CORTEX 130 (2020) 64–77

Available online at www.sciencedirect.com

ScienceDirect

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

Research Report

Characterizing developmental prosopagnosia beyond face perception: Impaired recollection but intact familiarity recognition

Anna Stumps [a,b], *Elyana Saad* [a,b], *David Rothlein* [a,b], *Mieke Verfaellie* [c,d] *and Joseph DeGutis* [a,b,*]

[a] *Boston Attention and Learning Laboratory, VA Boston Healthcare System, Boston, MA, USA*
[b] *Department of Psychiatry, Harvard Medical School, Boston, MA, USA*
[c] *Memory Disorders Research Center, VA Boston Healthcare System, Boston, MA, USA*
[d] *Boston University School of Medicine, Department of Psychiatry, USA*

ARTICLE INFO

Article history:
Received 20 November 2019
Reviewed 15 January 2020
Revised 2 March 2020
Accepted 24 April 2020
Action editor Holger Wiese
Published online 18 May 2020

Keywords:
Developmental prosopagnosia
Recognition memory
Receiver operating characteristic
Recollection
Familiarity

ABSTRACT

Converging lines of research suggests that many developmental prosopagnosics (DPs) have impairments beyond face perception, but currently no framework exists to characterize these impaired mechanisms. One potential extra-perceptual deficit is that DPs encode/retrieve faces in a distinct manner from controls that does not sufficiently support individuation. To test this possibility, 30 DPs and 30 matched controls performed an old/new face recognition task while providing confidence ratings, to which a model-based ROC analysis was applied. DPs had significantly reduced recollection compared to controls, driven by fewer ‘high-confidence target’ responses, but intact familiarity. Recollection and face perception ability uniquely predicted objective and subjective prosopagnosia symptoms, together explaining 51% and 56% of the variance, respectively. These results suggest that a specific deficit in face recollection in DP may represent a core aspect of the difficulty in confidently identifying an individual by their face.

Published by Elsevier Ltd.

1. Introduction

Quickly and successfully identifying familiar faces is fundamental to social functioning but poses significant challenges for individuals with developmental prosopagnosia (DP), a neurodevelopmental disorder that affects up to 2.5% of the population (Kennerknecht et al., 2006). DPs have lifelong deficits in recognizing faces but are neurologically intact and

---

Abbreviations: DP, developmental prosopagnosia; DPSD, dual-process signal detection; ROC, receiver operating characteristic.

[*] *Corresponding author*. Boston Attention and Learning Laboratory, VA Boston Healthcare System, 150 S. Huntington Avenue, Boston, MA 02130, USA.

E-mail address: degutis@hms.hardvard.edu (J. DeGutis).

https://doi.org/10.1016/j.cortex.2020.04.016
0010-9452/Published by Elsevier Ltd.

have otherwise normal socio-cognitive, intellectual, and visual functioning.

Though DP is typically diagnosed using face recognition *memory* tasks (e.g., Cambridge Face Memory Test Duchaine & Nakayama, 2006; Famous Faces Test; FFMT), numerous studies have focused on DPs' difficulties in face perception; the ability to perceptually code (e.g., holistically process) and discriminate simultaneously presented faces (e.g., face matching paradigms; Weigelt et al., 2014). DPs as a group have shown deficits in matching faces across expression and lighting changes (e.g., White, Rivolta, Burnton, Al-Janabi, & Palermo, 2017) as well as decreased holistic face processing, integrating feature and configural information into a unified percept (e.g., Avidan, Tanzer, & Behrmann, 2011; though see; Biotti, Gray, & Cook, 2017). This holistic deficit may be especially pronounced for the eye region (i.e., integrating the eyes in particular with other facial features into a holistic percept; Chapman, Bell, Duchaine, & Susilo, 2018; DeGutis, Cohan, Mercado, Wilmer, & Nakayama, 2012). Further, compared to controls, DPs have also shown reduced face-selectivity in core face-selective regions (e.g., fusiform and occipital face areas) when viewing faces and objects during fMRI (Jiahui, Yang, & Duchaine, 2018). During EEG, DPs have shown a reduced N170 face inversion effect (Towler, Gosling, Duchaine, & Eimer, 2012) and reduced 'super additivity' in the N250r for repeated whole versus partial faces (Towler, Fisher, & Eimer, 2018) in comparison to controls.

Despite these group-level perceptual differences between DPs and controls, other studies have found considerable heterogeneity in DPs' perceptual performance (Dalrymple, Garrido, & Duchaine, 2014; McKone et al., 2011; Ulrich et al., 2017) as well as in the perceptual neural mechanisms implicated (Towler, Fisher, & Eimer, 2017). Ulrich et al. (2017) tested 11 DPs on face perception measures commonly deficient in DPs (e.g., holistic processing, viewpoint changes). Though they found small but significant differences at the group level in several of the perceptual tasks, 6 of 11 DPs' scores were within 2 standard deviations of the control group mean on *all* of the tasks, and the remainder showed a heterogeneous performance profile. The lack of perceptual deficits in a considerable subgroup of DPs was also observed by Dalrymple et al. (2014) as well as McKone et al. (2011), who found that 62.50% (10/16) and 100% (6/6) of their DP sample scored within 2 standard deviations of the control group mean on the Cambridge Face Perception Test, respectively. Together, these results suggest that many DPs have intact face perception abilities, and additional deficits beyond face perception must explain their face recognition deficits.

To date, the field has defined extra-perceptual deficits as simply prosopagnosia in the *absence* of perceptual deficiencies (see Biotti, Gray, & Cook, 2019 for a discussion). However, it is crucial to characterize specific impairments beyond perception to develop a deeper understanding of DP and create more sophisticated models of DP subtypes. One candidate impairment is the ability to maintain a face representation over a short period of time, i.e., short-term memory for faces. However, several recent studies have failed to find evidence that short delays produce recognition impairments that were not present in no-delay conditions (Biotti et al., 2019; Jackson, Counter, & Tree, 2017; Shah, Gaule, Gaigg, Bird, & Cook, 2015). Another potential extra-perceptual deficit is that DPs encode and/or retrieve faces in a different manner than controls. Successful face encoding not only relies on perceptual processing, but also involves, among other processes, making semantic associations and judgments about the to-be-remembered face (e.g., trustworthiness) and comparing this face to representations stored in memory (e.g., Winograd, 1981). Failure to engage in these additional aspects of encoding/retrieval may independently contribute to DPs' face recognition deficits, particularly in DPs with minimal perceptual deficits. A study by Burns and colleagues (Burns, Tree, & Weidemann, 2014) used a Remember/Know (RK) procedure to measure the contribution of recollection and familiarity to recognition memory in a small sample of DPs ($N = 8$). Recollection is thought to reflect an all-or-none, threshold process that involves the retrieval of qualitative information, including perceptual, contextual or semantic details, associated with target items (Yonelinas, 2001). This associative process supports high confidence recognition decisions that are bolstered by elaborative encoding strategies (Yonelinas, 2002). Alternatively, participants can rely on familiarity, a signal-detection process that involves a 'feeling of knowing' without associated qualitative information (Yonelinas, 2002). Familiarity judgements are assessments of quantitative memory strength based on the assumption that previously studied items will be more familiar than unstudied items. These judgements tend to be related to the global similarity of test items with studied items, as changing the perceptual characteristics of items between study and test selectively diminishes familiarity (for a review see Yonelinas, 2002). In the study by Burns et al. DPs and controls performed multiple study-test blocks. At study, participants were presented with 6 faces repeated four times. At test, they saw 12 faces (6 old/6 new) and made "remember," "know," or "new" judgments for each face. Participants were instructed to select "remember" if they could recall contextual details about the studied face, "know" if the face evoked a sense of familiarity but was devoid of contextual details, or "new". The results showed that DPs had lower recognition accuracy than controls and notably, gave far fewer "remember" responses. The groups did not differ in their use of "know" responses. These results raise the possibility that DPs may have a selective deficit in recollection-based face recognition.

One drawback to the Burns et al. (2014) study is that for both groups, discrimination of old versus new faces on the basis of "know" responses was close to zero, yielding floor effects on familiarity. Thus, impaired recollection in DPs could simply reflect generally deficient face recognition. Further, it is not known whether DPs' recollection deficits were distinct from their face perception deficits, as an impoverished percept may contribute to impoverished face encoding and subsequent poor recollection. Finally, RK results must be interpreted with caution, as the task relies on introspective, subjective reports of underlying memory processes (Rotello, Macmillan, Reeder, & Wong, 2005), and it could be that DPs and controls interpret the RK instructions differently.

