Routledge
Taylor & Francis Group

Cognitive Neuropsychology

ISSN: 0264-3294 (Print) 1464-0627 (Online) Journal homepage: www.tandfonline.com/journals/pcgn20

An in-depth investigation of face perception in developmental prosopagnosia

Judith Lowes, Peter J. B. Hancock, Benjamin L. Armstrong & Anna K. Bobak

COGNITIVE NEUROPSYCHOLOGY
https://doi.org/10.1080/02643294.2026.2676551

An in-depth investigation of face perception in developmental prosopagnosia

Judith Lowes, Peter J. B. Hancock, Benjamin L. Armstrong and Anna K. Bobak

Faculty of Natural Sciences, University of Stirling, Scotland, Stirling, United Kingdom

ABSTRACT

Developmental prosopagnosia (DP) causes difficulty recognizing familiar faces, but little is known about which stage(s) of the face-processing system difficulties occur, or whether these are identical in all DP cases. Twenty DPs and matched controls completed six face perception tasks. The DP group performed significantly poorer than controls on face detection and gender categorization tasks and on a global (averaged) face perception measure. 40% of individual DPs showed significant impairments on tasks not involving identity perception (face detection and categorization, age and gender categorization). By contrast, 50% showed broadly typical face perception and thus support for the idea of apperceptive (affecting both face perception and face memory) and mnemonic (affecting only face memory) subtypes of DP. Intraclass correlations revealed clear inter- and intra-individual differences, confirming that data from a single perceptual task cannot be taken as evidence of broader face-perception ability spanning beyond the results of that test..

ARTICLE HISTORY
Received 15 October 2025
Revised 11 May 2026
Accepted 14 May 2026

KEYWORDS
Developmental prosopagnosia; face processing; speed–accuracy trade off; face perception; neurodevelopmental conditions; face recognition

1. Introduction

Developmental prosopagnosia is characterized by a severe, lifelong difficulty recognizing familiar faces, often including those of friends and family (Lowes et al., 2025), despite otherwise normal vision and IQ, and an absence of obvious brain damage (Behrmann & Avidan, 2005; Duchaine & Nakayama, 2006).

1.1. What is the nature of the impairment in DP? A memory impairment, a perceptual impairment – or both?

A dissociation between perception and memory has been identified in acquired prosopagnosia (AP) whereby memory for faces can be selectively impaired, leaving face perception unaffected (Davies-Thompson et al., 2014; De Renzi et al., 1991). This is perhaps unsurprising since AP is a result of brain lesions whose location and severity differ among individuals. Despite no visible brain damage in DP, a broad body of neuroscientific evidence shows that impairments in DP can also arise at multiple levels (for a review, see Towler et al., 2017) and can cause difficulties with face memory, or both face perception *and* face memory. By definition, face memory is always affected in prosopagnosia, but Towler and Eimer (2012) reported that, while the face-sensitive N170 component is present, typical N170 inversion effects are reduced or absent in DP, suggesting abnormal *perceptual* processing of upright face structure rather than a purely mnemonic problem. ERP studies also indicate that identity-relevant perceptual mechanisms are impaired in DP; Towler et al. (2018) showed that holistic face perception, integrating identity-diagnostic information across different facial regions, is disrupted, thus providing neural evidence that identity-specific holistic processing is compromised beyond basic feature encoding. At later stages of face processing, Eimer et al. (2012) reported that some individuals with DP show an N250 component to familiar faces (indicating activation of stored visual face memories) without a subsequent P600f, which is typically associated with conscious recognition and access to person-specific semantic information. This pattern suggests that, in at least a subset of DPs, impairment can occur at late recognition stages, after perceptual encoding has taken place. Towler and Eimer (2012) reviewed a broader ERP literature and emphasized that different individuals with DP can show impairments at different temporal stages, from early perceptual encoding to later recognition-related processes.

The existence of potential subtypes has important theoretical and applied implications for the understanding and treatment of DP, a lifelong condition estimated

CONTACT Judith Lowes judith.lowes@stir.ac.uk Psychology, Faculty of Natural Sciences, University of Stirling, Scotland, FK9 4LA, United Kingdom
Supplementary data for this article can be accessed online at https://doi.org/10.1080/02643294.2026.2676551.

to affect between 1% and 4% of the population (Bennetts, Murray, et al. 2017; DeGutis et al., 2023; Kennerknecht et al., 2008). Both face perception and face memory contribute to face identity recognition and a more precise understanding of the loci and nature of the deficits in DP, or in individual DPs, is needed to inform rehabilitation attempts. The most commonly-proposed subtypes of DP are (i) *mnemonic* (Biotti et al., 2019; Ulrich et al., 2017), in which individuals can extract structural descriptions of faces and discriminate faces when there are no, or minimal, memory demands, but have difficulties retrieving previously-seen faces from memory, and (ii) *perceptual* or *apperceptive*, in which individuals have difficulty forming robust facial percepts and thus are impaired at facial discrimination, even when faces are presented simultaneously, in addition to their severe face memory difficulty (Biotti et al., 2019; Ulrich et al., 2017).

Although the dominant Bruce and Young model of face recognition (Bruce & Young, 1986) is hierarchical in nature and suggests that impaired face perception leads to impaired face recognition (but not the other way around), a recent study by DeGutis et al. (2023) reported that, counter-intuitively, DPs with *milder* face memory impairments showed *greater* impairment on the Cambridge Face Perception Test (CFPT) than DPs with more severe face memory deficits, suggesting that severe early-stage perceptual impairment does not always translate into greater memory deficit. Despite the popularity of the CFPT, one important limitation is its relatively low internal reliability, .74 for upright faces and .50 for inverted faces (Bowles et al., 2009), making it arguably unsuitable for use as the *sole* measure of face perception ability.

Although the CFPT (Duchaine et al., 2007) is considered a test of face identity *perception*, principal components analysis of data from a large study ($n = 165$) of potential DPs (Bate et al., 2019) found that the CFPT and the Cambridge Face Memory Test (CFMT, Duchaine & Nakayama, 2006), testing novel face learning, loaded onto one factor, while long-term face memory, as measured by famous face test scores, loaded onto a separate factor. If this is the case, then influential early attempts (Dalrymple et al., 2014) to dissociate face perception and face memory in DP using the two tasks [CFMT and CFPT] that Bate and colleagues subsequently found to load on to the same factor become somewhat difficult to interpret.

1.2. Limitations of the literature to date and motivation for the present study

Some of the heterogenous findings in DP (for a review see Corrow et al., 2016) may partly be explained by methodological issues. For example, previous studies often administered only a small number of perceptual tests (e.g., Bate et al., 2019; Dalrymple et al., 2014), or used tests with lower reliability (e.g., CFPT), from which it is difficult to make wider generalizations about the condition.

Even among typical perceivers, correlations across procedurally similar face matching tasks are weak (Bobak et al., 2023) suggesting that stimulus properties and subtle paradigm differences must be considered when interpreting results. A similar lack of consistency across different face tasks has been found among super recognizers, leading to calls to administer a wider range of face tests to fully capture participants' true performance abilities at both ends of the ability spectrum (Bate et al., 2018; Biotti et al., 2019; Ramon, 2021). A broader, more systematic, approach seems likely to be able to provide a fuller explanation of DP than is possible if relying on a small number of face perception tasks.

A further limitation of the DP literature is that, until recently, it was common practice to exclude from research individuals who reported severe face recognition problems, but who did not score at least 2 SD below control means on a minimum of two objective face processing measures. However, recent studies suggest that DPs with milder face memory impairments appear to perform similarly to those with a more classical face recognition impairment (Burns et al., 2023; DeGutis et al., 2023, 2024) and should therefore be included in DP research. Recent work also suggests that slower reaction times can lead to "typical" accuracy scores and mask impairments in DP (Lowes et al., 2024a). Including these participants can improve our understanding of the full range of the condition and of patterns of face perception performance in DP and how these may differ across the full DP spectrum. To address this, we included participants with mild objective impairment, but who report severe day-to-day difficulties.

Reviews of DP (Bate & Tree, 2017; Robotham & Starrfelt, 2018) have highlighted the need to systematically test the different sub processes involved in face recognition to better understand the precise nature of the deficits that occur in DP. In addition to the neuroscience evidence briefly summarized above, recent behavioural work investigating face perception in DP (Biotti et al., 2019) argued that impairments at the perceptual encoding stage may be more widespread, and play a more important role, in DP than currently acknowledged, and that these impairments should be investigated in a more structured way. We address these recommendations in the present study.

1.3. The present study

We took a novel approach to characterizing perceptual impairment in DP by assembling a comprehensive test battery to hierarchically tap each of the multiple stages of face processing that were mapped in a recent literature review of tests of upright face processing used in DP (Robotham & Starrfelt, 2018), informed by leading models of face recognition (Bruce & Young, 1986; Young & Bruce, 2011). The battery allows a rich understanding of the precise nature of the face perception deficits in DP. Such insights can inform the design of bespoke remedial interventions aimed at improving face recognition.

In classifying our test battery as perceptual, we followed the terminology of Robotham and Starrfelt's (2018) literature review that informed our research design. The authors mapped paradigms testing upright face processing according to what they termed three "levels of processing": Perception, recognition, and identification. Face detection, face categorization, and categorization of facial age and gender paradigms are described by Robotham and Starrfelt as *perception tasks*. We also considered the configural/featural task as a perceptual in line with Yovel and Duchaine (2006). According to their work, performance on this task serves as evidence for a specialized face perception mechanism.

The study's aim was to precisely map the perceptual underpinnings of DP at both *group* and *individual* level. As recommended by Bate and Tree (2017), we tested participants across the adult lifespan (age 14–71 years). Our study design mitigated the risk of overestimating the prevalence of perceptual impairments in DP by using a subjective report and only face *memory* tests for classification (Lowes et al., 2024a). We therefore avoid the "double dipping" problem (Kriegeskorte et al., 2009) which occurs when a face perception test is included as part of the classification battery since this will artificially inflate the proportion of DPs in the sample who demonstrate face perception deficits.

The battery tested *non-identity face perception* (face detection, face categorization, gender and age categorization, feature and configuration change discrimination) followed by *face identity perception*. We then compared participants' performance on these perceptual measures to their performance on four face memory measures.

