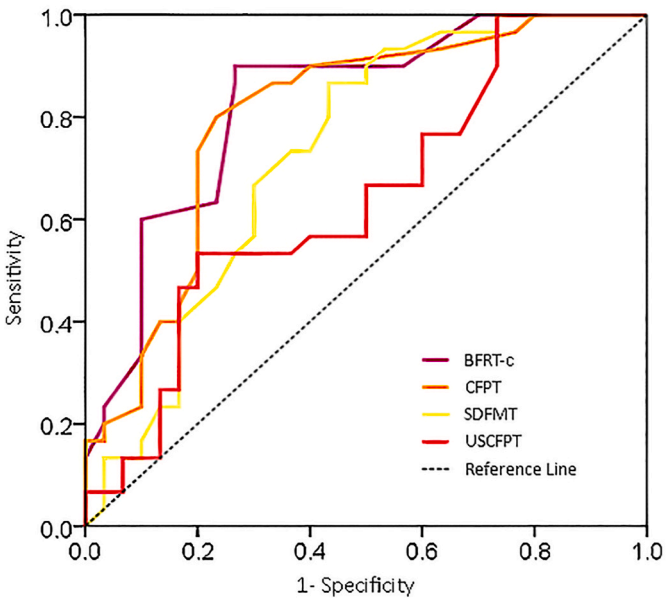


Fig. 6. ***ROC curve for the face perceptual tests***. The area under the curve (AUC) for four perceptual tests shows that BFRT-c and CFPT perform better than other tests.

3.4. How do the BFRT-c and CFPT compare in their ability to predict DP diagnosis and face recognition ability?

We next sought to directly compare the BFRT-*c* and CFPT in predicting DP vs. control group membership and face recognition ability. A logistic regression including both tests showed an overall significant model fit (χ^2 (57) = 26.87, p < .001, AIC for model fit = 62.31). Notably, the BFRT-*c* significantly predicted unique variance in DP diagnosis (p = .004, Odds Ratio = 3.19) while the CFPT did not predict unique variance (p = .12, Odds Ratio = 1.74). When using both measures to predict CFMT scores, both the BFRT-*c* (t = 4.23, p < .001) and CFPT (t = 2.71, p = .009) predicted unique variance in CFMT performance and combined predicted 50.3% of the variance (F (2, 57) = 28.79, p < .001). Similarly, the overall model was significant for predicting FFMT scores (F (2, 57) = 19.99, p < .001, R^2 = 0.42), with both BFRT-*c* (t = 2.78, p = .007) and CFPT (t = 3.04, p = .004) as significant independent predictors in the model. The model was also significant for predicting PI20 scores (F (2, 57) = 15.16, p < .001, R^2 = 0.35), with only BFRT-*c* (t = 3.50, p < .003) as a significant independent predictor in the model (CFPT, t = 1.4, p = .16). Together, this suggests that the BFRT-*c* outperformed the CFPT when predicting DP diagnosis and self-reported face recognition and the two tests performed similarly when predicting objective face recognition measures.

3.5. How do the face matching assessments compare in their ability to predict DP diagnosis and face recognition ability?

We next sought to perform exploratory analyses comparing the ability of all four tests to predict DP vs. control membership and face recognition. To achieve this, we first performed a logistic regression with the four perceptual tests predicting DP vs. control group membership. Results of the logistic regression indicated an overall significant relationship (χ^2 (55) = 27.20, p < .001, AIC for model fit = 65.97) between the perceptual tests (BFRT-*c*, CFPT, SDFMT, USCFPT) and the binary outcome variable (categorized as being a control/DP). However, only BFRT-*c* (W = 7.41, p = .006) predicted unique variance in the

diagnosis of DP.

We next compared the face perception tests in their ability to predict more fine-grained, continuous measures of objective (CFMT/FFMT) and subjective face recognition (PI20). Results of the multiple linear regression indicated that together, the four perceptual tests significantly predicted CFMT scores, (F (4, 55) = 14.70, $p < .001$, $R^2 = 0.52$). The BFRT-*c* ($t = 3.58$, $p < .001$) was the only significant unique predictor for CFMT performance with marginal independent variance explained by the CFPT ($t = 1.93$, $p = .059$). Similarly, the overall model was significant for predicting FFMT scores (F (4, 55) = 10.13, $p < .001$, $R^2 = 0.42$), with BFRT-*c* ($t = 2.29$, $p = .026$) and CFPT ($t = 2.41$, $p = .019$) as significant independent predictors in the model. The model was also significant for predicting PI20 scores (F (4, 55) = 7.46, $p < .001$, $R^2 = 0.35$), with only BFRT-*c* ($t = -3.09$, $p = .003$) as a significant independent predictor in the model. Notably, the USCFPT and SDFMT did not uniquely predict any of the continuous objective or subjective face recognition measures.

3.6. Comparing changes in lighting vs. viewpoint in discriminating between DPs and controls

We finally sought to determine whether lighting change or viewpoint change trials better discriminated between DPs and controls by examining separate trials of the BFRT-*c* and SDFMT tests (Fig. 7A). There were significant differences in performance between the control and DP group in the BFRT-*c* viewpoint change trials (t (58) = 4.66, $p < .001$, Cohen's $d = 1.2$) and BFRT-*c* lighting change trials (t (58) = 5.23, $p < .001$, Cohen's $d = 1.35$) conditions as well as the SDFMT viewpoint (t (58) = 2.67, $p = .010$, Cohen's $d = 0.69$) and SDFMT lighting change (t (58) = 3.23, $p = .002$, Cohen's $d = 0.84$) conditions.