In the current study, to more thoroughly examine the contribution of distinct memory mechanisms to DPs' performance, we obtained confidence ratings during an old/new face recognition task and applied a receiver operating characteristic (ROC) model-based approach. This approach is useful when comparing patients versus controls because, unlike the RK procedure, it does not require introspection about qualitative differences in memory judgment, and it accounts for response bias by relating the cumulative hit rate and cumulative false alarm rate (Yonelinas, Kroll, Dobbins, Lazzara, & Knight, 1998). This approach also allowed for an assessment of whether a dual-process model of recognition memory (Yonelinas et al., 2002) best accounted for the differences between DPs and controls. Finally, we also administered a validated battery of face perception tasks in order to examine whether DPs' face memory impairments explained additional variance in objective and subjective prosopagnosia symptoms beyond DPs' perceptual abilities.

2. Materials and methods

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

2.1. Participants

Participants were between the ages of 18 and 65 years old. Individuals with DP were recruited from four sources: a) Our database of Boston DPs who previously participated in laboratory studies, b) DPs referred to us from Dr. Matthew Peterson at MIT, who recently completed a Boston-area DP study (Peterson et al., 2019), c) Individuals referred to our lab from Dr. Brad Duchaine's website, www.faceblind.org, and d) Individuals responding 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. Participants were pre-screened over the phone and excluded from participation if they met *any* of the following criteria: a history of a significant neurological disorder, 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, and current dependence on alcohol or other substances. The minimum sample size for each group was defined based on Burns et al., 2014 (DP $N = 8$; control $N = 20$). However, we significantly increased our sample size to have the sensitivity to detect group differences in familiarity.

2.2. Qualifying as a DP or control participant

To qualify as a DP, we required individuals to report a lifelong history of face recognition difficulties (e.g., not resulting from an event such as a brain injury) and score > 65 on the Prosopagnosia Index questionnaire (PI-20; Shah, Gaule, Sowden, Bird, & Cook, 2015), a self-report measure of prosopagnosia symptoms. We also required z-scores of < -1 for mild DP and < -2 for major DP on both the Famous Faces test (FFMT) and the Cambridge Face Memory Test (CFMT; Duchaine & Nakayama, 2006; see Data and Digital Materials Availability), based on DSM-5 criteria for mild and major cognitive deficits (Sachdev et al., 2014). To rule out other causes of poor face recognition, participants had to score normally on a visual acuity/contrast sensitivity test (Functional Acuity Contrast Test [FRACT]; Ferris III, Kassoff, Bresnick, & Bailey, 1982) and a mid-level vision battery (Leuven Perceptual Organization Screening Test [L-POST]; Torfs, Vancleef, Lafosse, Wagemans, & De-Wit, 2014; see Data and Digital Materials Availability) and within the normal range on the Autism Spectrum Quotient questionnaire (<33; Baron-Cohen, Wheelwright, Skinner, Martin, & Clubley, 2001).

Controls had similar screening criteria to DPs except they could not complain of lifelong face recognition difficulties, score higher than 64 on the PI-20, or exceed the cutoffs on both the FFMT and CFMT. Based on these criteria, 32 DPs and 33 controls were invited to participate in all portions of the study. Two additional DPs and 3 additional control subjects were later removed based on their performance on the old/new recognition task (see Old/new face recognition paradigm). Prior to data collection, consent was obtained for all participants according to the Declaration of Helsinki. The study was approved by the VA Boston Healthcare System and Harvard Medical School Institutional Review Boards, and all study tasks were completed at the Boston VA Medical Center or the Harvard Decision Science Lab.

2.3. Old/new face recognition paradigm

We used 120 photographs of front-view, greyscale, neutral expression faces from the FERET face database (see Data and Digital Materials Availability; Phillips, Wechsler, Huang, & Rauss, 1998; Phillips, Moon, Rizvi, & Rauss, 2000) that served as targets (old faces) and lures (new faces) for the recognition test. Faces were cropped using an oval, removing clothing and shoulders. The size of each oval was 8 cm wide x 8.5 cm high and participants were seated 60 cm from the computer screen. Age, gender, and ethnicity were matched across lure and target faces. To further ensure similarity between target and lure faces, we attempted to match lures to target faces based on their verbal descriptions of distinctive features (e.g., blonde hair, very thin eyebrows).

The old/new face recognition paradigm was administered in PsychoPy version 1.85.4 (Peirce & MacAskill, 2018; see Data and Digital Materials Availability). During the study phase, participants were instructed to study the faces for a later memory test. Faces were presented one at a time in the center of the screen for 1.5 s, with the set of faces presented in the same order twice. All subjects received the same order of faces. Immediately after the study phase, participants were presented with the 60 target and 60 lure faces randomly intermixed, and on each trial were asked to rate on a scale of 1–6 their level of confidence in classifying each face as "old" or "new" (1 – Confident Old, 2 – Somewhat Sure Old, 3 – Guessing Old, 4 – Guessing New, 5 – Somewhat Sure New, 6 – Confident New). Confidence ratings appeared directly below each face. Participants were instructed to try and use all confidence ratings when responding, as selective use of only one or two response categories can distort individual ROC curves and add artifacts to group average ROC curves (Yonelinas & Parks, 2007).

Inspecting the confidence rating bins of all participants revealed that two control subjects and one DP showed extremely biased response patterns (e.g., only using two confidence bins for all items); these subjects were excluded from analyses. One control subject was excluded due to reportedly confusing the confidence ratings during the task, and one DP was excluded because their overall performance did not exceed chance. This left a final sample of 30 DPs and 30 controls.

2.4. *Face perception battery*

To examine if encoding and retrieval deficits in DPs potentially explain variance in prosopagnosia symptoms unique from face perception, we also included a battery of validated face perception tests that required participants to match simultaneously presented faces. All tasks were administered in PsychoPy version 1.85.4 (Peirce & MacAskill, 2018; see Data and Digital Materials Availability). We took the composite score of these tests, as it may be possible for DPs to score normally on any particular face perception test (Ulrich et al., 2017), and a composite score is often more reliable than the constituent measures. In particular, we administered the Cambridge Face Perception Test (CFPT; Duchaine, Germine, & Nakayama, 2007), Computerized Benton Face Recognition Test (BFRT-c; Benton et al., 1994; Rossion & Michel, 2018), and the University of Southern California Face Perception Test (USCFPT; Biederman, Margalit, Maarek, Meschke, & Shilowich, 2017). We also administered the Telling Faces Together (TFT) task, which is an adaptation of a task that previously showed perceptual matching deficits in DPs (White et al., 2017). In this task, participants perform same/different identity judgments of two face images shown from either different viewpoints or different lighting conditions (50% same identity/50% different identity). Scores from these face perception tests were converted to z-scores using the mean and standard deviation of the control sample and then averaged to create a face perception composite score. The CFPT was not included in the perceptual composite score, as it was only administered to DPs.

2.5. *Statistical analysis*

2.5.1. *ROC analysis model fitting*

We used the old/new recognition task confidence ratings and accuracy scores to create ROC curves separately for DPs and controls. Our analyses focused on the dual-process signal detection model (DPSD), since previous work suggested that DPs have a specific deficit in recollection (Burns et al., 2014). To be more comprehensive, we also performed ROC analyses using a popular single-process model, the unequal variance signal detection model (UVSD; see Supplementary Materials for the methods and results). We modeled the data using Koen et al.'s (2017) ROC toolbox function in Matlab version 2019a (see Data and Digital Materials Availability). Average ROC curves and z-ROC curves were predicted by plotting the cumulative hit rate (P("old"|old)) against the cumulative false alarm rate (P("old"|new)). The first point on the ROC graph represents the most conservative response criterion (1 − Confident Old), and as one moves along the graph, the following points represent a relaxation of the response criterion.

The DPSD model predicts the theoretical constructs of familiarity (F) and recollection of oldness (R_o; Yonelinas, 1994; Yonelinas & Parks, 2007). The recollection of newness (R_n) parameter in this model was constrained to equal zero, in line with the most common approach (Yonelinas & Parks, 2007), as new items have no previously encoded qualitative information associated with them, and thus cannot be recollected in item recognition paradigms. In this model, target items that are above the recollection threshold are classified as old, whereas target items that fall below this threshold will either be classified as old on the basis of a feeling of familiarity or misclassified as new. As R_n was constrained to zero, it is assumed that new items are classified on the basis of familiarity, such that items with a weaker familiarity signal will be classified as new. For item recognition tests, ROC curves that reflect familiarity are symmetrical and curvilinear, whereas recollection is characterized by asymmetrical curves with greater y-intercepts (see Yonelinas & Parks, 2007 for a review).