The pre-registered research questions and associated hypotheses (available at https://osf.io/m3wvg/overview) addressed in this study are outlined in Table 1. Our previous work reporting classification of this same sample (Lowes et al., 2024a) showed that the DP group adopted a different speed–accuracy strategy to the control group, preferring to proceed more slowly and carefully. To account for these differential speed–accuracy trade-offs which masked the true extent of face identity recognition ability deficits among DPs, we therefore also report an additional exploratory measure of the balanced integration score or BIS (Liesefeld & Janczyk, 2019, 2022) on some tasks; these are highlighted as exploratory in Table 1. See Section 2 for full details.

The paper is structured as follows. First (Section 3), we provide a general methodological overview of participants, procedures and statistical approaches that apply to all our research questions and experiments. Next (Section 4), we present each research question in turn, covering motivation, brief test methods, results and discussion. The section begins with a table summarizing results of each hypothesis. Additional methodological details for previously published tests are provided in Supplementary material. Section 4.6 considers the *patterns* of individual level results across all tasks. Formal single case analyses and dissociation analyses are provided in Supplementary Tables S4a–e. The final section (Section 5) is a general discussion.

2. Data transparency statement

Although this study is exploratory in nature, it tests many, sometimes competing, hypotheses. We preregistered the study (https://doi.org/10.17605/OSF.IO/M3WVG) after data collection had started but prior to analysis. Analysis plans, data and R scripts are available at https://osf.io/s57zq/overview.

We report an additional measure that was not preregistered, the balanced integration score or BIS (Liesefeld & Janczyk, 2019, 2022) on the following tasks: Face among non faces; gender categorization; age categorization; configural and featural house and face tasks. BIS helpfully adjusts accuracy (a preregistered measure) for response time (RT; another preregistered measure), creating a single integrated score, thereby controlling for differential speed–accuracy trade-offs within and between participants. Although confidence or motor factors may contribute to slower responses, prior work shows these are unlikely to explain group differences in the BIS. Specifically, studies reporting significantly slower RTs in DPs on a fluid reasoning task and a bicycle memory task (Lowes et al., 2024b) found that these differences disappeared on non-face tasks when using the BIS but persisted on face tasks (Lowes et al., 2024a). This pattern suggests that general confidence issues do not fully account for the BIS group differences. BIS is calculated as $Z_{accuracy} - Z_{RT}$ and the formal

Table 1. Research questions and hypotheses.

Research questions	Hypotheses	Stage of perceptual processing	Task	Outcome variable (s)
1. Are all, or some, DPs impaired at face detection?	1a: If DPs are impaired at detecting faces, then their RTs on face-present upright trials will be longer than those of controls.	Face detection	Faces among non faces	RT BIS (exploratory)
	1b: If DPs are impaired at holistic processing, then they will show a reduced inversion effect on "face present" trials compared with controls.	Face detection	Faces among non faces	Inversion effect subtraction method Inversion effect normalized inversion index (exploratory)
	1c: DPs will be less accurate than controls.	Face categorization	Mooney faces	% correct
2. Are all or some DPs impaired at facial gender categorization?	2: DPs will be less accurate than controls.	Non-identity encoding of facial gender	Gender categorization	% correct BIS (exploratory)
3. Are all or some DPs impaired at facial age categorization?	3: DPs will be less accurate than controls	Non-identity encoding of facial age	Age categorization	% correct BIS (exploratory)
4. Do DPs show deficits in configural face processing? And are any deficits observed domain specific or process specific?	4a: If DP is *domain specific* then DPs will be *more* impaired at detecting changes to upright *faces* than to upright houses compared with controls.	Non-identity encoding of configuration and features	Configural featural	% correct BIS (exploratory) Face advantage score
	4b: If DP is *process specific* and driven by a deficit in configural processing, then DPs will be *more* impaired at detecting changes to *spacing* than to features compared with controls across both faces and houses.	Non-identity encoding of configuration and features	Configural featural	% correct BIS (exploratory)
	4c: If DP is driven by a deficit in *domain-specific holistic processing*, then DPs will show a *reduced* inversion effect for faces, but not houses, compared with controls.	Non-identity encoding of configuration and features	Configural featural	% correct Face inversion subtraction method Face Inversion normalized inversion index (exploratory) House inversion
5. Is face perception ability related to face identity recognition?	5a: Face perception ability is related to face memory ability.	All	CFMT	BIS (exploratory)
			CBMT	CFMT-CBMT BIS (exploratory)
			Old New Faces	BIS exploratory
			FFT	% correct (Z)
			Mooney	% correct (Z)
			Faces among non faces	RT (Z)
			Configural featural	% correct (Z)
			CFPT	Upright accuracy (Z)

calculation for BIS is shown in Supplementary material 1. For RT, all trials < 250 ms were removed from analysis. The only perception task on which trials were removed based on RT was the configural featural faces task, here 18/12,800 trials were removed (0.14%).

For transparency we report both the original preregistered primary outcome measures (either accuracy or RT alone) and BIS for most tasks, We did not consider BIS an appropriate measure for the Mooney face task because the task was to click very precisely on the eyes of the image that contained a face and we reasoned that motor control could disproportionately influence RTs. We did not calculate BIS on the Cambridge Face Perception Test (CFPT) since task instructions emphasize accuracy over speed, and motor control and familiarity with a mouse/touchpad strongly influence RTs. The four face memory measures used to initially classify participants as DPs or typical controls reported elsewhere (Lowes 2024a) are provided in Table 2.

When testing inversion effects (hypothesis 1b) we calculated inversion effects using both the preregistered subtraction method as well as an exploratory normalized inversion method. See Section 4.1 for the rationale and comparison of results.

Table 2. Participant demographics.

		Group		
Age group	Gender	Control	DP	Total
Age 14–35	Female	10	5	15
	Male	11	2	13
	Total	21	7	28
Age 36–59	Female	15	7	22
	Male	6	1	7
	Total	21	8	29
Age 60–74	Female	9	2	11
	Male	9	3	12
	Total	18	5	23
Total	Female	34	14	48
	Male	26	6	32
	Total	60	20	80

The pre-registered research question 4 was specific to children and, because none of the children tested met the classification criteria for DP, this analysis was not possible. For the same reason, we were unable to look at child and adult performance separately as originally envisaged for hypothesis 5a, but we retained this analysis looking at adults only. Neither of these changes had any effect on interpretation since we also removed the age-matched child controls. Preregistered hypothesis 5d was ultimately not included in our battery for reasons of overall test duration.

The preregistered analysis plan specified one-way tests for measures testing a directional hypothesis, however we later decided to revert to the more conservative two-tailed tests for group analyses. Single case analyses are shown with both one-tailed and two-tailed results.

The authors declare that there are no competing interests. Approval for the study was granted by University of Stirling General University Ethics Panel (reference 204). Informed consent and child assent were obtained from all participants, and research was conducted in accordance with the principles of the Declaration of Helsinki.

3. General methods

3.1. Participants

DP participants (DPs) were 20 individuals aged 15–71 years (Table 2) and 60 age- and gender-matched controls. All participants lived in the UK, Ireland or USA and reported normal, or corrected-to-normal, vision and performed normally on five subtests of the Birmingham Object Recognition Battery (Riddoch & Humphreys, 1993). Participants reported they had no neurodevelopmental or neurological condition, learning difficulty (other than mild dyslexia), psychiatric illness, acquired prosopagnosia, or a history of brain injury or any concussion during the preceding 12 months. Participants were recruited though media coverage, social media, personal networks, and prosopagnosia support groups and offered a £10 gift voucher to recompense them for their time.

Results from four individual face memory measures that were used to classify this sample have been previously reported (Lowes et al., 2024a) and are summarized in Table 3. Participants were classified as DP if they reported lifelong subjective face recognition difficulties, scored ≥ 61 on the PI-20 questionnaire (Shah, Gaule, Sowden, et al., 2015) or an adapted version of the PI20 for parents of adolescents, and showed objective face memory impairments on a global face memory score which was calculated by computing an average age-standardized balanced integration score (BIS) from (1) Old New faces task (Dalrymple et al., 2014); (2) Cambridge Face Memory Test (CFMT; Duchaine & Nakayama, 2006); (3) a CMT difference score calculated by subtracting CFMT scores from Cambridge Bicycle Memory scores (Dalrymple et al., 2014); and (4) Famous Face Test accuracy (Lowes et al., 2024a).

Notably, 19/20 of the DPs reported here showed major impairment (≤ −1.7 SD) on two *separate* objective face memory measures. The z-scores for the final

Table 3. DP z-scores on tests used for classification.

ID	Age	PI-20	Global face memory (BIS)	Old New Faces (BIS)	FFT	CFMT (BIS)	CBMT (BIS)	CMT difference (BIS)	No. of individual face memory measures showing impairment (out of 4)
AF019	68	−7.75	**−3.90**	**−6.92**	**−5.22**	**−1.72**	0.02	**−1.74**	4
AF002	25	−7.28	**−2.24**	**−4.35**	0.41	**−2.70**	−0.37	**−2.33**	3
AF016	23	−7.01	**−2.03**	**−2.11**	**−2.63**	−1.52	0.35	**−1.87**	4
AF017	29	−6.60	−1.51	**−3.17**	−0.04	−0.17	2.50	**−2.67**	2
AF006	54	−6.16	−1.63	**−2.42**	−1.58	−1.16	0.18	−1.34	4
AF003	55	−6.05	−1.66	−0.60	−1.52	**−2.25**	0.04	**−2.29**	3
AF018	49	−5.72	**−2.40**	−1.17	**−4.30**	**−2.88**	**−1.65**	−1.23	4
CF059	16	−5.59	**−2.85**	NA	−0.58	**−5.53**	**−3.07**	**−2.46**	2 of 3
AF009	42	−5.17	**−2.72**	0.03	−0.63	**−5.56**	−0.85	**−4.71**	2
AF010	53	−5.17	**−5.65**	**−12.13**	**−2.63**	**−3.48**	0.89	**−4.37**	4
AF007	31	−4.95	**−2.76**	**−2.88**	0.34	**−4.67**	−0.85	**−3.82**	3
AF008	71	−4.78	**−1.84**	**−5.72**	−1.37	−1.23	**−2.19**	0.96	3
AF060	61	−4.65	−1.44	**−2.94**	**−1.92**	−0.07	0.77	−0.85	2
AF001	51	−4.62	−1.33	**−2.45**	−0.24	−0.35	1.92	**−2.26**	2
AF004	31	−4.54	**−1.95**	−0.67	**−1.87**	**−2.20**	0.85	**−3.05**	3
AF099	70	−4.38	**−2.40**	**−7.53**	NA	0.22	0.11	0.11	1 of 3
AF075	45	−3.64	**−2.44**	**−3.07**	−1.40	**−3.83**	**−2.37**	−1.46	4
AF021	50	−3.09	**−2.14**	−1.38	**−2.03**	**−2.03**	1.08	**−3.11**	4
AF098	67	−2.76	**−1.82**	**−3.84**	−1.45	−1.02	−0.05	−0.97	3
CF005	17	−0.66	**−2.97**	**−2.27**	**−1.98**	**−2.73**	2.18	**−4.90**	4