Given that the number of trials is reduced when dividing into lighting and viewpoint change conditions, we averaged the accuracy for SDFMT

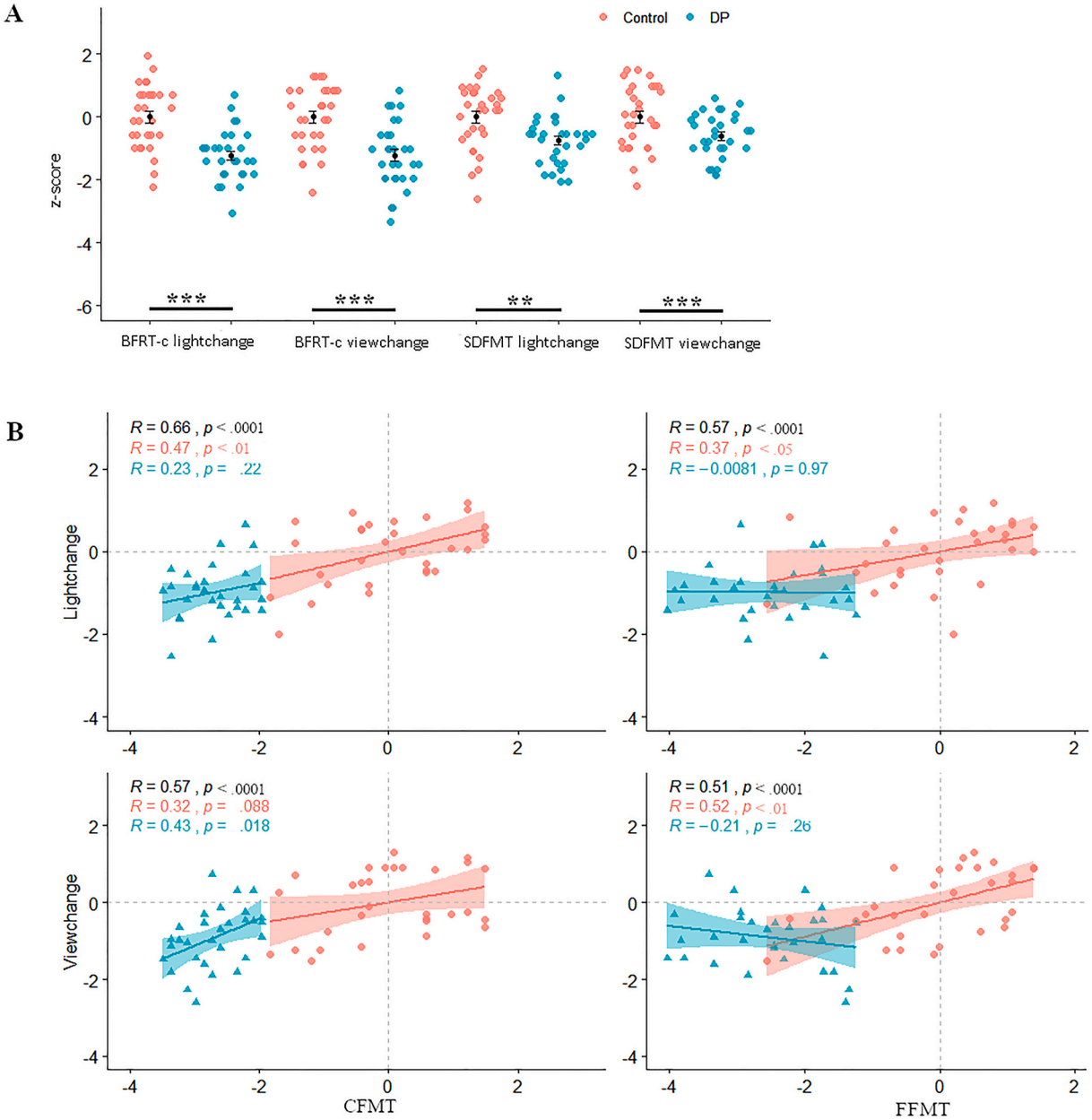


Fig. 7. ***Lighting and viewpoint change conditions.*** A) *z* scores of individual participants' data: Difference between DP and Control group. The black dot represents the mean and the error bars represent standard error of mean (SEM). B) Zero order correlations for lighting and viewpoint change across memory tests (CFMT and FFMT).

and BFRT-*c* for each condition. We then performed a logistic regression with the composite lighting change and viewpoint change trials as predictors. The model was significant (χ^2 (57) = 26.15, $p < .001$, AIC for model fit = 62.98), with only lighting change trials as a significant independent predictor for the diagnosis of DP (W = 7.83, $p = .005$). The ROC analysis in this case revealed that the combined lighting change trials (AUC = 0.84, $p < .001$) were the best predictor of all the measures, followed by all trials of the BFRT-*c* (AUC = 0.83, $p < .001$) and CFPT (AUC = 0.80, $p < .001$).

We further examined linear regressions with composite light and viewpoint change trials as predictors for objective and subjective face recognition scores (Fig. 7B). Overall regressions were significant for all three face recognition measures: CFMT (F (2, 57) = 26.71, $p < .001$, R^2 = 0.48), FFMT (F (2, 57) = 16.80, $p < .001$, R^2 = 0.37) and PI-20 (F (2, 57) = 15.7, $p < .001$, R^2 = 0.36). Examination of individual predictors showed that lighting change trials significantly predicted unique variance in the CFMT ($t = 4.13$, $p < .001$), FFMT ($t = 3.19$, $p = .002$) as well PI-20 ($t = -3.1$, $p = .004$), whereas viewpoint change trials only significantly predicted unique variance in the CFMT ($t = 2.39$, $p = .02$) and trended towards predicting unique variance in the FFMT ($t = 2.0$, $p = .051$) and PI-20 ($t = -1.97$, $p = .054$).