After fitting the DPSD model to the ROC curves, using RStudio version 1.1.383 we ran independent samples *t*-tests to compare recollection and familiarity across the groups. Statistical significance for all tests was determined at an alpha level of $p < .05$. Comparisons of performance and model parameters were repeated for a reduced sample of 15 controls and 15 DPs matched on overall old/new performance. To preview, the performance-matched sample results suggest that the dual-process model provides a better account for the data than the single-process model. Therefore, our analyses focused on the DPSD model. However, these group comparison analyses applied to the unequal variance signal detection (UVSD) model are also presented in the Supplementary Materials.

2.5.2. *Item analyses*

DPs and controls may qualitatively or quantitatively differ in how they attend to faces, which may be reflected in differing patterns of errors between DPs and controls. To test whether the pattern of item-level performance differed between DPs and controls, we measured the extent to which item difficulty (average accuracy for each item) was consistent within and across DP and control groups, separately for targets and lures. The difficulty of each face stimulus item was indexed as the percentage of participants who responded correctly to a given stimulus such that more challenging items would have a lower percentage correct. Additionally, since each target had a matched lure, we computed a third item difficulty measure consisting of the number of hits for each target minus the number of false alarms to its matched lure (hit rate - false alarm rate). Next, we determined how well aligned difficulty values for targets, lures, and target–lure pairs were within each group as well as between the two groups by computing the split-half reliability for DPs and controls as well as across DPs and controls. The within-group split-half reliability measure was computed by randomly splitting each group into two sub-groups of 15 participants (10,000 repetitions) and correlating (Pearson) the resulting set of item-difficulties from each group. The split-half reliability consisted of the average correlation coefficient across repetitions. Matched cross-group reliabilities were computed identically, except the subgroups were composed of randomly selected participants (without replacement) from each group (15 DPs and 15 controls). Finally, a group-label permutation test was used to test whether the within or cross-group split-half reliabilities were greater than would be expected assuming no distinction between DPs and controls. Here, group assignment (DP or control) was randomly permuted (10,000 repetitions), and the split-half reliability measures computed above were recomputed using these pseudo-groups and used to construct null permutation distributions. Significance (*p*) was quantified as the percentile of the real reliability value relative to its null permutation distribution.

2.5.3. *Recollection/familiarity predicting face perception and prosopagnosia symptoms*

To determine the relationship of both recollection and familiarity with perception, across the entire sample we computed a Pearson correlation for each memory parameter with the perceptual composite. In addition, across the entire sample we ran a series of separate linear regression models with both recollection and the perceptual composite predicting 1) the severity of prosopagnosia symptoms (PI-20), 2) face memory abilities (CFMT), and 3) famous face recognition abilities (FFMT) to determine if recollection and perception independently contribute to variance in prosopagnosia symptoms. We also ran a logistic regression model to examine if recollection and perception independently predicted group membership (DP *vs* control). Within DPs, we also explored whether recollection or familiarity were associated with subjective and objective prosopagnosia symptoms. Though we focus our analyses on the dual-process model, for a full examination of the data we also ran these linear and logistic regression models with the UVSD model parameter estimate, Vo, and the perceptual composite as predictors (see Supplementary Materials). Of note, no part of the study procedures or analyses were pre-registered prior to the research being conducted.

3. Results

3.1. *Demographics and diagnostic test performance*

Our sample included 30 DPs (22 females) and 30 controls (18 females) with a mean age of 38.03 (*SD* = 12.55) and 33.87 (*SD* = 13.23), respectively. The groups did not significantly differ by age or gender. Based on DSM-5 criteria of impairment (<-1 SD for mild, < −2 SD for major) on diagnostic face recognition measures, our DP sample included 3 mild DPs and 27 major DPs. As expected, the DP group performed significantly worse than controls on objective diagnostic measures of face recognition, CFMT and FFMT, and endorsed significantly more prosopagnosia symptoms on the PI-20 (see Table 1). We also found that DPs performed worse than controls on our face perception battery (see Table 1). Similar to several previous studies (e.g., Ulrich et al., 2017), these perceptual differences were less pronounced than those on the recognition tests (see Table 2). Analyses (see Supplementary Results and Figs. 3) revelated these results were consistent with a negatively shifted distribution of perceptual performance, similar to other DP studies (e.g., Biotti et al., 2019).

3.2. *Overall old/new recognition performance and confidence ratings*

As expected, DPs' overall accuracy (*M* = 63.37, *SD* = 5.39) was significantly lower than that of controls (*M* = 73.00, *SD* = 8.93, $p < .001$; see Table 3). This difference was driven by DPs' poorer ability to classify old faces as old (*M* = .57, *SD* = .09) compared to controls (*M* = .75, *SD* = .10, $p < .001$), as DPs' false alarm rates were very similar to controls' (DPs: *M* = .31, *SD* = .10; controls: *M* = .29, *SD* = .15, *p* = .658).

We next examined how the use of confidence ratings differed between DPs and controls. When judging targets, controls had significantly more "confident old" responses on average (*M* = 30.50, *SD* = 9.04) than DPs (*M* = 12.73, *SD* = 5.48, $p < .001$), whereas DPs had significantly more "somewhat sure old" and "guessing old" responses than controls (see Fig. 1). Interestingly, there were no significant differences between DPs' and controls' "confident new" responses when judging lure faces (DPs: *M* = 12.60, *SD* = 7.84; controls: *M* = 11.80, *SD* = 10.13, *p* = .734), nor were there differences in the number of "somewhat sure new" and "guessing new" responses (all *p*'s > .15). Importantly, these results demonstrate that DPs do not generally have less confidence than controls when making face recognition judgments, but that their reduced confidence is specific to recognizing *previously studied* faces.

3.3. *Group and individual ROC curves in DPs versus controls*

We used the combination of accuracy and confidence ratings to compare ROC curves between DPs and controls using a

Table 1 – Demographics, face recognition memory, and face perception performance.

Measure	DP	Control	*p*-value	Cohen's *d*
n	30	30	–	–
M:F	7:22	12:18	.107	–
Age	38.03 ± 12.55	33.87 ± 13.23	.216	.32
Autism Quotient – (AQ)	19.30 ± 9.03	–	–	–
Cambridge Face Memory Test Total Score – (CFMT)	39.37 ± 4.77	60.24 ± 5.94	<.001	3.87
Famous Faces – (FFMT %)	36.67 ± 13.32	84.12 ± 12.83	<.001	3.63
Prosopagnosia Index Total Score – (PI-20)	82.30 ± 9.21	35.27 ± 7.97	<.001	5.46
Perceptual Composite	−.90 ± .68	.00 ± .68	<.001	1.32

Note. Mean ± standard deviation. *p*-values are from *t*-tests and χ^2 tests comparing DPs and controls.

Table 2 – DPs' raw data: Demographics, face recognition memory and face perception performance.

Participants	Age	Gender	CFMT	FFMT	Perceptual Composite	CFPT	Recollection	Familiarity
1	22	F	34	.27	−.78	42	.17	.20
2	29	F	37	.35	−.42	70	.32	.60
3	34	F	39	.33	−.35	50	.03	.59
4	61	M	36	.29	−2.22	66	.00	.85
5	36	X[a]	35	.54	−1.59	36	.05	1.06
6	33	M	36	.53	−2.60	76	.03	.43
7	27	M	38	.19	−.03	32	.17	1.28
8	46	F	34	.39	−.84	64	.11	.59
9	53	F	35	.40	−1.35	40	.09	.54
10	26	F	42	.62	−.47	34	.08	.69
11	35	F	43	.45	−1.32	44	.15	.82
12	30	F	41	.43	−.19	54	.24	.92
13	32	F	40	.56	−.28	60	.08	.08
14	36	F	47	.50	−.59	44	.13	.54
15	27	F	44	.29	.18	40	.00	.43
16	63	F	32	.38	−.90	104	.09	.16
17	30	F	37	.20	−.60	52	.21	.34
18	55	F	47	.35	−.59	62	.13	.58
19	39	F	33	.47	−1.73	52	.12	.48
20	37	M	33	.47	−1.02	50	.14	.44
21	28	F	33	.35	−.99	46	.16	.01
22	64	F	39	.47	−.17	58	.09	.90
23	30	M	43	.25	.07	56	.00	.74
24	52	F	49	.43	−1.28	86	.13	.58
25	51	F	38	.12	−1.17	48	.00	.70
26	25	F	45	.33	−2.14	46	.13	.53
27	51	M	44	.27	−.95	52	.00	.54
28	39	F	44	.20	−1.08	30	.18	.04
29	23	F	42	.08	−.93	62	.15	.30
30	27	F	41	.50[b]	−.80	70	.12	.24
DPs	38.03 ± 12.55	7:22	39.37 ± 4.77	36.67 ± 13.32	−.90 ± .68	54.20 ± 16.27	.11 ± .08	.54 ± .30
Controls	33.87 ± 13.23	12:18	60.24 ± 5.94	84.12 ± 12.83	.00 ± .68		.38 ± .18	.71 ± .56

Note. Summary data for DPs and controls are represented by the mean ± standard deviation. CFMT: Cambridge Face Memory Test, FFMT: Famous Faces Test, CFPT: Cambridge Face Perception Test. Control subjects were not administered the CFPT.