Note: FFT = Famous Faces Test; BIS = balanced integration score; CFMT = Cambridge Face Memory Test; CBMT = Cambridge Bicycle Memory Test; CMT difference = (CFMT BIS – CBMT BIS); global face memory = average of Old new faces, FFT, CFMT, CMT difference. Z-scores in bold indicate major impairment of at least − 1.7 SD.

participant (AF006) were −2.42 on Old New Faces and between −1 and −1.7 on *each of* the CFMT, FFT and CMT difference score. This participant was therefore considered to have consistently below average objective face memory performance and classified as DP.

We follow previous DP literature (DeGutis et al., 2013, 2014; DeGutis, Cohan, et al., 2012; Lowes et al., 2024a) in selecting ≤ −1.7 SD as our cut-off. We refer to this as a major impairment to distinguish it from milder impairment of between −1 and −1.69 SD, consistent with our previous work. This cut-off represents performance approximately in the bottom 5% of the population on a single (two-tailed) test and we believe is an appropriate and psychometrically balanced criterion due to the large number of tests in our battery and our use of two-tailed tests (see arguments by Gerlach et al. (2024) and the percentile-based impairment modelling in Grant et al. (2024), both of which support the adoption of a slightly less restrictive threshold than 2 SD when multiple related tests are used, as we do here). The choice of age group broadly followed previous literature (e.g., Bate et al., 2019; Bowles et al., 2009). It was also partially driven by practical consideration as we aimed to have 20 controls in each age group. When deciding which tests to administer to adolescents aged 14–17 years we followed Bate et al. (2015) who administered the CFMT adult version and Cambridge Face Perception Test (CFPT) and Bennetts, Mole, et al. (2017) who administered the CFMT+ to a 15 year old prosopagnosic and 14 year old super recognizer (respectively) and typical controls aged 14 and 15 years.

Four additional potential DPs with subjective difficulties completed the test battery but are *not* included here because they did not consistently show objective impairment on the classification measures (i.e., they did not meet the pre-registered classification criteria of a score ≤ −1 SD on two separate face memory measures). Some controls did not complete all tasks due to participant or computer errors, or dropped out part way through the battery, and because we were interested in patterns of performance across tasks we excluded from analysis participants who failed to complete > 1 of the perception tasks ($n = 5$ controls). This left a total 60 age- and gender-matched typical control participants aged 14–74 years (26 men and 34 women). For this final sample, group comparison analyses are calculated using pairwise deletion[1] to maximize power since the sample sizes in each age group were relatively small. Inspection of data showed that the Missing Completely at Random Assumption was highly plausible, thus the overall effect on results was minimal since the only data missing were for Mooney faces ($n = 1$ control) and CFPT ($n = 2$ controls).

3.2. Overview of tasks

Figure 1 gives an overview of the tasks, the perceptual sub process that each investigates, and the decision participants must make.

3.3. Procedure

Data were collected online using Testable (www.testable.org, Rezlescu et al., 2020). Participants received written task instructions containing links to each task and were instructed to complete the tasks in a fixed order.[2] Full onscreen instructions were also provided. Testing took place over at least three self-paced sessions. Participants were advised to take a break between sessions of at least an hour (adults) or 12 hours for those under 18 years. Participants could take additional breaks at the end of any task if desired.

The DP sample size was determined following previous literature (e.g., Duchaine & Nakayama, 2004; Esins et al., 2016; Huis in 't Veld et al., 2012; Marsh et al., 2019; Schmalzl, Palermo, and Coltheart, 2008; Ulrich et al., 2017) and our control sample size target (20 participants per age group) was informed by McIntosh and Rittmo (2021), who recommend a neuropsychological control sample of 16 participants, but that increasing control sample size above 16 does little to meaningfully increase power.

3.4. Statistical analyses

There were no gender differences in any of the pre-registered outcome variables (all $ps > .05$) and no differences in gender distribution between the DP and control groups ($X^2 = 1.11$, $p = .292$, and all $ps > .247$ in individual age groups) so we collapsed data across genders for analysis.

The significance level was set at $p < .05$ (two-tailed) for all measures. To control the False Discovery Rate (FDR) we applied the Benjamini Hochberg (BH) correction (Benjamini & Hochberg, 1995) to the 13 primary outcome measures (Faces among non faces RT, Faces among non faces RT normalized inversion index, Mooney face accuracy, age categorization BIS, gender categorization BIS, Cambridge Face Perception Test (CFPT) accuracy, CFPT normalized inversion index, configural face BIS, configural featural house BIS, configural featural face advantage score, configural featural spacing advantage score, configural featural face inversion index, configural featural house inversion index) and set a false discovery rate of 0.05.

Data were analysed in R (R Core Team, 2021) using the Tidyverse, jmv, and irr packages (Gamer et al., 2019; R Core Team, 2021; Selker et al., 2022; Signorelli, 2025;

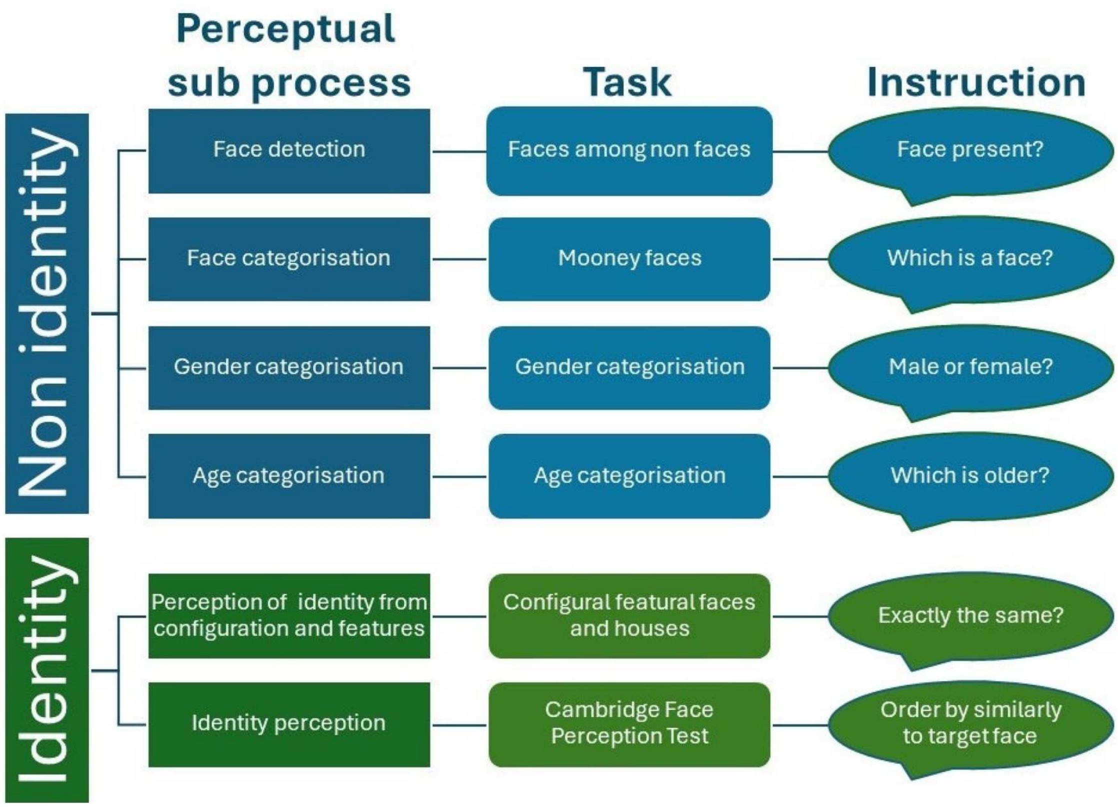

Figure 1. Overview of tasks.

Wickham et al., 2019). Cronbach's alpha was calculated in SPSS 28.0.0.0 (IBM) and Bayesian analyses in JASP (JASP Team, 2023).

3.4.1. Group analyses

We performed Bayesian independent samples *t*-tests on the standardized scores using the default Cauchy prior with a scale of 0.707 and Welch's *t*-tests which are recommended in independent subject designs with different experimental group sizes and/or unequal group variance (Ruxton, 2006) and are more conservative than student's *t*-tests. Where necessary, we used the non-parametric alternative Wilcoxon Rank Sum tests and Bayesian Mann Whitney *U* tests with five chains of 1000 iterations as non-parametric alternatives. As in our previous work, individual scores were converted to z-scores centred on age group control means. We used Fisher's exact tests to compare the proportion of each group showing impairment on the global face perception measure and calculated confidence intervals for the proportion of each group showing impairment on the individual task using the Clopper Pearson exact interval method.