4. Discussion

The current study helps to address the existing gap in the literature regarding which tests are best for identifying face perception impairments. When diagnosing DP as well as predicting objective and subjective face recognition, the BFRT-*c* and CFPT consistently showed robust predictive abilities and clearly outperformed the other measures. When we directly compared the BFRT-*c* and CFPT, the BFRT-*c* outperformed the CFPT, solely predicting unique variance in DP diagnosis and self-reported face recognition. The two tests both predicted unique variance in objective face recognition measures (CFMT, FFMT). This suggests that the BFRT-*c* is as good as, if not better than, the widely used CFPT in measuring face perception deficits and provides complementary information to the CFPT. In exploratory analyses, we also found that accuracy for making perceptual judgments of faces across lighting changes better predicted DP vs. control membership as well as subjective and objective face recognition compared to viewpoint changes. These results have both important clinical and theoretical implications in assessing populations with face processing difficulties.

The results from our study demonstrate that the BFRT-*c* is a sensitive and reliable measure for assessing face perception impairments in DPs, contrasting previous studies which used the original version of the BFRT (Albonico et al., 2017; Duchaine and Nakayama, 2004; Duchaine and Weidenfeld, 2003). Using the original BFRT, previous studies found deficient performance, *z*-score $\leq -.97$, in only 2 out of 11 (Duchaine and Nakayama, 2004) and 5 out of 23 DPs (Albonico et al., 2017), whereas 19 out of 30 DPs in our study showed a *z*-score ≤ -1. Further, 10 out of 30 DPs in the current study had *z*-scores < -1.7 and 6 of 30 had *z*-scores < -2. One likely reason why the current BFRT-*c* results differ from results with the original BFRT is that the BFRT-*c* version included speeded task instructions, which led both DPs and controls to make faster judgments. Unconstrained stimuli presentation times could lead to using image-based matching (Duchaine and Nakayama, 2004; Nunn et al., 2001) or more detailed serial processing of facial features (Behrmann et al., 2005; Stollhoff et al., 2010) like using eyebrows and hairline to match the faces (Duchaine and Weidenfeld, 2003). Providing speeded instructions in the current study likely reduced DPs' (as well as controls') usage of feature-based strategic approaches (Delvenne et al., 2004; Duchaine and Nakayama, 2005; Nunn et al., 2001). Indeed, previous studies with DPs and other clinical groups have observed only mild accuracy deficits but extremely slow performance in original version of the BFRT (Benton and Van Allen, 1972; Busigny et al., 2009; De Renzi et al., 1991; Geskin and Behrmann, 2018; Schretlen et al., 2001) whereas other studies using the original BFRT failed to report the time taken to perform the task (Albonico et al., 2017; Duchaine and Nakayama, 2004; Duchaine and Weidenfeld, 2003). In contrast to those studies, using the BFRT-*c* we did not find significant RT differences between DPs and controls. Together, this suggests that the speeded instructions of the BFRT-*c* (Dzhelyova et al., 2020; Rossion and Michel, 2018) resulted in faster face matching judgments and made the BFRT-*c* a better face perception assessment than the original.

We found that the BFRT-*c* performed as well as, if not better than, the widely used CFPT, particularly when diagnosing DP and predicting subjective face recognition, and both measures were much better than the USCFPT and SDFMT tests at differentiating between DP and control groups. One explanation for the better performance of the BFRT-*c* is that it requires identity-based matching while the CFPT and other tests might rely more on image-based matching. The BFRT-*c* exploits the natural variability in face identity by manipulating lighting and viewpoint, which are both considered to be necessary elements in face learning (Burton, 2013; Jenkins et al., 2018; Kramer et al., 2017; Ritchie and Burton, 2017). This is much less evident in the CFPT that use morphs and computer-generated faces, making it more of a similarity judgement rather than an identity matching/discrimination task (Rossion and Michel, 2018). It is possible that after selecting the first match face in CFPT, participants use this front-view image to judge image-specific similarity to order the rest of the front-view faces. Further, some studies have suggested that the selecting and arranging of faces on the CFPT requires additional computer literacy than typical face-matching tests and may involve higher-level executive processes (Biotti et al., 2019; Huis in 't Veld et al., 2012; Logan et al., 2016; White et al., 2017). These strategic responses may lead to greater variability in performance as seen in our study as well as reported previously (Garrido et al., 2008) and may make it a particularly poor test for older adults (Bowles et al., 2009) and those experiencing cognitive decline.

In the case of the USCFPT's poor performance, one potential factor is that the faces were generated using two dimensions namely, distance between eyes and mouth and height of cheekbones. This may have allowed DPs to be accurate using a simple feature-based strategy (e.g., judging the size of the eyes). Apart from this, the synthetic faces used in USCFPT were front-view and lacked natural variability (changes in hair, skin texture, light, and viewpoint changes), enabling DPs to potentially use more of an image-matching strategy. Critically, the faces used in the USCFPT were presented for a long duration (5 s), which may have made feature-by-feature comparisons an available strategy. These factors likely combined to make the USCFPT far less ideal for assessing face perception deficits in DPs. Furthermore, though the SDFMT task exploits changes in naturalistic face images, it was similarly not a particularly sensitive test to measure DPs' impairments. Two important differences between the BFRT-*c* and SDFMT are that the former uses much more drastic lighting and viewpoint changes and, further, the BFRT-*c* is not affected by potential response biases that could be present in same/different matching tasks like the SDFMT.