[a] This participant's gender identity is nonbinary.

[b] This participant was administered a different version of the FFMT than the rest of the sample (i.e., included different celebrities' faces).

dual-process model of recognition memory (dual-process signal detection, DPSD). We also performed analyses using a popular single-process model (unequal variance signal detection, UVSD; see Supplementary Materials and Figs. 1). However, we focus on the DPSD approach because when equating DPs and controls on overall accuracy, the DPSD model better fit the results (see section below).

Quantitative measures of goodness-of-fit suggested the DPSD model fit the overall results well (AIC = 369.10, BIC = 388.61). The model fit was significantly better for controls (AIC = 348.52, BIC = 368.03) than DPs (AIC = 389.68, BIC = 409.19, $p < .001$). The ROC curve analyses, as can be seen in Fig. 2a, showed that DPs displayed a more *symmetrical* probability ROC curve, whereas the controls' curve was more *asymmetric* and had a higher y-axis intercept. These group

Table 3 – Old/new performance and dual-process model parameter estimates.

Measure	DP	Control	*p*-value	Cohen's *d*
Overall Performance—(%)	63.37 ± 5.39	73.00 ± 8.93	<.001	1.31
Hit Rate	.57 ± .09	.75 ± .10	<.001	1.89
False Alarm Rate	.31 ± .10	.29 ± .15	.658	.16
Discriminability (HR-FAR)	.27 ± .11	.46 ± .18	<.001	1.27
d'	.72 ± .31	1.34 ± .62	<.001	1.26
Area Under the Curve	.67 ± .07	.78 ± .09	<.001	1.36
Recollection	.11 ± .08	.38 ± .18	<.001	1.94
Familiarity	.54 ± .30	.71 ± .56	.159	.38

Note. Mean ± standard deviation. *p*-values are from *t*-tests comparing DPs and controls. HR - FAR: Hit Rate - False Alarm Rate.

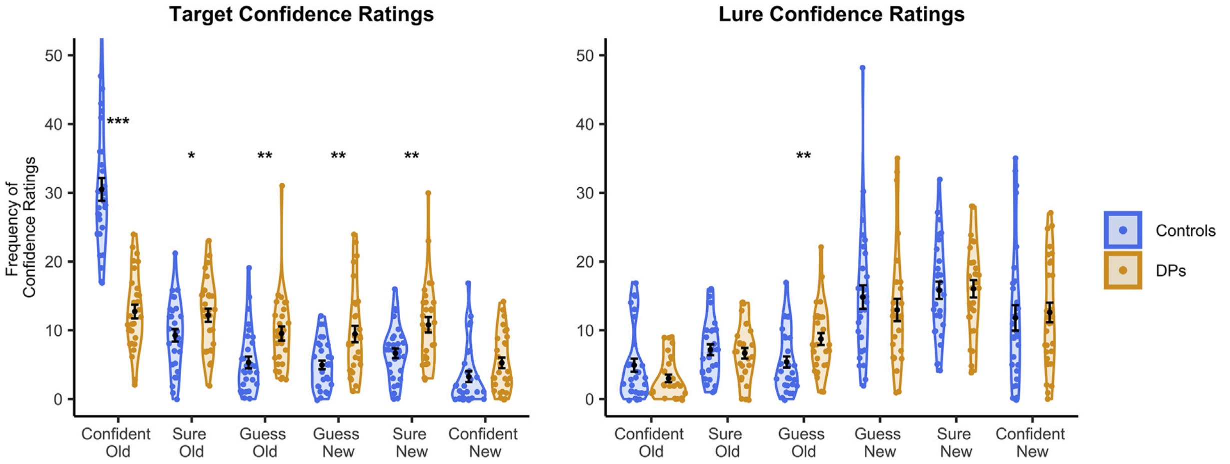
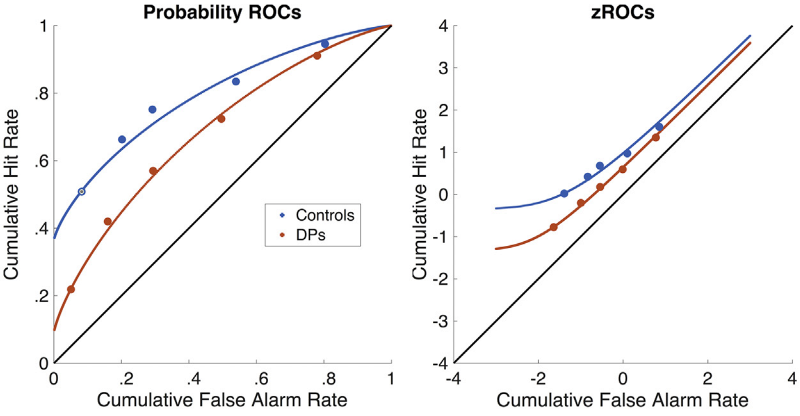
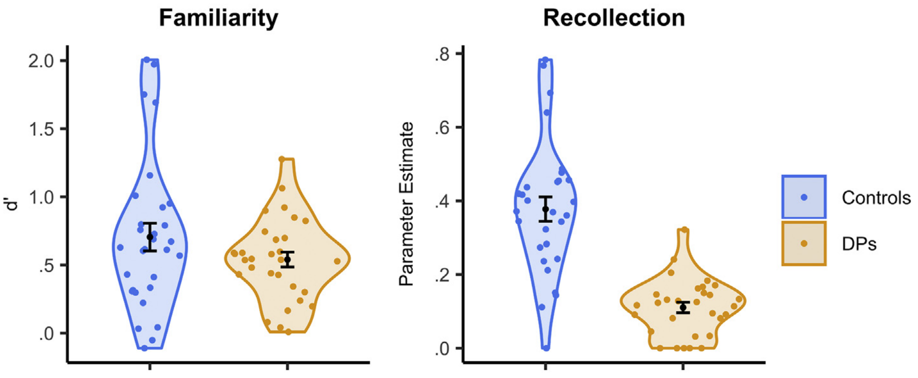


Fig. 1 – Total frequency of each confidence rating response for all target and lure items across DPs and controls. $^{*}p < .05$, $^{**}p < .01$, $^{***}p < .001$. Error bars represent the standard error of the mean.

differences were apparent in each groups' ROC curves at the individual level as well (Fig. 3). In terms of the model parameters, this manifested as DPs having reduced recollection compared to controls (DPs: $M = .11$, $SD = .08$; controls: $M = .38$, $SD = .18$, $p < .001$), but similar familiarity (DPs: $M = .54$, $SD = .30$; controls: $M = .71$, $SD = .56$, $p = .159$) (Fig. 2b).



(a) Group ROC Curves



(b) Parameter Estimates

Fig. 2 – A. Dual-process signal detection (DPSD) average probability ROC and zROC curves for DPs and controls. B. Distribution of familiarity and recollection model parameter estimates across DPs and controls. Error bars represent the standard error of the mean.

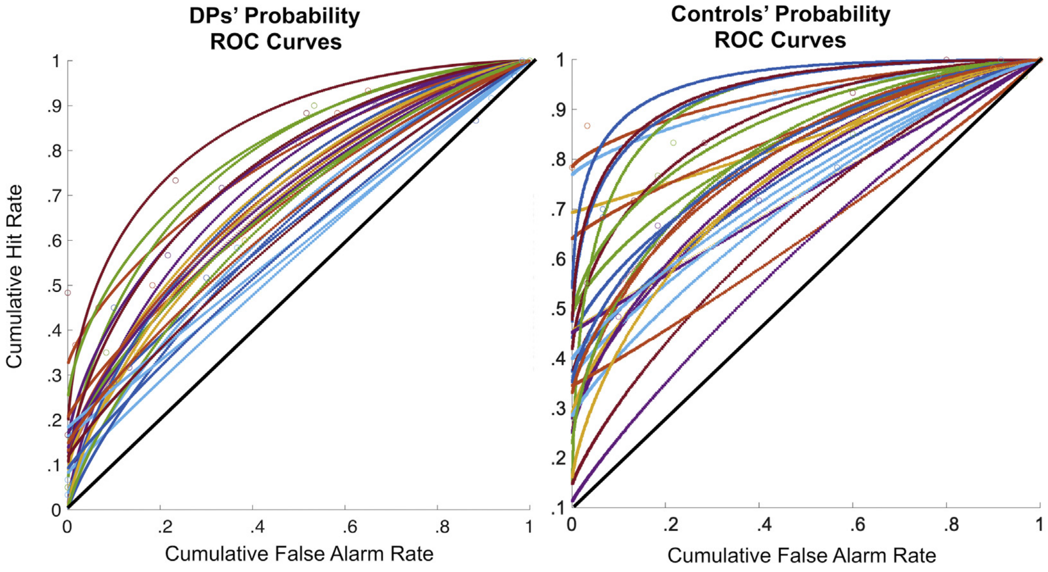


Fig. 3 – DPSD probability ROC curves for each individual DP and each individual control.