3.4.2. Individual cases

We were interested in whether the DPs who ranked lowest on various face perception measures would be the same individuals who ranked lowest on the four face memory measures we used to classify the DP group, in other words was the lowest performing DP on one task also ranked the lowest on others? To answer this, we estimated Intraclass Correlation Coefficients (ICC) and their 95% confidence intervals based on mean rating using *consistency* and a two-way mixed effects model (model type C1). We calculated ICCs separately for the DP and control groups using individual z-scores from each upright face perception task (Mooney face accuracy, Faces among non faces BIS, gender categorization BIS, age categorization BIS, Configural featural faces BIS, CFPT upright accuracy) as well as the four face memory measures used to classify the DP group (Old New Faces BIS, CFMT BIS, CMT difference BIS and Famous Faces Test accuracy, Lowes et al., 2024a). We then compared the correlations by checking for overlapping confidence intervals. Overlapping confidence intervals suggest that there is no meaningful difference between groups. We use Koo and Li's (2016) guidelines to interpret the strength of the effect, where less than .5 is poor, between .5 and .75 is moderate, between .75 and .9 is good and greater than .9 is excellent. The ICC results using the original accuracy (or RT) measures are shown in Supplementary Table S3.

At the individual participant level, we performed single case analyses using the SingleBayes_ES.exe computer program for Bayesian tests of deficit (BTD), or the

Table 4. Test reliability.

Cronbach's alpha	Accuracy			RT		
	Full sample	Control	DP	Full sample	Control	DP
Mooney faces	.91	.91	.91			
Faces among non faces				.92	.88	.95
Gender categorization	.40	.20	.58	.85	.85	.85
Age categorization	.63	.14	.86	.90	.90	.86
Configural featural	.85	.86	.83			
CFPT	.72	.69	.68			

Note: CFPT = Cambridge Face Perception Test.

DiffBayes_ES.exe program for Bayesian standardized difference test (BSDT) (Crawford et al., 2010; Crawford & Garthwaite, 2007) with an alpha level of .05. Formal single case analyses are reported in full in the Supplementary materials.

3.4.3. Test reliability

Internal reliability (Table 4) for most tests was good (> .8), apart from the CFPT, which was adequate. Age and gender categorization accuracy scores were close to ceiling, which constrained variability and thus Cronbach's alpha for these accuracy measures are quite low. Internal reliability for RT was good or excellent (Cronbach's alpha RT > .85 on each task) so we also calculated BIS which adjusts accuracy to take account of RT.

4. Research questions and results

To foreshadow the detailed results, we provide an overview of the results of each hypothesis in Table 5. A summary at individual case level is provided at the end of the results section in Table 15.

4.1. Face detection and categorization

4.1.1. Motivation

Face detection is the earliest stage of face processing and is thought to be relatively automatic in typical perceivers. For efficient face recognition to take place, presumably individuals must first be able to rapidly *detect* the presence of a face within a scene or among other objects, and also *categorize* a face as a face (i.e., decide if a given stimulus is a face or not) in order to recruit the specialist face recognition system. A deficit at these very early stages of face processing would suggest an apperceptive subtype of DP whereby individuals have difficulty encoding, as well as recognizing, faces.

We tested detection using the Faces among non faces task (Garrido et al., 2008) since they found group level RT impairments on this task in adult DPs. Further, a child study using this task (Dalrymple & Duchaine, 2016) reported that three of seven children with DP showed impaired face detection. Together, these findings suggest that some, but perhaps not all, DPs are impaired at face detection. Mooney faces were chosen as these lack obvious featural cues that DPs may rely on for face identification, thus theoretically forcing participants to rely on configural or holistic processing.

4.1.2. Procedure

4.1.2.1. Faces among non faces task. For a full description of methods, please see Garrido et al. (2008). Briefly, participants were shown a 5 × 5 grid of images, pre-presented for 8 s. A single face is present in 25/37 upright and 25/37 inverted trials. The remaining trials contain no face. The task is to indicate by key press as quickly as possible when a face is present. DPs were predicted to take longer to detect the presence of a face and to show smaller inversion effects versus controls. Inversion effects were calculated as follows [RT on correct upright face-present trials *minus* RT correct inverted face-present trials].

4.1.2.2. Mooney faces. We administered a shorter 10-item version of the original task (see Supplementary materials) which is described in detail in Verhallen and

Table 5. Summary of hypothesis testing results at group level.

Hypotheses: Compared with age matched typical controls	Data support hypothesis?	Cohen's *d*	*rbb*	*p*	BF_{10}	% of DPs impaired
1a Face detection: DP will be slower to detect faces	**Yes**		.40	**.008****	1.88	50%
1b Face detection: DPs will show reduced normalized face inversion effects	**Yes**		.35	**.020***	1.08	40%
1c Mooney faces: DPs will be less accurate	No		.28	.062	0.60	15%
2 Gender categorization: DPs will be less accurate (BIS)	**Yes**		.37	**.015***	1.49	40%
3 Age categorization: DPs will be less accurate (BIS)	No			.129	0.71	35%
4a Configural featural: DPs will be *more* impaired at faces than houses	**Yes**	0.67		**.005****	2.49	
4b Configural featural: DPs will be *more* impaired at spacing than features	No	−0.22		.430	0.38	
4c Configural featural: DPs will show reduced face inversion effects	No	0.30		.245	0.47	10%
5a Face perception ability is related to face memory ability	No (ICC < .21)					

Note: *$p < .05$, ** $p < .01$, *** $p < .001$. Effect sizes are expressed as Cohen's *d* (*d*), or Glass's rank biserial correlation (*rbb*) for non-parametric tests, with their Bayesian equivalents (BF_{10}). The Benjamini Hochberg (BH) method was used to correct for multiple comparisons across 13 measures with the false discovery rate set at 0.05, *p* values in bold remained significant after BH adjustment (Supplementary Table S2). The gender and age categorization outcome measures are Balanced Integration Score due to accuracy ceiling effects. Impaired = $Z \leq -1$.

Table 6. Faces among non faces descriptive statistics.

	DP				Control		
	N	Z	M	SD	*N*	M	SD
RT (ms)							
14–35 years	7	0.77	1106	502	21	936	222
36–59 years	8	1.34	1544	786	21	1136	304
60–74 years	5	1.34	1795	351	17	1308	362
Total	20	1.14	1453	642	59	1114	328
Balanced Integration Score							
14–35 years	7		−1.86	5.68	21	0.00	1.65
36–59 years	8		−1.12	2.59	21	0.00	1.73
60–74 years	5		−2.75	2.84	17	0.00	1.42
Total	20		−1.79	3.84	59	0.00	1.59
RT Inversion effect (ms)							
14–35 years	7		−750	549	21	−815	332
36–59 years	8		−668	411	21	−929	326
60–74 years	5		−710	639	17	−932	452
Total	20		−707	495	59	−888	367

Note: Mean RT in ms calculated on correct, upright target-present trials. Because BIS is a standardized measure, BIS scores for controls are 0 or very close to 0. Inversion effect RT is a difference score calculated by subtraction [mean RT (ms) upright trials – mean RT (ms) inverted trials] with negative values indicating an inversion effect.

Mollon (2016). Participants had to indicate which of three simultaneously presented images contained a face by clicking on the eye area of the face as quickly as possible. Stimuli remained on screen until a response was made. The outcome measure was accurate identification and chance is 33%. Note that one control participant's data were highly atypical showing higher accuracy and faster RT for inverted versus upright trials; because this is the opposite of the pattern robustly observed on face perception tasks among typical perceivers we removed this outlier's data from the analysis.

4.1.3. Results

Descriptive statistics for the faces among non faces task are shown in Table 6.

4.1.3.1. 1a: Faces among non faces (upright faces).

The DP group was significantly slower to detect the presence of upright faces than the age-matched control group, $W = 826$, $p = .008$, which remained significant after Benjamini Hochberg adjustments (Figure 2). Bayesian Mann Whitney tests provided anecdotal evidence for a true between-group difference, $BF_{10} = 1.88$.

To account for possible differential speed accuracy trade-offs (SATs) that might otherwise mask accuracy impairment (see Section 2), we calculated BIS. A Wilcoxon rank sum test with continuity correction showed that the DP group produced, on average, significantly lower BIS than the control group, $W = 361$, $p = .010$, with anecdotal evidence for a group difference $BF_{10} = .1.94$.

Taken together, the results supported the hypothesis that DPs would show an impairment for upright face detection. We next analysed inversion effects.

4.1.3.2. 1b: Faces among non faces inversion effects.

Inspection of individual data showed that 19/20 DPs showed the expected face inversion effect, that is they were slower to detect the presence of inverted than upright faces. When the face present within the array was inverted, the DP group showed an attenuated face inversion effect ($M_{diff} = -707$ ms, SD = 494) versus the control group ($M_{diff} = -888$ ms, SD = 367). In other words, although both groups took longer to correctly detect the presence of an inverted compared with upright face (thus showing a FIE), RT increased *less* among DPs than among typical controls.

Due to the observed RT differences between groups in the upright condition (see above), we computed a *normalized* inversion RT index following previous literature (Avidan et al., 2011), calculated as follows: Inversion

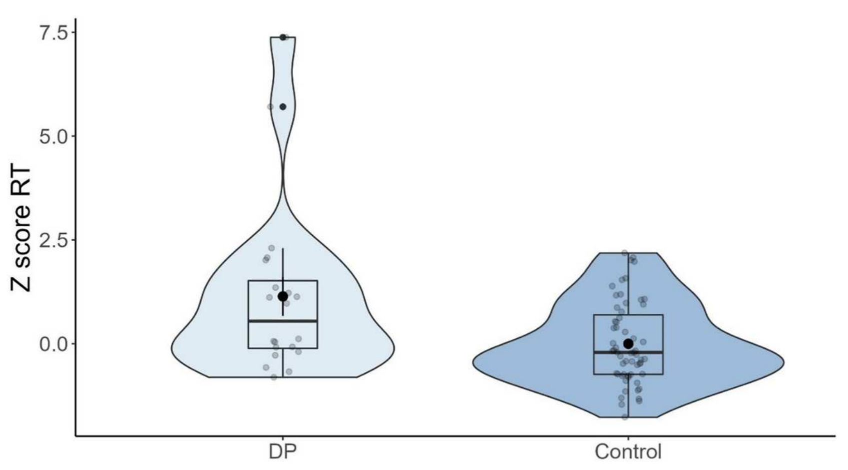

Figure 2. DP group were slower to detect upright faces: Faces among non faces.

Note: Mean RT is calculated on upright correct trials. In this and similar following figures, each dot represents a single participant, the box shows the interquartile range (IQR), and the midline indicates the group median score. The end of each whisker line represents 1.5 × the IQR.