Besides highlighting the strengths of the BFRT-*c*, our results demonstrate that, compared to *viewpoint* changes, *lighting* changes not only better classify DP vs. control group membership, but also significantly predict performance on objective (CFMT) and subjective (PI20) face recognition tests. This is consistent with anecdotal observations of DPs' difficulty with face matching across lighting changes (e.g., Duchaine and Nakayama, 2004) and, to our knowledge, this is the first study to demonstrate this at the DP group level. Our findings are in line with previous studies in healthy controls reporting that lighting changes dramatically impact face matching performance (Adini et al., 1997; Braje, 2003; S. Favelle et al., 2017; Rossion and Michel, 2018) and significantly reduce performance in short-term and long-term memory for faces (Braje et al., 1998). Moreover, Liu et al. (2009) showed it is more difficult to generalize/transfer face identities learnt through variations in lighting changes than viewpoint changes. One explanation is that lighting changes lead to significant changes in face shape, surface texture, and/or 3D appearance of faces, impacting both configural and

featural information (Favelle et al., 2011, 2017). Thus, lighting changes may better bring out DPs' impairments than viewpoint changes because comparing faces across lighting changes taxes multiple face-selective processes, being able to identify DPs with either deficits in featural (e.g., eyes, Fisher et al., 2016) or holistic processing (DeGutis et al., 2012) or some milder combination of both. Future studies would be useful to investigate the specific contributions of these deficient processes to DPs' lighting change trial impairments Finally, evidence suggests that the well-studied acquired prosopagnosic PS, unlike DPs, is more impaired with viewpoint changes than lighting changes in the BFRT (Rossion and Michel, 2018). It would be interesting to determine if DPs and APs differ in their relative lighting vs. viewpoint deficits.

The current results have important implications for DP. First, our results confirm that DPs have clear, group-level perceptual deficits when using sensitive and reliable measures, contrasting with studies that have suggested DPs may have limited perceptual deficits (e.g., Ulrich et al., 2017). Our results are also consistent with the assertion that face recognition deficits in some DPs can be primarily driven by face perception impairments, i.e., apperceptive prosopagnosia (Biotti et al., 2019), whereas for others it may be due to more associative mechanisms. Though this study helps in characterizing perceptual deficits, to better differentiate between apperceptive vs. associative prosopagnosia, it would be useful to develop a better understanding of what specific associative mechanisms are impaired and how to measure them rather than simply classifying associative prosopagnosics as the 'absence of perceptual impairments' (Biotti et al., 2019). To this end, Stumps et al. (2020) recently demonstrated that DPs have impairments in face recollection, the all-or-none recognition of a face with semantic and contextual details, while showing intact familiarity, or feeling of knowing. Importantly, face recollection and face perception deficits uniquely predicted DP vs. control group membership and continuous measures of face recognition ability, suggesting that face perception and face memory deficits represent distinct contributions to DP (for an alternative viewpoint that all APs have some level of face perceptual deficits, see Busigny et al., 2014).

Beyond DPs, the current results have important implications for the broader area of characterizing face perception impairments. Previous studies attempting to identify face perception impairments have either used measures of unknown reliability and validity or have only used a single measure (e.g., Biotti et al., 2019). According to DSM-5 criteria, identifying deficient cognitive processes in individuals requires multiple impaired measures (Sachdev et al., 2014). The current results suggest that the BFRT-*c* and CFPT are two highly valid and sensitive face perception measures that predict unique variance in objective measures of face recognition. Combined, they can help determine whether face perception deficits play a role in impaired face recognition ability across a broad spectrum of psychiatric and neurological populations and can help make important treatment recommendations. Finally, the current results suggest that as researchers and clinicians develop novel face perception tests to detect impairments, close consideration should be paid to applying time pressure and incorporating lighting changes across faces.

The results of the current study are compelling but have some limitations. First, though we compared the BFRT-*c* to the most widely used assessment to measure face perception deficits in DPs, the CFPT, it would be useful to perform future studies comparing the BFRT-*c* to other face matching tests that use real faces, include lighting/viewpoint changes (or other variations), and that avoid response biases that are common in same/different matching tasks. Second, though our controls showed comparable performance to larger normative samples, it was relatively small. Finally, though DPs are an excellent test group to assess the sensitivity of face perception assessments, replicating the current study in additional populations (e.g., Alzheimer's, autism, age-related cognitive decline) would be useful to assess the generalizability of the BFRT-*c* and CFPT in assessing face perception deficits.

5. Conclusion

Research has shown that face perception and face memory are interconnected but distinct processes. Yet, in contrast to face memory assessments that are better-established, studies have yet to examine the best assessments to reliably measure face perception impairments. Encoding faces accurately is an initial step in developing face familiarity and later stages of face memory, person recognition, and integration with subsequent cognitive processes. The current study examines multiple face perception tests and demonstrates that the BFRT-*c* and, to a lesser extent, the CFPT are valid and sensitive face perception tests and combined can be used to effectively assess face perception impairments.

Credit author statement

Maruti Mishra: Conceptualization, Data Curation, Formal analysis, Visualization, Writing - Original Draft, Writing - Review & Editing. Regan Fry: Investigation, Data curation, Writing- Review & Editing. Elyana Saad: Methodology, Software. Joseph Arizpe: Methodology, Software, Writing- Review & Editing. Yuri-Grace Ohashi: Investigation. Joseph M DeGutis: Conceptualization, Methodology, Resources, Writing - Original Draft, Writing - Review & Editing, Supervision, Project administration, Funding acquisition.

Declaration of competing interest

We have no known conflicts of interest to disclose.

Acknowledgement

We thank Alice E. Lee and volunteers at HDSL for data collection of DP and control participants, respectively. We want to thank developmental prosopagnosics and control participants for completing our challenging battery of tasks.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.neuropsychologia.2021.108067.

Funding

This study was funded by R01 from the National Eye Institute grant #R01EY026057 awarded to JD. The contents of this manuscript do not represent the views of the U. S. Department of Veterans Affairs or the United States Government.