3.4. *Group and individual ROC curves in performance-matched DPs versus controls*

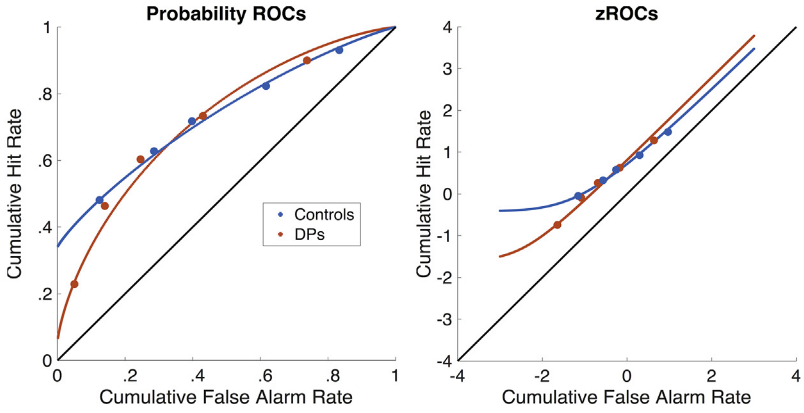
Although these results suggest that DPs and controls use different memory processes for face recognition, a potential alternative explanation, consistent with a single-process model, is that DPs' reduced recollection simply reflects poorer overall performance. By this alternative view, controls with poorer performance should show similarly reduced recollection to DPs. To evaluate these two accounts, we compared performance of the 15 best performing DPs (overall accuracy $M = 67.89$, $SD = 3.27$), and the 15 worst performing controls (overall accuracy $M = 66.00$, $SD = 4.98$) – participants were matched in age (M age $= 37$, $p = .989$). These subgroups did not differ in overall accuracy ($p = .231$), albeit controls had significantly higher hit rates (controls: $M = .72$, $SD = .11$; DPs: $M = .60$, $SD = .10$, $p = .006$), and DPs had significantly lower false alarm rates (controls: $M = .40$, $SD = .13$; DPs: $M = .25$, $SD = .09$, $p = .001$). Critically, comparing the ROC curves for these DPs and controls, we again found that DPs had a more symmetrical probability ROC curve, whereas the controls' curve had a higher y-axis intercept (Fig. 4a). This was reflected in a significant dissociation between recollection and familiarity across groups that is difficult to account for by a single-process signal detection model. DPs showed significantly *lower* recollection ($M = .11$, $SD = .09$) than controls ($M = .37$, $SD = .18$, $p < .001$), but *higher* familiarity ($M = .75$, $SD = .22$) than controls ($M = .36$, $SD = .30$, $p < .001$). These findings argue against the notion that the impairment in recollection in DPs is simply a consequence of overall poorer performance. Rather, whereas the best performing DPs use familiarity to support their recognition judgments (as does the DP group as a whole), even the worst performing controls rely on recollection (Fig. 4b). This dissociation provides additional preliminary evidence that DPs may rely on a qualitatively different memory mechanism than controls (Tian et al., 2020), but future work is required to more thoroughly address this hypothesis.

3.5. *Item analyses*

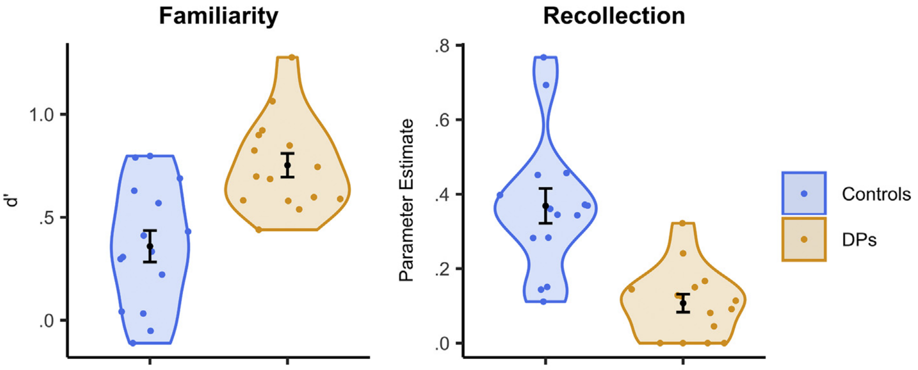
One potential explanation for the reduced contribution of recollection to DPs' face recognition memory is that DPs attended to different aspects of the faces than did controls and, as a result, encoded perceptually impoverished face representations. For example, it has been shown that DPs have specific difficulty encoding the eye region compared to controls but relatively normal processing of the mouth region (DeGutis et al., 2012). It could be that DPs had particular difficulty encoding faces with more distinctive eye regions – faces that would yield high recollection in controls. Arguing against this explanation, we found a strong association between the item-level accuracies between DPs and controls using a cross-group split-half reliability measure for targets ($r = .61$), lures ($r = .75$), and hits minus false alarms ($r = .69$). These between-group associations were not significantly different than the within-group associations for DPs (targets: $r = .67$; lures: $r = .76$; hits minus false alarms: $r = .71$) or controls (targets: $r = .59$; lures: $r = .80$; hits minus false alarms: $r = .68$), suggesting that DPs and controls did not differ in the items that they found more memorable and the items they found difficult (see Figs. 2). This suggests that ROC differences between the DP and control groups were not driven by relative item difficulty differences.

3.6. *Do recollection and face perception independently predict prosopagnosia symptoms?*

Finally, we determined if DPs' recollection deficit was independent from their face perception difficulties. Across the full



(a) Performance Matched Group ROC Curves



(b) Performance Matched Parameter Estimates

Fig. 4 – A. DPSD average probability ROC and zROC curves for the reduced, performance matched sample of the 15 best performing DPs and 15 worst performing controls. B. Distribution of familiarity and recollection model parameter estimates across the reduced, performance matched sample of DPs and controls. Error bars represent the standard error of the mean.

sample, there was a significant correlation between the recollection parameter and the face perception composite measure ($r = .59$, $p < .001$), whereas the familiarity parameter and face perception composite were only weakly related ($r = .19$, $p = .151$). We ran a series of linear regression models across the full sample with recollection and face perception composite scores as predictors and objective (CFMT and FFMT) and subjective (PI-20) face recognition measures as dependent variables. Recollection and face perception scores together explained 51%, 31%, and 56% of the variance in CFMT, FFMT, and PI-20, respectively (see Table 4). Notably, recollection explained unique variance in both objective and subjective prosopagnosia symptoms *independent* of, and to a *greater* extent than, perceptual abilities. Similarly, a logistic regression showed that perception and recollection together strongly predicted prosopagnosia diagnosis (Nagelkerke

Table 4 – Overall sample linear and logistic regression models.

Models	Outcome Variables	Predictors		Model *p*-value	Model Adj R^2
		Recollection	Perceptual Composite		
Model 1	Prosopagnosia Index – (PI-20)	.52***	−.32**	<.001	.56
		(.11)	(.11)		
Model 2	Cambridge Face Memory Test – (CFMT)	.49***	.32**	<.001	.51
		(.11)	(.11)		
Model 3	Famous Faces Memory Test – (FFMT)	.44**	.19	<.001	.31
		(.13)	(.13)		
Model 4[1]	DP Diagnosis	2.92***	1.19	<.001	.72
		(.85)	(.69)		

Note. Standardized betas reported. All values in parentheses represent the coefficient's associated standard error. [1] Model 4 represents the logistic regression model predicting DP diagnosis (R^2 = Nagelkerke R^2). $^{*}p < .05$, $^{**}p < .01$, $^{***}p < .001$.