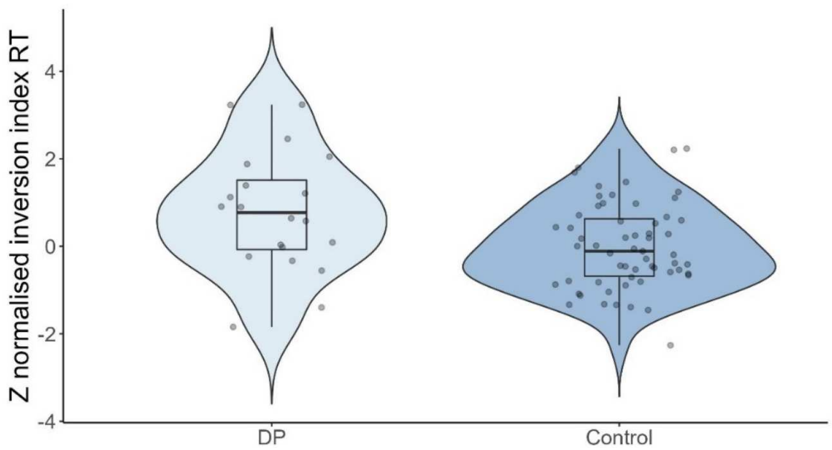

Figure 3. DPs showed significantly reduced face inversion effects: Faces among non faces.

Note: Faces among non Faces Inversion Index = (upright RT – inverted RT) / (upright RT + inverted RT). Scores > 0 indicate a relatively smaller inversion effect versus age-matched controls.

Index = [(upright RT – inverted RT)/(upright RT + inverted RT)] and again standardized this index based on control age group means. The normalized inversion index measures the *proportional* effect of inversion on individuals' performance. Crucially, it accounts for the fact that an identical absolute increase in RT of say 500 ms on inverted versus upright trials would indicate *a relatively larger face inversion effect* for a hypothetical control participant with a mean upright RT of 1000 ms (representing a 50% increase in RT) versus a hypothetical DP participant with a mean upright RT of 2000 ms (representing a relatively smaller increase in RT of 25%). This exploratory inversion index data showed that the DP group's mean RT time was *relatively* less affected by inversion compared with age-matched controls and a Wilcoxon rank sum test with continuity correction indicated this group difference was statistically significant, $W = 797$, $p = .020$, rank biserial correlation = .26 (Figure 3).

For completeness we repeated the inversion effect analysis using the preregistered subtraction method (upright RT *minus* inverted RT). A Wilcoxon rank sum test with continuity correction indicated that the group difference in inversion effect was NS, $W = 753$, $p = .067$ and Bayesian Mann–Whitney tests provided converging weak evidence for the null hypothesis, $BF_{10} = 0.59$.

Table 7. Descriptive statistics Mooney faces.

Proportion correct	DP				Control		
	N	Z	M	SD	*N*	M	SD
14–35 years	7	0.16	.89	0.11	21	.83	0.33
36–59 years	8	−0.52	.78	0.40	21	.90	0.25
60–74 years	5	−2.13	.80	0.34	17	.95	0.07
Total	20	−0.69	.82	0.03	59	.89	0.25

Note: Total number of trials = 10.

In summary, results indicated mixed support for hypothesis 1b. The DP group showed reduced, although not absent, face inversion effects compared with controls. The DP group's face inversion effect was significantly *reduced* versus controls when using the exploratory normalized method but was not significantly different when using the pre-registered subtraction method.

4.1.3.3. 1c: Mooney Faces.

A Wilcoxon rank sum test with continuity correction indicated no significant group accuracy difference (Table 7) for face categorization ($W = 753$, $p = .062$, rank biserial correlation = 0.21) thus we were unable to reject the null hypothesis. A Bayesian Mann Whitney test provided converging anecdotal evidence for the null hypothesis ($BF_{10} = 0.62$). Single case analysis is shown in Supplementary Table S4a and summarized in Table 15.

4.1.4. Discussion of face detection and categorization

The DP group was significantly slower than controls to *detect* a face among objects replicating the original behavioural study (Garrido et al., 2008). This indicates impairment at an early stage of face processing, before identity is perceived.

On the Mooney task, we found no significant group-level difference between DPs and controls, which is consistent with previous work (Towler et al., 2016; Ulrich et al., 2017), albeit these studies used different test versions and outcome measures to ours, thus preventing direct comparison. Note that this evidence for a deficit in the first stage of face processing – face detection – does not imply a deficit early in the whole stream of visual processing. The impairment might derive, for example, from poorer holistic processing (Leong et al., 2024) that would depend on integration of the face percept into a whole.

4.2. Facial gender categorization

4.2.1. Motivation

Accurate face identity recognition presumably requires the target's gender and approximate age (see 4.3 age categorization) be judged quite rapidly. Categorization of facial gender and age does not require any decisions about identity to be made and therefore provides insight into whether lower-level face processing abilities are impaired. Previous literature found that gender judgements can be impaired in DP (Esins et al., 2016; Jones & Tranel, 2001; Marsh et al., 2019) but other studies have reported typical performance (Ariel & Sadeh, 1996; Chatterjee, 2012; DeGutis, Chatterjee, et al., 2012; Duchaine et al., 2006; Le Grand et al., 2006) and this therefore remains an open question.

4.2.2. Procedure

For a full description of methods, see Supplementary Materials 4. Briefly, participants were presented with 30 colour face images on a black background, and the task was to judge whether the face was male or female. Hair was cropped but hairline retained, none of the stimuli showed makeup, facial hair or jewellery. Stimuli were presented one at a time and remained on screen until a response was made. The outcome variables are proportion correct and BIS.

4.2.3. Results

Perhaps due to mean accuracy being close to ceiling in both groups (Table 8), Wilcoxon rank sum tests indicated no significant group difference in accuracy ($p = .696$, $BF_{10} = 0.34$) but showed that the DP group (mean RT 2186 ms, SD 863 ms) was significantly *slower* to make a correct response versus typical controls (mean RT 1669 ms, SD 527 ms), $W = 863$, $p = .004$.

We therefore repeated the analysis using BIS which confirmed a significant group difference, $W = 380$, $p = .015$, rank biserial correlation = .27 with the DP group significantly less accurate at judging facial gender versus age matched controls once RT and differential speed accuracy trade-offs were accounted for using BIS. Bayesian Mann Whitney tests provided converging anecdotal support for a true group BIS difference, $BF_{10} = 1.499$ (Figure 4).

Table 8. Descriptive statistics gender categorization task.

	DP				Control		
	N	Z	M	SD	*N*	M	SD
Proportion correct							
14–35 years	7	−1.78	.94	0.08	21	.98	0.02
36–59 years	8	0.47	.98	0.03	21	.96	0.04
60–74 years	5	−0.97	.92	0.04	18	.96	0.04
Total	20	−0.68	.95	0.06	60	.96	0.03
BIS				Z			
14–35 years	7	−2.80		5.38	21	0.00	1.47
36–59 years	8	−0.53		1.88	21	0.00	1.77
60–74 years	5	−2.26		2.75	19	0.00	1.63
Total	20	−1.76		3.63	60	0.00	1.48

Note: Accuracy scores are unstandardized and chance (proportion correct) = 0.5. Because BIS is a standardized measure, BIS scores for controls are 0, or very close to 0, by definition.

4.2.4. Discussion of gender categorization

Results provided support for hypothesis 2. The DP group was significantly slower than age-matched controls to correctly categorize facial gender. An integrated measure (BIS) showed DPs were significantly more impaired than controls once response time was accounted for. Note that the accuracy celling effects in this task mean that BIS differences are likely to reflect differences in RT rather than providing new information. These findings are in line with Esins et al. (2016), who reported group level ($n = 16$) gender categorization impairment in DP as well as another recent study (Marsh et al., 2019).

4.3. Facial age categorization

4.3.1. Motivation

As with the gender categorization task, we included this facial age categorization task to investigate whether the DP group – or individual DPs – would show impaired performance on a naturalistic face perception task that did not involve face identity recognition. We reasoned that including a second perceptual categorization task could also provide converging evidence of low-level face perception impairment. There is some single case evidence of impaired ability to judge facial age in a child with DP (Ariel & Sadeh, 1996) but also studies reporting no group-level difference between DPs and typical controls (Chatterjee & Nakayama, 2012; Dobel et al., 2007) and thus whether DPs can make accurate age categorization decisions when presented with naturalistic faces remains another open question.

4.3.2. Procedure

For a full description of methods, see Supplementary Materials 5. Briefly, participants had to judge which of two simultaneously presented full colour faces was older and indicate their choice by clicking on the older face. The position of the older face was counterbalanced. Stimuli remained on screen until a response was made. The outcome variables are accuracy and BIS.

4.3.3. Results

Most participants produced accuracy scores at, or close to ceiling, meaning data were not normally distributed (Table 9). A Wilcoxon rank sum test with continuity correction indicated no significant group difference in

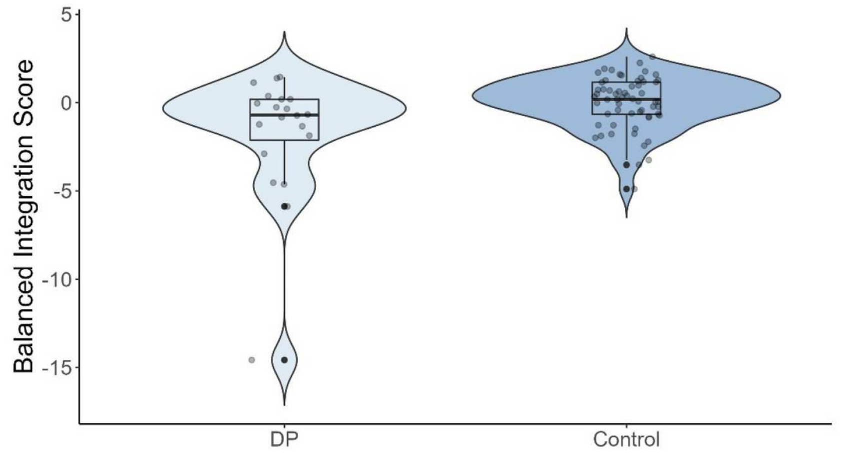

Figure 4. DPs show significantly poorer gender categorization performance.