Data availability

The raw data and materials for all experiments are available upon request to the corresponding author.

References

Adini, Y., Moses, Y., Ullman, S., 1997. Face recognition: the problem of compensating for changes in illumination direction. IEEE Trans. Pattern Anal. Mach. Intell. 19 (7), 721–732. https://doi.org/10.1109/34.598229.

Albonico, A., Barton, J., 2019. Progress in Perceptual Research: the Case of Prosopagnosia, F1000Research, p. 8. https://doi.org/10.12688/f1000research.18492.1.

Albonico, A., Malaspina, M., Daini, R., 2017. Italian normative data and validation of two neuropsychological tests of face recognition: Benton Facial Recognition Test and Cambridge Face Memory Test. Neurol. Sci. 38 (9), 1637–1643. https://doi.org/10.1007/s10072-017-3030-6.

Arizpe, J.M., Saad, E., Douglas, A.O., Germine, L., Wilmer, J.B., DeGutis, J.M., 2019. Self-reported face recognition is highly valid, but alone is not highly discriminative of prosopagnosia-level performance on objective assessments. Behav. Res. Methods 51 (3), 1102–1116. https://doi.org/10.3758/s13428-018-01195-w.

Baron-Cohen, S., Wheelwright, S., Skinner, R., Martin, J., Clubley, E., 2001. The autism-spectrum quotient (AQ): evidence from asperger syndrome/high-functioning autism, malesand females, scientists and mathematicians. J. Autism Dev. Disord. 31 (1), 5–17. https://doi.org/10.1023/A:1005653411471.

Barton, J.J.S., Cherkasova, M.V., Hefter, R., Cox, T.A., O'Connor, M., Manoach, D.S., 2004. Are patients with social developmental disorders prosopagnosic? Perceptual heterogeneity in the Asperger and socio-emotional processing disorders. Brain 127 (8), 1706–1716. https://doi.org/10.1093/brain/awh194.

Bate, S., Bennetts, R.J., 2014. The rehabilitation of face recognition impairments: a critical review and future directions. Front. Hum. Neurosci. 8 https://doi.org/10.3389/fnhum.2014.00491.

Behrmann, M., Avidan, G., Marotta, J.J., Kimchi, R., 2005. Detailed exploration of face-related processing in congenital prosopagnosia: 1. Behavioral findings. J. Cognit. Neurosci. 17 (7), 1130–1149. https://doi.org/10.1162/0898929054475154.

Benton, A.L., Van Allen, M.W., 1968. Impairment in facial recognition in patients with cerebral disease. Cortex 4 (4), 344. https://doi.org/10.1016/S0010-9452(68)80018-8. IN1.

Benton, A.L., Van Allen, M.W., 1972. Prosopagnosia and facial discrimination. J. Neurol. Sci. 15 (2), 167–172. https://doi.org/10.1016/0022-510X(72)90004-4.

Benton, A.L., Sivan, A.B., Hamsher, K.D., Varney, N.R., Spreen, O., 1994. Contributions to Neuropsychological Assessment: A Clinical Manual. Oxford University Press, USA.

Biederman, I., Margalit, E., Maarek, R.S., Meschke, E.X., Shilowich, B.E., Juarez, J.J., Hacker, C.M., Seamans, T.J., Herald, S.B., 2017. What is the nature of the perceptual deficit in congenital prosopagnosia. Proceedings of the Annual Meeting of Vision Sciences Society.

Biotti, F., Gray, K.L.H., Cook, R., 2019. Is developmental prosopagnosia best characterized 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. Q. J. Exp. Psychol. 70 (2), 201–217. https://doi.org/10.1080/17470218.2016.1161059.

Boutet, I., Taler, V., Collin, C.A., 2015. On the particular vulnerability of face recognition to aging: a review of three hypotheses. Front. Psychol. 6, 1139. https://doi.org/10.3389/fpsyg.2015.01139.

Bowles, D.C., McKone, E., Dawel, A., Duchaine, B., Palermo, R., Schmalzl, L., Rivolta, D., Wilson, C.E., Yovel, G., 2009. Diagnosing prosopagnosia: effects of ageing, sex, and participant-stimulus ethnic match on the Cambridge face memory test and Cambridge face perception test. Cogn. Neuropsychol. 26 (5), 423–455. https://doi.org/10.1080/02643290903343149.

Braje, W.L., 2003. Illumination encoding in face recognition: effect of position shift. J. Vis. 3 (2) https://doi.org/10.1167/3.2.4, 4–4.

Braje, W.L., Kersten, D., Tarr, M.J., Troje, N.F., 1998. Illumination effects in face recognition. Psychobiology 26 (4), 371–380. https://doi.org/10.3758/BF03330623.

Bruce, V., Young, A., 1986. Understanding face recognition. Br. J. Psychol. 77 (3), 305–327. https://doi.org/10.1111/j.2044-8295.1986.tb02199.x.

Bruyer, R., Brysbaert, M., 2011. Combining speed and accuracy in cognitive psychology: is the inverse efficiency score (IES) a better dependent variable than the mean reaction time (RT) and the percentage of errors (PE)? Psychol. Belg. 51 (1), 5–13. https://doi.org/10.5334/pb-51-1-5.

Burton, A.M., 2013. Why has research in face recognition progressed so slowly? The importance of variability. Q. J. Exp. Psychol. 66 (8), 1467–1485. https://doi.org/10.1080/17470218.2013.800125.

Burton, A.M., White, D., McNeill, A., 2010. The glasgow face matching test. Behav. Res. Methods 42 (1), 286–291. https://doi.org/10.3758/BRM.42.1.286.