$R^2 = .72$) and that recollection was a substantially better predictor than perception. Taken together, these results provide compelling evidence that DPs' recollection deficits account for their objective and subjective prosopagnosia symptoms beyond their face perception problems. Finally, we explored whether recollection or familiarity predicted subjective and objective prosopagnosia symptoms *within* DPs. We did not find significant associations with CFMT (recollection: $r = -.07$; familiarity: $r = .02$), FFMT (recollection: $r = -.02$; familiarity: $r = .06$), or PI-20 (recollection: $r = -.17$; familiarity: $r = .22$), all p's $> .24$. This suggests that recollection primarily differentiates DPs from controls rather than DPs from each other.

4. Discussion

The aim of this study was to better understand the nature of the face recognition memory deficit in DP and its contribution to prosopagnosia symptoms. On an old/new recognition test assessing memory for a set of 60 newly encoded faces, DPs had a significantly decreased hit rate and far fewer 'high confidence target' responses than controls. Applying a dual-process recognition model to these results revealed that the contribution of recollection to face recognition memory was greatly reduced in DPs, whereas the contribution of familiarity was intact. Importantly, we found that DPs' impairment in recollection went beyond their face perception deficit, with recollection and a composite measure of face perception *uniquely* predicting DP diagnosis as well as objective and subjective measures of face recognition. Together, these results suggest that DPs not only have poor face recognition, but also have deficits in encoding and/or retrieving a face in an individuated manner that would provide for high confidence recollection. These findings have important mechanistic, theoretical, and treatment implications for DP.

Our findings of impaired recollection and preserved familiarity in DPs build upon work by Burns et al. (2014), who used a Remember/Know (RK) procedure and showed reduced 'remember' responses in DPs. The current study expands on these findings by using a much larger DP sample and an ROC approach that eliminates introspective responses associated with the RK procedure, which could differ between DPs and controls. Though Burns et al. reported a recollection deficit in DPs, because 'know' discrimination was close to zero for both DPs and controls, they were not able to unequivocally assess the status of familiarity. The current study demonstrates that familiarity is intact in DPs, a finding that was especially clear in the comparison of performance-matched DPs and controls, where the best-performing DPs had significantly greater familiarity than the worst-performing controls. Notably, the number of faces to be learned in the current study was much larger than that in Burns et al. (60 *vs* 6, respectively). Despite these differences in memory load, a similar recollection deficit was observed in both studies, suggesting that DPs' recollection deficit in the current study is not likely due solely to enhanced sensitivity to interference associated with the large number of faces (Podd, 1990). Instead, the current results, combined with those of Burns et al., suggest that DPs have a specific deficit in associating studied faces with contextual/semantic information that can provide accurate 'high-confidence old' or 'remember' responses, and instead rely on familiarity or a 'feeling-of-knowing' to determine whether or not they saw a face.

In addition to demonstrating differences between DPs and controls in recollection, the current results show that face recollection was strongly related to several objective and subjective measures of face recognition ability. When combined with the face perception composite, recollection predicted 51% of the CFMT variance, 31% of the FFMT variance, and 56% of the PI-20 variance. Importantly, recollection explained variance in these measures independent from, and to a consistently *greater* degree than, face perception ability. The finding that recollection makes an independent contribution to face recognition may explain why a subset of DPs perform within the normal range on face perception tasks, yet still have severe face recognition deficits (Dalrymple et al., 2014; McKone et al., 2011; Ulrich et al., 2017). This suggests that recollection of recently learned faces may represent a fundamental deficit in developmental prosopagnosia that affects recognition of highly familiar faces as well as newly encountered faces in everyday life.

These findings raise the question of what are the potential mechanisms of DPs' observed recollection deficits. One possible contributing factor may be related to atypical representational 'face-space' in DPs. Face-space is conceptualized as a multi-dimensional representational space with an average of previously encountered faces at the center (for a recent review, see Valentine, Lewis, & Hills, 2016). This framework suggests a norm-based coding of faces such that face identities are compared to the average face, which results in less distinctive faces clustering around the average face, whereas more distinctive identities are represented in lower-density regions, making these identities easier to individuate. In individuals without face recognition deficits, differences in multi-dimensional face space are present when comparing own- to other-ethnicity face processing (MacLin & Malpass, 2001), with other-ethnicity faces clustering more densely in the middle of face space while own-ethnicity faces are more dispersed (Papesh & Goldinger, 2010). Notably, recognition memory for other- compared to own-ethnicity faces has been associated with reduced recollection but intact familiarity (Marcon, Susa, & Meissner, 2009), consistent with the notion that recollection is a function of the distinctiveness of to-be-remembered stimuli (e.g., Mäntylä, 1997; Rajaram, 1998; Yonelinas et al., 2002). Considering these findings, it could be that DPs' impaired recollection but preserved familiarity is in part related to differences in face-space.

Face-space is typically probed with adaptation paradigms in which participants adapt to a distinct individual face (e.g., "Dan"), and this adaptation period temporarily shifts individuals' average face representation in the direction of "Dan," making identity neutral test faces appear to look more similar to the opposite identity (e.g., "antiDan") (Nishimura, Doyle, Humphreys, & Behrmann, 2010; Palermo, Rivolta, Wilson, & Jeffery, 2011). Face-space studies have shown that DPs have significant face identity aftereffects (Nishimura et al., 2010), but as the test faces' identity strength approaches 0 (neutral) such that their identity is less perceptually distinctive from Dan, DPs no longer show an identity aftereffect similar to controls (Palermo et al., 2011). These results, as well as results of our item difficulty analyses showing high correspondence between easier and difficult items in DPs and controls, suggest that DPs' face-space has a similar *layout* to that of controls, but DPs face-space may be more densely clustered or have fewer dimensions (Cenac, Biotti, Gray, & Cook, 2019). Thus, the recollection signal typically cued by perceptual distinctiveness may be slightly weaker in DPs due to the potentially dense clustering of the face representations. The hypothesized relationship between atypical face-space and recollection impairments may explain our observed association between recollection and the perceptual composite. However, as recollection and perception explain unique variance in DPs' prosopagnosia symptoms, atypical face-space does not likely explain all of DPs' recollection deficits. Additional studies relating face-space measures, such as face identity aftereffects, to recollection/familiarity would be useful to determine if face-space differences contribute to DPs' recollection deficits.

Another possible factor contributing to DPs' poor recollection is that they have difficulty conceptually elaborating on perceptual face representations. That is, DPs may encode novel faces only in a *perceptual* manner, whereas controls may generate person-related semantic information that is encoded with the face as well (e.g., he looks trustworthy or like my cousin David; Schwartz & Yovel, 2016, 2019). Notably, the formation of person-related conceptual associations at encoding significantly boosts face recognition compared to perceptually-focused or passive encoding (Schwartz & Yovel, 2019). DPs' difficulty making person-related conceptual associations with faces may stem in part from weakened connections between core face-perception regions and extended face-processing regions, such as anterior temporal regions involved in representing person-related knowledge (Avidan et al., 2013). One implication of this explanation is that DPs' recollection deficit could potentially be ameliorated by using approaches that foster conceptual processing and the formation of person-related associations to faces at encoding (e.g., repetition lag training with faces; Jennings & Jacoby, 2003; Jennings, Webster, Kleykamp, & Dagenbach, 2005).

Though the face-space and person-related conceptual processing accounts of DPs' recollection deficits focus on faces, an important outstanding question concerns the generality of these recollection deficits. DPs typically do not complain of general episodic memory difficulties, indicative of global recollection deficits, and have shown a normal ability to learn and remember new voices (Liu, Corrow, Pancaroglu, Duchaine, & Barton, 2015). Further, DPs have repeatedly been shown to perform in the normal range in word learning/recognition tasks (e.g., Rubino, Corrow, Corrow, Duchaine, & Barton, 2016). However, a recent meta-analysis suggests that nearly half of DPs have difficulty with learning and recognizing new objects (Behrmann & Geskin, 2017). Repeating the current paradigm with novel objects would help determine if the recollection impairment observed here is selective to faces or also affects DPs' recognition of objects.

The finding of preserved familiarity-based recognition in DP also raises interesting questions about the nature of the familiarity signal and the aspects of the face that support this signal. For example, Towler et al. (2018) found that DPs tend to rely on external features when making facial recognition judgments, and DeGutis et al. (2012) showed that DPs may have particular difficulty with processing the eye region. As the stimuli in our study included external features, it is possible that at least some part of the familiarity signal in DPs was driven by a sense of fluency in processing the stimuli's external features. Modifying the faces during the testing period, such as eliminating external features, or examining eye movements during encoding and retrieval could help better elucidate whether the source of familiarity differs in DPs and controls. Another outstanding question concerns the relationship between the familiarity process that supports recognition of novel faces and the mechanism that underlies intact covert recognition of famous faces that are not explicitly recognized (for a review see Rivolta, Palermo, & Schmalzl, 2013). Our findings raise the possibility that the familiarity mechanism identified here also supports the formation of enduring representations of faces of known individuals that can support access to familiarity in covert face recognition tasks.