Note: In both groups (DP and control) mean accuracy was close to ceiling so BIS was calculated ($Z_{accuracy} - Z_{RT}$) to account for DPs' slower RTs and to control for differential speed accuracy trade-offs.

standardized accuracy ($p = .147$) or BIS ($p = .129$). This absence of observed effect is unlikely to be a result of low power since Bayesian Mann Whitney tests (five chains of 1000 iterations) showed converging support for the null hypothesis for both accuracy ($BF_{10} = 0.33$) and BIS ($BF_{10} = 0.58$).

4.3.4. Discussion of age categorization

Contrary to the prediction of hypothesis 3, we found no evidence for a group difference in ability to accurately judge facial age, including when accuracy was adjusted to account for response time and differential speed accuracy trade-offs using BIS. Similar to other studies (Dobel et al., 2007; Nunn et al., 2001), the absence of significant group difference could be due to accuracy ceiling effects as our test was initially designed to be suitable for children. Nevertheless, reduced BIS performance was observed in the DP group ($Z = -1.61$), suggesting reduced DP ability versus controls. As noted in Section 4.2, when accuracy is at ceiling, BIS differences are likely to reflect differences in RT rather than providing new information. Interestingly, two of the three participants who were severely impaired when categorizing age *also* showed a significant impairment when judging facial gender (Table 15) suggesting these individuals have more generalized difficulty encoding early-stage non-identity facial information.

In summary, the DP group performed significantly worse at facial gender discrimination versus controls but not at facial age categorization. Having tested several aspects of non-identity face perception, we next proceeded to test whether participants were able to discriminate changes to features and the spacing between features in a task where the facial identity was held constant.

Table 9. Descriptive statistics age categorization task.

	DP				Control		
	N	Z	M	SD	*N*	M	SD
Proportion correct							
14–35 years	7	−.22	.97	0.08	21	.99	0.02
36–59 years	8	.15	1.00	0.00	21	.99	0.01
60–74 years	5	.09	1.00	0.00	18	.99	0.01
Total	20	.02	.99	0.04	60	.99	0.02
BIS							
14–35 years	7	−1.07	−4.27	8.72	21	0.00	2.05
36–59 years	8	−0.14	−0.49	1.32	21	0..00	1.80
60–74 years	5	0.07	0.34	0.62	18	0.00	1.45
Total	20	−1.61	−1.61	5.37	60	0.00	1.29

Note: Chance accuracy = .5. Because BIS is a standardized measure, BIS scores for controls are 0, or close to 0, by definition.

Table 10. Descriptive statistics: Configural featural task face and house accuracy.

	DP				Control		
	N	Z	M	SD	*N*	M	SD
Faces							
14–35 years	7	−0.83	.66	.12	21	.76	.12
36–59 years	8	−0.44	.70	.08	21	.74	.09
60–74 years	5	−0.35	.60	.10	18	.64	.12
Total	20	−0.56	.66	.10	60	.72	.12
Houses							
14–35 years	7	−0.03	.79	.15	21	.80	.07
36–59 years	8	0.32	.80	.09	21	.77	.10
60–74 years	5	−0.04	.72	.10	18	.72	.11
Total	20	0.11	.78	.12	60	.76	.10
Face advantage							
14–35 years	7	−0.71	−.09	.03	21	−.03	.09
36–59 years	8	−0.53	−.06	.05	21	−.02	.09
60–74 years	5	−0.34	−.09	.08	18	−.06	.09
Total	20	−0.54	−.08	.05	60	−.03	.09

Note: Faces = proportion of upright face trials correct. Houses = proportion of upright house trials correct. Chance accuracy = .5. Face advantage is a normalized *difference* score calculated as [(faces − houses)/(faces + houses)]. Face advantage scores > 0 indicates relatively better performance for *faces* than houses whereas a face advantage score <0 indicates that a participant performed worse on faces *relative to* their performance on houses. Face advantage scores < 0 indicate that the magnitude of the relative impairment for faces is greater than that observed in typical controls.

Table 11. Comparing DP and control group performance: Configural featural task.

	DP M (SD)	Control M (SD)	BF_{10}	*p*	Effect size
Face advantage	−0.54 (0.60)	0.00 (0.98)	2.49	**.005****	*d* = 0.67
Face accuracy	−0.56 (0.87)	0.00 (0.98)	2.21	**.021***	*d* = 0.60
House accuracy	0.11 (1.41)	0.00 (0.98)	0.29	.331	*rbb* = 0.11
Exploratory analyses					
Face BIS	−1.10 (1.39)	0.00 (1.52)	7.62	**.005****	*d* = 0.76
House BIS	−0.21 (1.95)	0.00 (1.57)	0.25	.890	*rbb* = 0.02

Note: **p* < .05, ** *p* < .01, *** *p* < .001.

4.4 Configural and featural processing: Are any deficits observed domain specific or process specific?

4.4.1. Motivation

Given that the most common approach to rehabilitation of DP to date has been to try to improve discrimination of face spacing or features (e.g., Brunsdon et al., 2006; DeGutis et al., 2007; 2014; Schmalzl et al. 2008, 2008understanding whether either, or both, abilities are impaired in DP in general, or in individual DPs, can inform more personalized approaches to rehabilitation.

We used this task to test three competing accounts of the face perception impairments observed in DP (hypotheses 5a–5c above, for a discussion and full methods, see Yovel & Duchaine, 2006). Correctly discriminating changes to the spacing between features tests participants' ability to process second-order spatial relations, i.e., configural processing, which is thought to be essential for efficient face recognition (for a review, see Maurer et al., 2002). It is possible that face recognition difficulties in DP are due to a face-selective configural processing deficit. If inversion effects in the house spacing condition are *smaller* than inversion effects in the face spacing condition (or are absent), this would suggest that face processing mechanisms are domain-specific (i.e., specialized for faces) as opposed to process-specific (i.e., specialized for configural processing more generally), thus supporting the *face-specific holistic hypothesis* (Yovel & Duchaine, 2006).

Table 12. Descriptive statistics: Configural and featural task spacing and features accuracy.

	DP			Control		
	N	M	SD	*N*	M	SD
Spacing						
14–35 years	7	0.73	0.12	21	0.78	0.10
36–59 years	8	0.73	0.08	21	0.74	0.10
60–74 years	5	0.70	0.12	18	0.66	0.13
Total	20	0.72	0.13	60	0.73	0.12
Features						
14–35 years	7	0.72	0.09	21	0.78	0.09
36–59 years	8	0.77	0.10	21	0.78	0.07
60–74 years	5	0.61	0.13	18	0.70	0.10
Total	20	0.71	0.12	60	0.75	0.09
Spacing advantage.						
14–35 years	7	0.00	0.03	21	0.00	0.07
36–59 years	8	−0.02	0.05	21	−0.03	0.07
60–74 years	5	0.07	0.16	18	−0.03	0.09
Total	20	0.01	0.11	60	−0.02	0.08

Note: Spacing = proportion of upright spacing change (face + house) trials correct. Features = proportion of upright features change (face + house) trials correct. Chance accuracy = .5. Spacing advantage is a normalized difference score calculated as [(spacing − features)/(spacing + features)]. A spacing advantage z-score > 0 indicates that a participant showed *relatively* better performance in the spacing versus the feature condition in comparison with a typical age matched control while spacing advantage z-scores < 0 indicates a *relatively* better performance in the feature condition.

4.4.2. Procedure

Full methods are described in Yovel and Duchaine (2006). Briefly, participants were presented with 160 pairs of greyscale images, either faces or houses, and had to decide if the pairs were identical (80 pairs) or different (80 pairs), indicating their answer by keypress. The faces (Ann spacing and Ann Part-S from the original Experiment 2) and houses presented showed the same identity throughout the test but had been manipulated to create differences either in the spacing between features or to features themselves. Due to feedback during piloting, we simplified the instructions (see Supplementary materials) and showed participants examples of the types of changes they might see but using a different facial identity to the one used at test. RT and accuracy data were checked for suboptimal effort, defined as performance at, or below, chance *and* RT < 250 ms on more than 10% of trials. No participant was removed on this basis.

We calculated a face advantage score to measure the extent to which participants are *relatively* more accurate at detecting changes to faces versus houses (a within-participant measure) compared to typical controls of similar age (a between-participant measure). This tests the domain-specific hypotheses and was calculated as follows [(upright spacing accuracy − upright feature accuracy)/(upright spacing accuracy + upright feature accuracy)].

4.4.3. Results

4.4.3.1. 4a Face advantage score. Face advantage scores were normally distributed (Shapiro wilk *p* = .964) and descriptive statistics for the component and derived scores are provided in Table 10. At group level, DPs were significantly impaired versus controls in the face condition, but not in the house condition.

Analysis of face advantage scores (Table 11) provided support for a *domain-specific* account of DP. On average, the individuals in the DP group showed significantly reduced ability to discriminate faces compared to houses, relative to age-matched controls (including after Benjamíni Hochberg adjustments were applied for multiple comparisons). Bayesian analysis provided

Table 13. Descriptive statistics Cambridge face perception test.

	DP				Control			
	N	Z	M	SD	N	Z	M	SD
Upright accuracy								
14–35 years	7	−1.38	0.63	0.07	20	0.00	0.74	0.08
36–59 years	8	−1.65	0.61	0.11	21	0.00	0.75	0.09
60–74 years	5	−0.25	0.64	0.08	17	0.00	0.68	0.11
Total	20	−1.20	0.63	0.08	58	0.00	0.73	0.10
Inversion index								
14–35 years	7	0.81	−0.17	0.07	20	0.00	−0.30	0.17
36–59 years	8	1.51	−0.08	0.12	21	0.00	−0.30	0.15
60–74 years	5	−0.45	−0.23	0.10	17	0.00	−0.17	0.15
Total	20	0.77	−0.15	0.11	58	0.00	−0.26	0.16

Note: Upright accuracy = [(maximum errors − participant error)/maximum errors]; inversion index = [(upright errors − inverted errors) / (upright + inverted errors)]. Z-scores are centred on control age group means to allow comparison with typical controls of similar age. A positive z-score for the inversion index means less of an inversion effect than controls.

converging anecdotal support for true group differences in face discrimination ability and moderate evidence for *no* group difference when discriminating upright houses. Together, the component and derived scores provide converging evidence for a *domain-specific deficit* in the DP group affecting faces but not houses. Individual level data is provided in Supplementary Table S4b and summarized in Table 15.