Busigny, T., Robaye, L., Dricot, L., Rossion, B., 2009. Right anterior temporal lobe atrophy and person-based semantic defect: a detailed case study. Neurocase 15 (6), 485–508. https://doi.org/10.1080/13554790902971141.

Busigny, T., Van Belle, G., Jemel, B., Hosein, A., Joubert, S., Rossion, B., 2014. Face-specific impairment in holistic perception following focal lesion of the right anterior temporal lobe. Neuropsychologia 56, 312–333. https://doi.org/10.1016/j.neuropsychologia.2014.01.018.

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.

De Renzi, E., Faglioni, P., Grossi, D., Nichelli, P., 1991. Apperceptive and associative forms of prosopagnosia. Cortex 27 (2), 213–221. https://doi.org/10.1016/S0010-9452(13)80125-6.

DeGutis, J., Cohan, S., Kahn, D.A., Aguirre, G.K., Nakayama, K., 2013. Facial expression training improves emotion recognition and changes neural tuning in a patient with acquired emotion recognition deficits and prosopagnosia. J. Vision 13 (9), 993–993.

DeGutis, J., Cohan, S., Mercado, R.J., Wilmer, J., Nakayama, K., 2012. Holistic processing of the mouth but not the eyes in developmental prosopagnosia. Cogn. Neuropsychol. 29 (5–6), 419–446. https://doi.org/10.1080/02643294.2012.754745.

DeGutis, J., Cohan, S., Nakayama, K., 2014a. Holistic face training enhances face processing in developmental prosopagnosia. Brain 137 (6), 1781–1798. https://doi.org/10.1093/brain/awu062.

DeGutis, J.M., Chiu, C., Grosso, M.E., Cohan, S., 2014b. Face processing improvements in prosopagnosia: successes and failures over the last 50 years. Front. Hum. Neurosci. 8, 561. https://doi.org/10.3389/fnhum.2014.00561.

Delvenne, J.-F., Seron, X., Coyette, F., Rossion, B., 2004. Evidence for perceptual deficits in associative visual (prosop)agnosia: a single-case study. Neuropsychologia 42 (5), 597–612. https://doi.org/10.1016/j.neuropsychologia.2003.10.008.

Demirci, E., Erdogan, A., 2016. Is emotion recognition the only problem in ADHD? Effects of pharmacotherapy on face and emotion recognition in children with ADHD. ADHD Attention Deficit and Hyperactivity Disorders 8 (4), 197–204. https://doi.org/10.1007/s12402-016-0201-x.

Dimitriou, D., Leonard, H.C., Karmiloff-Smith, A., Johnson, M.H., Thomas, M.S.C., 2015. Atypical development of configural face recognition in children with autism, Down syndrome and Williams syndrome. J. Intellect. Disabil. Res. 59 (5), 422–438. https://doi.org/10.1111/jir.12141.

Duchaine, B.C., 2000. Developmental prosopagnosia with normal configural processing. Neuroreport 11 (1), 79–83. https://doi.org/10.1097/00001756-200001170-00016.

Duchaine, B.C., Nakayama, K., 2004. Developmental prosopagnosia and the Benton facial recognition test. Neurology 62 (7), 1219–1220. https://doi.org/10.1212/01.wnl.0000118297.03161.b3.

Duchaine, B.C., Weidenfeld, A., 2003. An evaluation of two commonly used tests of unfamiliar face recognition. Neuropsychologia 41 (6), 713–720. https://doi.org/10.1016/S0028-3932(02)00222-1.

Duchaine, Brad, Nakayama, K., 2005. Dissociations of face and object recognition in developmental prosopagnosia. J. Cognit. Neurosci. 17 (2), 249–261. https://doi.org/10.1162/0898929053124857.

Duchaine, Brad, 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.

Duchaine, Bradley, Germine, L., Nakayama, K., 2007. Family resemblance: ten family members with prosopagnosia and within-class object agnosia. Cogn. Neuropsychol. 24 (4), 419–430. https://doi.org/10.1080/02643290701380491.

Dwyer, P., Xu, B., Tanaka, J.W., 2019. Investigating the perception of face identity in adults on the autism spectrum using behavioural and electrophysiological measures. Vis. Res. 157, 132–141. https://doi.org/10.1016/j.visres.2018.02.013.

Dzhelyova, M., Schiltz, C., Rossion, B., 2020. The relationship between the Benton face recognition test and electrophysiological unfamiliar face individuation response as revealed by fast periodic stimulation. Perception 49 (2), 210–221. https://doi.org/10.1177/0301006619897495.

Favelle, S., Hill, H., Claes, P., 2017. About face: matching unfamiliar faces across rotations of view and lighting. I-Perception 8 (6). https://doi.org/10.1177/2041669517744221, 2041669517744221.

Favelle, S.K., Palmisano, S., Avery, G., 2011. Face viewpoint effects about three axes: the role of configural and featural processing. Perception 40 (7), 761–784. https://doi.org/10.1068/p6878.

Ferris, F.L., Kassoff, A., Bresnick, G.H., Bailey, I., 1982. New visual acuity charts for clinical research. Am. J. Ophthalmol. 94 (1), 91–96. https://doi.org/10.1016/0002-9394(82)90197-0.

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

Fry, R., Wilmer, J., Xie, I., Verfaellie, M., DeGutis, J., 2020. Evidence for normal novel object recognition abilities in developmental prosopagnosia. Royal Society open science 7 (9), 200988. https://doi.org/10.1098/rsos.200988.

Fysh, M.C., Bindemann, M., 2018. The kent face matching test. Br. J. Psychol. 109 (2), 219–231. https://doi.org/10.1111/bjop.12260 (London, England: 1953).

Gainotti, G., 2007. Different patterns of famous people recognition disorders in patients with right and left anterior temporal lesions: a systematic review. Neuropsychologia 45 (8), 1591–1607. https://doi.org/10.1016/j.neuropsychologia.2006.12.013.