Though the results of the current study are compelling, one might question the sensitivity of confidence judgments in DP, given that their lifelong experience of not recognizing faces could lead to a lack of confidence in making any memory judgment about a face. This does not appear to be the case, as DPs' reduced confidence in their memory judgments was specific to previously studied faces; DPs had similar confidence as controls when making judgements about foil stimuli. Nonetheless, converging evidence from other methods to separate the contribution of recollection and familiarity that do not rely on confidence judgments (e.g., the process dissociation procedure; Jacoby, 1991) could be useful.

The current results raise exciting new avenues for future research. First, findings from this study provide a theoretically-motivated framework to further investigate extra-perceptual deficits in DPs. Previously, extra-perceptual deficits have been defined as an *absence* of perceptual deficiencies rather than the presence of a specific deficit (Biotti et al., 2019). The current approach provides a framework for studying extra-perceptual mechanisms in DP that also may be involved when DPs learn faces over time and make judgments about familiar faces. Second, the current study also has important implications for studying individual differences in face recognition in non-prosopagnosics and super face-recognizers. Interestingly, there are anecdotal reports that super recognizers readily encode and retrieve contextual and person-related semantic information (e.g., 60 min interview Feb 18, 2012; CBS, 2012) suggestive of reliance on recollection. However, no study to date has examined the relative contribution of recollection and familiarity to superior face recognition abilities in this population. Finally, the current results may inform DP treatment approaches. Whereas the majority of current DP treatments target enhancing perceptual processing of faces (Davies-Thompson et al., 2017; DeGutis, Bentin, Robertson, & D'Esposito, 2007; DeGutis, Cohan, & Nakayama, 2014), complementary training of memory mechanisms may be of added value.

In sum, the current study provides a compelling behavioral demonstration of deficits in DPs beyond face perception. This is an important step towards understanding the mechanisms of DP and face recognition in general and opens up new possibilities for enhancing face recognition.

5. Data and Digital Materials Availability

5.1. *Data*

Data supporting these findings are available at https://osf.io/dah4n/

5.2. *Stimuli*

The images used in the old/new recognition task are from the FERET database of facial images collected under the FERET program, sponsored by the DOD Counterdrug Technology Development Program Office (Phillips et al., 1998, 2000). The images are not to be distributed, published, copied or further redistributed. Specific directions on how to obtain these images can be found at: https://www.nist.gov/itl/products-and-services/color-feret-database. The facial stimuli used in the USCFPT paradigm were obtained from direct contact with Biederman et al. (2017) and cannot be restributed. Finally, the images used in the TFT task are from the CMU Multi-PIE Face Database (Gross, Matthews, Cohn, Kanade, & Baker, 2010) and cannot be redistributed. Information to obtain these images can be found at: http://www.cs.cmu.edu/afs/cs/project/PIE/MultiPie/Multi-Pie/Home.html.

5.3. *Paradigms*

The CFMT was administered using a jar file obtained from direct contact with Duchaine et al. (2007), and the FFMT was presented in Qualtrics. The visual acuity/contrast sensitivity test (Functional Acuity Contrast Test; Ferris III, Kassoff, Bresnick, & Bailey, 1982) and mid-level vision battery (Leuven Perceptual Organization Screening Test; Torfs et al., 2014) are freely available at https://michaelbach.de/fract/ and http://gestaltrevision.be/tests/lpost_consent.php, respectively. The old/new recognition paradigm and battery of face perception tests (USCFPT, TFT and BFRT-c) were presented in PsychoPy version 1.85.4 (Peirce & MacAskill, 2018). The BFRT-c task is available on request to Rossion and Michel (2018). Code supporting the presentation of the old/new recognition task, USCFPT and TFT is available at https://osf.io/dah4n/.

5.4. *Analysis code*

We modeled the data using Koen et al.'s (2017) ROC toolbox function in Matlab version 2019a. The current release of the toolbox can be found at https://github.com/jdkoen/roc_toolbox/releases. Code supporting these specific analyses is available at https://osf.io/dah4n/.

Funding

This study was funded by R01 from the National Eye Institute grant #RO1EY026057 awarded to JD. MV's effort was supported by a Senior Research Career Scientist award from the Clinical Science Research and Development Service, Department of Veterans Affairs. The contents of this manuscript do not represent the views of the U. S. Department of Veterans Affairs or the United States Government.

Author contributions

Stumps, Anna: Methodology, Validation, Formal Analysis, Investigation, Data Curation, Writing – Original Draft and Review and Editing, and Visualization. **Saad, Elyana**: Conceptualization and Methodology. **Rothlein, David**: Formal Analysis, Investigation, and Visualization. **Verfaellie, Mieke**: Writing – Review and Editing and Supervision. **DeGutis, Joseph:** Conceptualization, Methodology, Validation, Formal Analysis, Investigation, Resources, Data Curation, Writing – Original Draft and Review and Editing, Supervision, Project Administration, and Funding Acquisition.

Open practices

The study in this article earned an Open Data badge for transparent practices.

Declaration of Competing Interest

The authors report no competing interests.

Acknowledgements

We want to thank our developmental prosopagnosic and control participants for completing our challenging battery of tasks.

Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.cortex.2020.04.016.

REFERENCES

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

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

Baron-Cohen, S., Wheelwright, S., Skinner, R., Martin, J., & Clubley, E. (2001). The autism-spectrum quotient (AQ): Evidence from asperger syndrome/high-functioning autism, malesand females, scientists and mathematicians. *Journal of Autism and Developmental Disorders, 31*(1), 5–17.

Behrmann, M., & Geskin, J. (2017). Are all visual objects created equal? *Journal of Vision, 17*(10), 1223.

Benton, A. L., Abigail, B., Sivan, A. B., Hamsher, K. D., Varney, N. R., & Spreen, O. (1994). *Contributions to neuropsychological assessment: A clinical manual*. In Oxford University Press.

Biederman, I., Margalit, E., Maarek, R., Meschke, E., & Shilowich, B. (2017). What is the perceptual deficit in developmental prosopagnosia? *Journal of Vision, 17*(10), 619.

Biotti, F., Gray, K. L., & Cook, R. (2017). Impaired body perception in developmental prosopagnosia. *Cortex, 93*, 41–49. https://doi.org/10.1016/j.cortex.2017.05.006.

Biotti, F., Gray, K. L. H., & Cook, R. (2019). Is developmental prosopagnosia best characterised as an apperceptive or mnemonic condition? *Neuropsychologia, 124*, 285–298.

Burns, E. J., Tree, J. J., & Weidemann, C. T. (2014). Recognition memory in developmental prosopagnosia: Electrophysiological evidence for abnormal routes to face recognition. *Frontiers in Human Neuroscience, 8*, 622. https://doi.org/10.3389/fnhum.2014.00622.

CBS. (2012). *Internet dating when you're a super-recognizer*.

Cenac, Z., Biotti, F., Gray, K. L., & Cook, R. (2019). Does developmental prosopagnosia impair identification of other-ethnicity faces? *Cortex, 119*, 12–19.

Chapman, A., Bell, L., Duchaine, B., & Susilo, T. (2018). Varieties of holistic processing deficits in developmental prosopagnosia. *Journal of Vision, 18*(10), 917.

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.

Davies-Thompson, J., Fletcher, K., Hills, C., Pancaroglu, R., Corrow, S. L., & Barton, J. J. S. (2017). Perceptual learning of faces: A rehabilitative study of acquired prosopagnosia. *Journal of Cognitive Neuroscience, 29*(3), 573–591.

DeGutis, J., Bentin, S., Robertson, L. C., & D'Esposito, M. (2007). Functional plasticity in ventral temporal cortex following cognitive rehabilitation of a congenital prosopagnosic. *Journal of Cognitive Neuroscience, 19*(11), 1790–1802.

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

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

Duchaine, B., Germine, L., & Nakayama, K. (2007). Family resemblance: Ten family members with prosopagnosia and within-class object agnosia. *Cognitive Neuropsychology, 24*(4), 419–430. https://doi.org/10.1080/02643290701380491.

Duchaine, B., & Nakayama, K. (2006). The Cambridge Face Memory Test: Results for neurologically intact individuals and an investigation of its validity using inverted face stimuli and prosopagnosic participants. *Neuropsychologia, 44*, 576–585. https://doi.org/10.1016/j.neuropsychologia.2005.07.001.