All scores are z-scores centred on control age group means; therefore, by definition, control group mean scores will be zero. Face advantage = [(upright face accuracy − upright house accuracy)/(upright face accuracy + upright house accuracy)]. BIS = Balanced Integration Score ($Z_{accuracy}$ − Z_{RT}) using upright trials only.

Group differences in standardized scores were calculated using Welch's independent samples t tests or Wilcoxon rank sum tests Effect Sizes: *d* = Cohen's d, *rbb* = rank biserial correlation. *P* values in bold remained significant after applying Benjamini Hochberg (BH) correction with a False Discovery Rate = .05, number of measures corrected for = 13. BH adjustments were applied only to one measure per hypothesis, so BIS scores were not included.

4.4.3.2. 4b spacing advantage score. Results (Table 12) did not show support for the process-specific hypothesis. 5b. No significant difference in spacing advantage was observed between DPs and controls, *W* = 513.5 *p* = .337, rank biserial correlation (rbb) = 0.11. A Bayesian Mann Whitney test showed moderate converging support for the null hypothesis (BF_{10} = 0.29). At individual case level (see Supplementary Table S4b), we again found no support for hypothesis 5b. None of the 20 DPs showed *only* a general spacing (configural) impairment. By contrast, a minority of DPs (3/20) showed significant *featural* impairment. Only one participant (DP) showed a significant dissociation; they performed significantly poorer when discriminating feature changes than spacing changes.

4.4.3.3. 4c inversion effects (domain-specific holistic processing hypothesis 5c).. Data did *not* show support for the domain-specific holistic processing account since groups showed similar standardized inversion effects for both faces *and* houses.

In the face condition, the DP and control groups *both* produced less accurate performance on inverted compared with upright trials and an independent samples *t*-test showed no significant between-group difference, *t*(33.3 = 1.18), *p* = .245 Cohen's *d* = 0.30. Bayesian independent *t*-tests provided converging anecdotal evidence for the null hypothesis, BF_{10} = 0.47.

Table 14. Intraclass correlation estimates.

		95% confidence interval		F test with true value 0			
	Intraclass Correlation (model type)	Lower Bound	Upper Bound	Value	*df*1	*df*2	*p*
Pre-registered	0.105 (C,1)	0.016	0.285	20.17	17	153	.007
DP group *all* individual measures							
Control group *all* individual measures	0.209 (C,1)	0.130	0.321	30.65	50	450	<.001
Exploratory							
DP group individual face *perception* measures only	0.343 (C,1)	0.167	0.577	40.13	19	95	<.001
Control group individual face *perception* measures only	0.209 (C,1)	0.114	0.334	20.59	56	280	<.001
DP group *average* face perception & face memory	0.373 (C,2)	−0.584	0.742	10.59	19	19	.159
Control group *average* face perception & face memory	0.749 (C,2)	0.580	0.850	30.98	59	59	<.001

Intraclass correlation (consistency) is a reliability measure that indicates whether participants' performance ranking is similar across different face processing tasks. For single measures we input all the individual face memory measures (CFMT BIS, Old New face BIS, FFT accuracy, CMT difference BIS) and face perception measures (Faces among non faces RT, Mooney faces accuracy, age categorization BIS, gender categorization BIS, configural/featural face BIS, CFPT upright accuracy).

Table 15. DP z-scores.

ID	Mooney	Faces among non faces		Age categorization		Gender categorization		Configural featural faces		CFPT	Global face perception		Global face memory	No. perception tests impaired	
	Acc.	RT	BIS	Acc.	BIS	Acc.	BIS	Acc.	BIS	Acc.	Acc.	BIS	BIS	Major	Mild
CF059	0.18	5.91	−15.3	−8.79	−22.4	−9.94	−14.7	−2.80	−4.10	−2.49	−4.96	−9.81	−2.85	5	0
AF098	−10.49	2.14	−2.57	0.43	0.97	−1.46	−0.73	−1.63	−3.03	−0.80	−2.68	−2.78	−1.82	3	1
AF009	0.38	2.13	−1.92	0.55	−2.47	−1.50	−4.62	−0.48	−1.35	−0.67	−0.64	−1.78	−2.72	3	1
AF010	−2.44	7.49	−7.28	0.55	−1.34	0.25	−0.83	0.06	−0.11	−1.30	−1.73	−2.22	−5.65	2	2
AF007	0.18	0.03	0.18	0.49	−3.70	−0.43	−1.82	0.13	−3.92	−0.63	−0.05	−1.62	−2.76	3	0
AF018	−3.65	0.15	0.06	0.55	−1.13	1.12	−0.26	−1.83	−2.03	−0.99	−0.83	−1.33	−2.40	3	0
AF017	0.18	−0.88	1.09	0.49	1.32	−3.60	−2.90	−1.13	−0.43	−1.93	−0.85	−0.45	−1.51	2	2
AF019	−0.74	1.86	−1.59	0.43	0.30	−0.64	−5.88	−0.78	−0.46	−1.05	−0.77	−1.57	−3.90	2	1
AF006	0.38	−0.54	0.75	0.55	0.84	0.25	0.19	−1.56	−2.22	−4.10	−0.66	−0.69	−1.63	2	0
AF008	−0.74	1.10	−0.82	0.43	−0.23	−2.29	−4.53	0.49	−0.84	0.32	−0.48	−1.14	−1.84	1	1
CF005	−0.13	0.01	0.20	0.49	−0.59	−0.43	−1.29	−1.13	0.05	−2.49	−0.62	−0.71	−2.97	1	2
AF016	−0.44	−0.73	0.94	0.49	−0.43	1.15	1.47	−0.71	−0.40	−1.75	−0.09	−0.10	−2.03	1	0
AF001	0.38	−0.24	0.46	0.55	1.08	1.12	1.13	0.06	0.05	−2.70	−0.06	0.07	−1.33	1	0
AF075	0.38	1.16	−0.95	0.55	−1.58	0.25	−1.23	−0.21	−0.84	−1.14	−0.22	−0.89	−2.44	0	4
AF004	0.50	1.14	−0.93	0.49	−1.08	−0.43	−0.62	−0.50	−0.61	−0.07	−0.19	−0.47	−1.95	0	2
AF003	0.38	1.02	−0.80	0.55	−0.08	1.12	−0.04	−0.21	−0.25	−0.99	−0.03	−0.30	−1.66	0	1
AF021	−0.02	−0.04	0.25	0.55	0.74	1.12	1.45	0.61	0.02	−1.30	0.17	0.19	−2.14	0	1
AF002	0.50	−0.12	0.33	0.49	−0.19	1.15	0.45	0.13	−1.28	−0.63	0.29	−0.14	−2.24	0	1
AF060	0.66	1.22	−1.65	0.43	−0.31	0.18	0.19	0.28	0.59	0.57	0.15	0.01	−1.44	0	1
AF099	0.66	−0.27	0.54	0.43	0.97	−0.64	−0.35	−0.14	−0.04	−0.30	0.05	0.25	−2.40	0	0
% DPs major impairment [a]	**15 [3,38]**	**25 [9,49]**	**20 [6,44]**	**5 [0, 25]**	**15 [3,38]**	**15 [3,38]**	**30 [12,54]**	**10 [1,32]**	**25 [9,49]**	**30 [12,54]**	**15 [3,38]**	**20 [6,44]**			
Controls	10 [4,21]	7 [2,16]	8 [3,18]	17 [8,29]	10 [4,21]	7 [2,16]	15 [11,32]	5 [1,14]	18 [10,30]	5 [1, 14]	0 [0,6]	3 [0,12]			
% DPs any impairment [a]	**15 [3,38]**	**50 [27,73]**	**30 [12,54]**	**5 [0,25]**	**35 [15,59]**	**25 [9,49]**	**40 [19,64]**	**30 [12,54]**	**35 [15,59]**	**50 [27,73]**	**15 [3,38]**	**40 [19,64]**			
Controls	14 [6,25]	20 [11,32]	13 [6,25]	20 [11,32]	2 [12,34]	13 [6,25]	20 [11,32]	17 [8/29]	23 [13,36]	15 [7,27]	3 [0,16]	8 [3,18]			

Note: [a] showing 95% upper and lower confidence intervals. Acc = accuracy. BIS = Balanced Integration Score Major impairment $z \leq -1.7$ shown in dark grey; mild impairment $-1.69 \leq z \leq -1$ shown in light grey. Broadly, scores in dark grey would be considered significantly impaired when using formal single case methods (see Supplementary Tables S4a–S4e). All scores are standardizd and centred on age-matched control means. A positive RT z-score indicates below average performance. Global face perception is the average of the individual perception measures; when computing this global score RT z-scores were multiplied by −1 so that $z < 0$ always indicates below average performance. Global face memory z-scores (average of 4 BIS measures) were used, alongside self-report, to classify participants as DP or typical control.

In the house condition, the DP (M = -0.05, SD = 1.03) and control groups (M < .001, SD = 0.98) showed similar house inversion effects, and an independent t-test confirmed the group difference was not significant, $t(31.4) = 0.19$, $p = .847$ Cohen's $d = .05$. Bayesian independent samples t-tests showed converging moderate support for the null hypothesis, $BF_{10} = 0.27$. Face and house inversion data were normally distributed (Shapiro Wilk $ps > .93$).

Taken together, results indicate that at group level DPs showed broadly typical inversion effects for faces and houses. Single case dissociation analyses (Crawford et al., 2010; Crawford & Garthwaite, 2005) indicated that face and house inversion effects did not dissociate in the DP group.

4.4.4. Discussion of configural featural processing

We found support only for the *domain-specific* account of DP. The overall group accuracy scores in this sample were very similar to those reported by Yovel and Duchaine (2006) for both faces and houses and, consistent with their findings, we observed a significant group difference for face discrimination but no significant difference on the matched house task (though see Gerlach & Starrfelt, 2021 who used a different stimulus set). Our findings support the view that DP does not result from overall poorer configural processing but instead is face-selective.