Garrido, L., Duchaine, B., Nakayama, K., 2008. Face detection in normal and prosopagnosic individuals. J. Neuropsychol. 2 (1), 119–140. https://doi.org/10.1348/174866407X246843.

Geskin, J., Behrmann, M., 2018. Congenital prosopagnosia without object agnosia? A literature review. Cogn. Neuropsychol. 35 (1–2), 4–54. https://doi.org/10.1080/02643294.2017.1392295.

Gonzalez-Perez, M., Wakui, E., Thoma, V., Nitsche, M.A., Rivolta, D., 2019. Transcranial alternating current stimulation (tACS) at 40 Hz enhances face and object perception. Neuropsychologia 135, 107237. https://doi.org/10.1016/j.neuropsychologia.2019.107237.

Grill-Spector, K., Weiner, K.S., Kay, K., Gomez, J., 2017. The functional neuroanatomy of human face perception. Annual Review of Vision Science 3 (1), 167–196. https://doi.org/10.1146/annurev-vision-102016-061214.

Gross, R., Matthews, I., Cohn, J., Kanade, T., Baker, S., 2010. Multi-pie. Image Vision Comput. 28 (5), 807–813.

Hajian-Tilaki, K., 2013. Receiver operating characteristic (ROC) curve analysis for medical diagnostic test evaluation. Caspian Journal of Internal Medicine 4 (2), 627–635.

Haxby, J.V., Gobbini, M.I., 2011. Distributed neural systems for face perception. In: Oxford Handbook of Face Perception. https://doi.org/10.1093/oxfordhb/9780199559053.013.0006.

Haxby, J.V., Hoffman, E.A., Gobbini, M.I., 2000. The distributed human neural system for face perception. Trends Cognit. Sci. 4 (6), 223–233. https://doi.org/10.1016/S1364-6613(00)01482-0.

Huis in 't Veld, E., Van den Stock, J., de Gelder, B., 2012. Configuration perception and face memory, and face context effects in developmental prosopagnosia. Cogn. Neuropsychol. 29 (5–6), 464–481. https://doi.org/10.1080/02643294.2012.732051.

Jenkins, R., Dowsett, A.J., Burton, A.M., 2018. How many faces do people know? Proc. Biol. Sci. 285, 20181319. https://doi.org/10.1098/rspb.2018.1319, 1888.

Kennerknecht, I., Grueter, T., Welling, B., Wentzek, S., Horst, J., Edwards, S., Grueter, M., 2006. First report of prevalence of non-syndromic hereditary prosopagnosia (HPA). Am. J. Med. Genet. 140A (15), 1617–1622. https://doi.org/10.1002/ajmg.a.31343.

Kramer, R.S.S., Jenkins, R., Young, A.W., Burton, A.M., 2017. Natural variability is essential to learning new faces. Vis. Cognit. 25 (4–6), 470–476. https://doi.org/10.1080/13506285.2016.1242522.

Lavallée, M.M., Gandini, D., Rouleau, I., Vallet, G.T., Joannette, M., Kergoat, M.-J., Busigny, T., Rossion, B., Joubert, S., 2016. A qualitative impairment in face perception in Alzheimer's disease: evidence from a reduced face inversion effect. J. Alzheim. Dis. 51 (4), 1225–1236. https://doi.org/10.3233/JAD-151027.

Liu, C.H., Bhuiyan, M.A.-A., Ward, J., Sui, J., 2009. Transfer between pose and illumination training in face recognition. J. Exp. Psychol. Hum. Percept. Perform. 35 (4), 939–947. https://doi.org/10.1037/a0013710.

Liesefeld, H.R., Janczyk, M., 2019. Combining speed and accuracy to control for speed-accuracy trade-offs (?). Behav. Res. Methods 51 (1), 40–60. https://doi.org/10.3758/s13428-018-1076-x.

Liesefeld, H.R., Fu, X., Zimmer, H.D., 2015. Fast and careless or careful and slow? Apparent holistic processing in mental rotation is explained by speed–accuracy trade-offs. J. Exp. Psychol. Learn. Mem. Cognit. 41, 1140–1151. https://doi.org/10.1037/xlm0000081.

Liu, J., Harris, A., Kanwisher, N., 2002. Stages of processing in face perception: an MEG study. Nat. Neurosci. 5 (9), 910–916.

Logan, A.J., Wilkinson, F., Wilson, H.R., Gordon, G.E., Loffler, G., 2016. The Caledonian face test: a new test of face discrimination. Vis. Res. 119, 29–41. https://doi.org/10.1016/j.visres.2015.11.003.

McKone, E., 2008. Configural processing and face viewpoint. J. Exp. Psychol. Hum. Percept. Perform. 34 (2), 310–327. https://doi.org/10.1037/0096-1523.34.2.310.

Margalit, E., Biederman, I., Herald, S.B., Yue, X., von der Malsburg, C., 2016. An applet for the Gabor similarity scaling of the differences between complex stimuli. Atten. Percept. Psychophys. 78 (8), 2298–2306. https://doi.org/10.3758/s13414-016-1191-7.

Mayer, E., Rossion, B., 2007. Prosopagnosia. In: The Behavioral and Cognitive Neurology of Stroke. Cambridge University Press, pp. 315–334. https://doi.org/10.1017/CBO9780511544880.017.

Mishra, M.V., Likitlersuang, J., Wilmer, J.B., Cohan, S., Germine, L., DeGutis, J.M., 2019. Gender differences in familiar face recognition and the influence of sociocultural gender inequality. Sci. Rep. 9 (1), 1–12.