Ferris, F., III, Kassoff, A., Bresnick, G., & Bailey, I. (1982). *New visual acuity charts for clinical research* (Vol. 94, pp. 91–96). Elsevier, 1 https://www.sciencedirect.com/science/article/pii/0002939482901970.

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

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

Jacoby, L. L. (1991). A process dissociation framework: Separating automatic from intentional uses of memory. *Journal of Memory and Language, 30*(5), 513–541.

Jennings, J. M., & Jacoby, L. L. (2003). Improving memory in older adults: Training recollection. *Neuropsychological Rehabilitation, 13*(4), 417–440.

Jennings, J. M., Webster, L. M., Kleykamp, B. A., & Dagenbach, D. (2005). Recollection training and transfer effects in older adults: Successful use of a repetition-lag procedure. *Aging, Neuropsychology, and Cognition, 12*(3), 278–298.

Jiahui, G., Yang, H., & Duchaine, B. (2018). Developmental prosopagnosics have widespread selectivity reductions across category-selective visual cortex. *Proceedings of the National Academy of Sciences, 115*(28).

Kennerknecht, I., Grueter, T., Welling, B., Wentzek, S., Horst, J., Edwards, S., et al. (2006). First report of prevalence of non-syndromic hereditary prosopagnosia (HPA). *American Journal of Medical Genetics Part A, 140A*(15), 1617–1622. https://doi.org/10.1002/ajmg.a.31343.

Koen, J. D., Barrett, F. S., Harlow, I. M., & Yonelinas, A. P. (2017). The ROC toolbox: A toolbox for analyzing receiver-operating characteristics derived from confidence ratings. *Behavior Research Methods, 49*(4), 1399–1406.

Liu, R. R., Corrow, S. L., Pancaroglu, R., Duchaine, B., & Barton, J. J. S. (2015). The processing of voice identity in developmental prosopagnosia. *Cortex, 71*, 390–397.

MacLin, O. H., & Malpass, R. S. (2001). Racial categorization of faces: The ambiguous race face effect. *Psychology, Public Policy, and Law, 7*(1), 98.

Mäntylä, T. (1997). Recollections of faces: Remembering differences and knowing similarities. *Journal of Experimental Psychology. Learning, Memory, and Cognition, 23*(5), 1203.

Marcon, J. L., Susa, K. J., & Meissner, C. A. (2009). Assessing the influence of recollection and familiarity in memory for own-versus other-race faces. *Psychonomic Bulletin & Review, 16*(1), 99–103.

McKone, E., Hall, A., Pidcock, M., Palermo, R., Wilkinson, R. B., Rivolta, D., et al. (2011). Face ethnicity and measurement reliability affect face recognition performance in developmental prosopagnosia: Evidence from the Cambridge Face Memory Test–Australian. *Cognitive Neuropsychology, 28*(2), 109–146.

Nishimura, M., Doyle, J., Humphreys, K., & Behrmann, M. (2010). Probing the face-space of individuals prosopagnosia. *Neuropsychologia, 48*(6), 1828–1841.

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

Papesh, M. H., & Goldinger, S. D. (2010). A multidimensional scaling analysis of own-and cross-race face spaces. *Cognition, 116*(2), 283–288.

Peirce, J. W., & MacAskill, M. R. (2018). *Building experiments in PsychoPy*. London: Sage.

Peterson, M. F., Zaun, I., Hoke, H., Jiahui, G., Duchaine, B., & Kanwisher, N. (2019). Eye movements and retinotopic tuning in developmental prosopagnosia. *Journal of Vision, 19*(9), 7.

Phillips, J. P., Moon, H., Rizvi, S. A., & Rauss, P. J. (2000). The FERET evaluation methodology for face-recognition algorithms. *IEEE Trans. Pattern Analysis and Machine Intelligence, 22*, 1090–1104.

Phillips, J. P., Wechsler, H., Huang, J., & Rauss, P. J. (1998). The FERET database and evaluation procedure for face-recognition algorithms. *Image and Vision Computing, 16*(5), 295–306. https://doi.org/10.1016/S0262-8856(97)00070-X.

Podd, J. (1990). Th effects of memory load and delay on facial recognition. *Applied Cognitive Psychology, 4*(1), 47–59.

Rajaram, S. (1998). The effects of conceptual salience and perceptual distinctiveness on conscious recollection. *Psychonomic Bulletin & Review, 5*(1), 71–78.

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

Review, 23(2), 111–116. https://doi.org/10.1007/s11065-012-9223-0.

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.

Rotello, C. M., Macmillan, N. A., Reeder, J. A., & Wong, M. (2005). The remember response: Subject to bias, graded, and not a process-pure indicator of recollection. *Psychonomic Bulletin & Review, 12*(5), 865–873.

Rubino, C., Corrow, S. L., Corrow, J. C., Duchaine, B., & Barton, J. J. S. (2016). Word and text processing in developmental prosopagnosia. *Cognitive Neuropsychology, 33*(5–6), 315–328.

Sachdev, P. S., Blacker, D., Blazer, D. G., Ganguli, M., Jeste, D. V., Paulsen, J. S., et al. (2014). Classifying neurocognitive disorders: The DSM-5 approach. *Nature Reviews Neurology, 10*(11), 634.

Schwartz, L., & Yovel, G. (2016). The roles of perceptual and conceptual information in face recognition. *Journal of Experimental Psychology. General, 145*(11), 1493.

Schwartz, L., & Yovel, G. (2019). Learning faces as concepts rather than percepts improves face recognition. *Journal of Experimental Psychology. Learning, Memory, and Cognition, 45*(10), 1733.

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

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

Tian, X., Wang, R., Zhao, Y., Zhen, Z., Song, Y., & Liu, J. (2020). Multi-item discriminability pattern to faces in developmental prosopagnosia reveals distinct mechanisms of face processing. *Cerebral Cortex, 30*(5), 2986–2996.

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, J., Fisher, K., & Eimer, M. (2017). The cognitive and neural basis of developmental prosopagnosia. *The Quarterly Journal of Experimental Psychology, 70*(2), 316–344.

Towler, J., Fisher, K., & Eimer, M. (2018). Holistic face perception is impaired in developmental prosopagnosia. *Cortex, 108*, 112–126.

Towler, J., Gosling, A., Duchaine, B., & Eimer, M. (2012). The face-sensitive N170 component in developmental prosopagnosia. *Neuropsychologia, 50*, 3588–3599. https://doi.org/10.1016/j.neuropsychologia.2012.10.017.

Ulrich, P. I., Wilkinson, D. T., Ferguson, H. J., Smith, L. J., Bindemann, M., Johnston, R. A., et al. (2017). Perceptual and memorial contributions to developmental prosopagnosia. *The Quarterly Journal of Experimental Psychology, 70*(2), 289–315. https://doi.org/10.1080/17470218.2016.1177101.

Valentine, T., Lewis, M. B., & Hills, P. J. (2016). Face-space: A unifying concept in face recognition research. *The Quarterly Journal of Experimental Psychology, 69*(10), 1996–2019.

Weigelt, S., Koldewyn, K., Dilks, D. D., Balas, B., McKone, E., & Kanwisher, N. (2014). Domain-specific development of face memory but not face perception. *Developmental Science, 17*(1), 47–58.

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

Winograd, E. (1981). Elaboration and distinctiveness in memory for faces. *Journal of Experimental Psychology: Human Learning and Memory, 7*(3), 181.

Yonelinas, A. P. (1994). Receiver-operating characteristics in recognition memory: Evidence for a dual-process model. *Journal of Experimental Psychology. Learning, Memory, and Cognition, 20*(6), 1341.

Yonelinas, A. P. (2001). Components of episodic memory: The contribution of recollection and familiarity. *Philosophical Transactions of the Royal Society of London Series B, 356*(1413), 1363–1374.

Yonelinas, A. P., Kroll, N. E. A., Dobbins, I., Lazzara, M., & Knight, R. T. (1998). Recollection and familiarity deficits in amnesia: Convergence of remember-know, process dissociation, and receiver operating characteristic data. *Neuropsychology, 12*(3), 323–339.

Yonelinas, A. P., & Parks, C. M. (2007). Receiver operating characteristics (ROCs) in recognition memory: A review. *Psychological Bulletin, 133*(5), 800–832. https://doi.org/10.1037/0033-2909.133.5.800.

Yonelinas, A. P., Scheib, J., Kroll, N., Lazzara, M., Quamme, J., Kishiyama, M., et al. (2002). The nature of recollection and familiarity: A review of 30 Years of research. *Journal of Memory and Language, 46*, 441–517. https://doi.org/10.1006/jmla.2002.2864.