Nunn, J.A., Postma, P., Pearson, R., 2001. Developmental prosopagnosia: should it be taken at face value? Neurocase 7 (1), 15–27. https://doi.org/10.1093/neucas/7.1.15.

Palermo, R., Rossion, B., Rhodes, G., Laguesse, R., Tez, T., Hall, B., Albonico, A., Malaspina, M., Daini, R., Irons, J., Al-Janabi, S., Taylor, L.C., Rivolta, D., McKone, E., 2017. Do people have insight into their face recognition abilities? Q. J. Exp. Psychol. 70 (2), 218–233. https://doi.org/10.1080/17470218.2016.116 1058.

Peterson, M.F., Zaun, I., Hoke, H., Jiahui, G., Duchaine, B., Kanwisher, N., 2019. Eye movements and retinotopic tuning in developmental prosopagnosia. J. Vis. 19 (9) https://doi.org/10.1167/19.9.7, 7–7.

Rezlescu, C., Pitcher, D., Duchaine, B., 2012. Acquired prosopagnosia with spared within-class object recognition but impaired recognition of degraded basic-level objects. Cogn. Neuropsychol. 29 (4), 325–347.

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. J. Exp. Psychol. Hum. Percept. Perform. 43 (12), 1961–1973. https://doi.org/10.1037/xhp0000400.

Richler, J.J., Cheung, O.S., Gauthier, I., 2011. Holistic processing predicts face recognition. Psychol. Sci. 22 (4), 464–471.

Ritchie, K.L., Burton, A.M., 2017. Learning faces from variability. Q. J. Exp. Psychol. 70 (5), 897–905. https://doi.org/10.1080/17470218.2015.1136656.

Rossion, B., Michel, C., 2018. Normative accuracy and response time data for the computerized Benton Facial Recognition Test (BFRT-c). Behav. Res. Methods 50 (6), 2442–2460. https://doi.org/10.3758/s13428-018-1023-x.

Ruiz, S., Chen, X., Rebuschat, P., Meurers, D., 2019. Measuring individual differences in cognitive abilities in the lab and on the web. PLoS One 14 (12), e0226217. https://doi.org/10.1371/journal.pone.0226217.

Sachdev, P.S., Blacker, D., Blazer, D.G., Ganguli, M., Jeste, D.V., Paulsen, J.S., Petersen, R.C., 2014. Classifying neurocognitive disorders: the DSM-5 approach. Nat. Rev. Neurol. 10 (11), 634–642. https://doi.org/10.1038/nrneurol.2014.181.

Schretlen, D.J., Pearlson, G.D., Anthony, J.C., Yates, K.O., 2001. Determinants of Benton facial recognition test performance in normal adults. Neuropsychology 15 (3), 405–410. https://doi.org/10.1037/0894-4105.15.3.405.

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. https://doi.org/10.1098/rsos.140343.

Stollhoff, R., Jost, J., Elze, T., Kennerknecht, I., 2010. The early time course of compensatory face processing in congenital prosopagnosia. PLoS One 5 (7). https://doi.org/10.1371/journal.pone.0011482.

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. https://doi.org/10.1016/j.cortex.2020.04.016.

Tanaka, J.W., Farah, M.J., 1993. Parts and wholes in face recognition. The Quarterly Journal of Experimental Psychology Section A 46 (2), 225–245. https://doi.org/10.1080/14640749308401045.

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. Behav. Res. Methods 46 (2), 472–487. https://doi.org/10.3758/s13428-013-0382-6.

Towler, A., White, D., Kemp, R.I., 2017. Evaluating the feature comparison strategy for forensic face identification. J. Exp. Psychol. Appl. 23 (1), 47. https://doi.org/10.1037/xap0000108.

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. Q. J. Exp. Psychol. 70 (2), 298–315 https://doi.org/10.1080%2F17470218.2016.1177101.

Watson, T.L.W., 2013. Implications of holistic face processing in autism and schizophrenia. Front. Psychol. 4, 414. https://doi.org/10.3389/fpsyg.2013.00529.

Weigelt, S., Koldewyn, K., Dilks, D.D., Balas, B., McKone, E., Kanwisher, N., 2014. Domain-specific development of face memory but not face perception. Dev. Sci. 17 (1), 47–58. https://doi.org/10.1111/desc.12089.

Weigelt, S., Koldewyn, K., Kanwisher, N., 2012. Face identity recognition in autism spectrum disorders: a review of behavioral studies. Neurosci. Biobehav. Rev. 36 (3), 1060–1084. https://doi.org/10.1016/j.neubiorev.2011.12.008.

White, D., Rivolta, D., Burton, A.M., Al-Janabi, S., Palermo, R., 2017. Face matching impairment in developmental prosopagnosia. Q. J. Exp. Psychol. 70 (2), 287–297. https://doi.org/10.1080/17470218.2016.1173076.

Yerys, B.E., Herrington, J.D., Bartley, G.K., Liu, H.-S., Detre, J.A., Schultz, R.T., 2018. Arterial spin labeling provides a reliable neurobiological marker of autism spectrum disorder. J. Neurodev. Disord. 10 (1), 32. https://doi.org/10.1186/s11689-018-9250-0.

Yue, X., Biederman, I., Mangini, M.C., Malsburg, C. von der, Amir, O., 2012. Predicting the psychophysical similarity of faces and non-face complex shapes by image-based measures. Vis. Res. 55, 41–46. https://doi.org/10.1016/j.visres.2011.12.012.

Zhao, Y., Li, J., Liu, X., Song, Y., Wang, R., Yang, Z., Liu, J., 2016. Altered spontaneous neural activity in the occipital face area reflects behavioral deficits in developmental prosopagnosia. Neuropsychologia 89, 344–355. https://doi.org/10.1016/j.neuropsychologia.2016.05.027.