Yovel and Duchaine (2006) did not report inversion effects among DPs but our finding that DPs did not show significantly reduced face inversion effects (FIE) compared with typical controls is consistent with previous work (Dalrymple et al., 2014; Klargaard et al., 2018). Others, however, have reported reduced, or absent, FIE in DPs (Avidan et al., 2011; Russell et al., 2009; Shah, Gaule, Gaigg, et al., 2015). The face inversion effect (Yin, 1969) has been widely replicated but it has recently been suggested, using tasks not employed in the present study, that FIE may be task-dependent among typical perceivers (Gerlach & Mogensen, 2024). If true, this task specificity may help account for the variability in outcomes reported across the literature.

4.5. Is face perception ability related to face identity recognition?

4.5.1. Motivation

Overall, we were interested in the *consistency* of participants' face processing ability across the ten tasks they completed (six perceptual, four memory). In other words, would the participant with the lowest (or highest) z-score on one face perception task also be the lowest (highest) scoring on an alternative face perception task, or on one of the face memory tasks?

4.5.2. Procedure

To test consistency of performance across the multiple tasks in our test battery, we calculated the intraclass correlation coefficient (ICC, see Section 3.4.2). The ICC results using the original accuracy (or RT) measures are shown in Supplementary Table S3.

We included upright accuracy on the Cambridge Face Perception Test (CFPT) as one of the ICC input measures because it is a very commonly used measure of face perception in DP, despite its relatively low internal consistency, which makes it less than ideal as a sole measure of face perception. Full methods are described in the original paper (Duchaine et al., 2007). Briefly, participants have 60 s to arrange six grayscale faces with morphed identities in correct order of similarity to a target face at the top of the screen. The primary DV is error score (upright trials only); the number of errors (maximum 144) is calculated from the deviation of each image from its correct position. For ease of comparison with other tasks, we converted the error score into an upright accuracy score following previous literature (Dalrymple et al., 2014; Rezlescu et al., 2012), i.e., [(maximum errors − participant errors)/maximum errors]. The secondary outcome measure of interest is the inversion score [(upright errors − inverted errors) / (upright + inverted errors)] following Avidan et al. (2011). A negative score indicates a face inversion effect, i.e., the participant is more accurate when ordering upright than inverted faces (see Table 13). Consistent with previous literature, independent t-tests showed that the DP group was significantly less accurate than controls on upright CFPT trials, $t(30.3) = 4.37$, $p < .001$, $BF_{10} = 31.2$ and this statistical difference survived correction for multiple comparisons.

4.5.3. Results

ICC estimates (Table 14) did not support the hypotheses that an individual's face perception (rank) ability would be related to their face memory ability. Using Koo and Li's (2016) cut-offs to interpret the reliability of the estimates, the correlation estimates were poor (<.5), meaning that there was no reliable relationship between individuals' rank performance across tasks.

The upper bound of the DP group ICC estimate overlapped with the lower bound of the control group estimate in each of the analyses, indicating that the correlations were not significantly different between groups.

Because the ICC estimates across the 10 variables that we input into the preregistered ICC analyses were unreliable, we conducted additional exploratory post-hoc analyses to determine whether there would be greater consistency across the multiple *face perception* measures only i.e., removing the face memory measures. We also calculated the ICC of the preregistered *global* (averaged) *face memory* and *global face perception* scores because taking the average performance across multiple tasks can provide a more robust, less noisy, performance measure. In all cases we computed ICCs separately for the DP and control groups since we were interested to see whether the consistency of ranking would be similar between groups. The exploratory results showed that the correlations, again, did not differ between groups and none of the correlation estimates reached the reliability criterion of 0.5 except for the global face perception and global face memory scores which showed good reliability in the *control group* only.

To summarize, similar inconsistencies in rank performance were observed for both the DP and control groups, in other words the inconsistencies were not specific to the group with poorer face processing ability (DP).

4.5.4. Discussion of the relationship between face perception and face identity recognition ability

The lack of consistency among DPs supports the view that DP is a highly heterogeneous condition. This heterogeneity claim has been made frequently in the literature based on data from behavioural (e.g., Dalrymple et al., 2014), eye tracking (e.g., Wilcockson et al., 2020), and neuroimaging (e.g., Minnebusch et al., 2007) studies but, to our knowledge, ours is the first study to statistically test this claim using such a large number of theoretically driven tasks designed to tap the different stages of face processing outlined in Bruce and Young's (1986) influential model of face processing.

Although the first ICC analyses comprised a large number of measures (10) which might potentially contribute to the observed unreliability, the final analyses comprised only two averaged measures (global face perception and global face memory) yet even this latter ICC estimate would be considered unreliable in the DP group, despite showing moderate-to-good reliability among controls. In other words, within the DP group, ranking high, or low, on global face perception was not related to similar ranking on global face memory. The inconsistency of DP performance across face perception and face memory task can also be seen visually in the table showing individual DP z-scores (Table 15).

Like DPs, typical controls also showed heterogenous performance when the 10 individual measures were input, in line with recent work which used ICC to examine the consistency of face matching performance among typical perceivers and found a similar lack of consistency (Bobak et al., 2023). However, when global (average) face perception and face memory measures were used as input measures, then correlation estimates among typical controls showed good-to-moderate reliability. These findings suggest that, in typical perceivers, there is a relationship between *overall* face perception and face memory ability, but this relationship is not observed in DPs. This finding supports the idea of apperceptive and mnemonic subtypes of DP (Biotti et al., 2019; Stollhoff et al., 2011; Ulrich et al., 2017).

4.6. Do all DPs show face perception impairments?

4.6.1. Motivation

DP studies reporting individual case level data (e.g., Bate et al., 2019; Dalrymple & Duchaine, 2016; Ulrich et al., 2017) found that perceptual deficits are common, though not universal, in DP. Neuropsychology sample sizes are typically small and so group differences can be driven by one or two participants with extreme impairments. The likelihood of detecting significant *group* differences may depend therefore on the make-up of a given sample (Bennetts et al., 2022). Inspection of individual level data can therefore provide useful insights into milder, though common impairments, and elucidate the perceptual heterogeneity in DP. Formal single case and dissociation analyses for all tasks are provided in Supplementary materials.

4.6.2. Results: individual DP cases

Table 15 lists individual DP z-scores from each task and illustrates patterns of upright face perception in the DP group. For comparison this includes the proportion of controls who showed impairment on these measures. Note that we are not suggesting that a mild atypical score on a task indicates impairment in the absence of any other face recognition difficulties; instead, these scores more likely reflect what "typical" levels of inattention (lack of motivation, tiredness, etc.,) may look like. Looking overall at face perception ability indexed by the global face perception score, exploratory analysis using Fisher's exact tests showed that DPs were significantly more likely than controls to show a major perceptual impairment ($p = .014$ and $p = .032$ for the exploratory BIS measure). Repeating these analyses for *any* impairment (mild or major) showed that DPs were significantly more likely to show impairment on the integrated BIS measure ($p = .002$) but not on the original measures (accuracy and RT), likely because of the previously mentioned accuracy ceiling effects on the age and gender tasks.

5. General discussion

Twenty DP participants and 60 age-matched controls completed a battery of perceptual tasks. At group level, DPs showed impairments at early stage, non-identity sub processes of face detection and gender categorization, as well as significantly reduced ability versus controls to discriminate changes to the configuration and parts of faces. We found that a significantly higher proportion of DPs than controls showed major impairment on the global face perception measure (15% vs 0%, respectively) and the global face perception BIS measure (20% vs 3%). Overall, 50% of DPs showed impairment of at least –1 SD on a minimum of three separate perception tasks vs 10% of controls. The remaining 10 DPs (50%) showed face memory difficulties without accompanying perceptual problems. Across the full test battery, high levels of inter-individual and intra-individual variability were observed. Our findings reinforce the need to inspect individual level data since an absence of group difference does not mean that all DPs are unimpaired and reporting only group-level analysis can mask findings with important theoretical implications. Although, as expected (DeGutis et al., 2024), some controls also showed impairment on the tasks, the key finding was that *no* control showed a major global face perceptual impairment.

Across the full test battery, high levels of inter-individual and intra-individual variability were observed. Our findings reinforce the need to inspect individual level data since an absence of group difference does not mean that all DPs are unimpaired and reporting only group-level analysis can mask findings with important theoretical implications.

Previous work has highlighted how bias can be introduced by conservative selection criteria in relation to face memory cut-offs and/or measurement error if using only one perpetual and one memory task (Biotti et al., 2019), however these considerations cannot account for our findings since the study design mitigated both these potential biases.

Our behavioural findings are consistent with ERP research showing that many individuals with DP exhibit atypical early, pre-identity perceptual encoding of faces. Towler et al. 2012, for example, found that, although the face-sensitive N170 component is present, the usual N170 inversion effect is reduced or absent in DP. This suggests an abnormal processing of upright facial structure rather than a purely mnemonic deficit. Additional work by Fisher et al. (2016) demonstrated decreased sensitivity to contrast information from the eye region, and Towler et al. (2016) reported that people with DP are less responsive to the typical spatial arrangement of facial features. Together, these findings point to impairments in DP at early perceptual processing stages.

Stumps et al. (2020) suggest that DP's face recollection difficulties may be due to them encoding faces in a solely perceptual way, unlike typical perceivers who may be simultaneously encoding perceptual and semantic information (e.g., "she looks friendly" or "he looks like my old uncle John") and thus are able to generate a *conceptual* encoding of a face (Schwartz & Yovel, 2016). The Stumps et al. (2020) study used only *identity* perception tasks but our finding that DPs are impaired when encoding semantic, *non-identity* aspects of faces (e.g., gender) is consistent with this notion and suggests that encoding differences occur at an earlier stage of face processing than face identity perception. Further research is needed to understand the precise mechanisms that link impaired perceptual non-identity processing with impaired face identity recognition.

5.1. DP showed impairments in non-identity face perception

Research questions 1–3 investigated the early, non-identity aspects of face perception and, overall, results indicate that early-stage face perception deficits are common in DP. At the group level, the DP group performed significantly worse than controls at face detection (Faces among non faces) and on a facial gender categorization task using an integrated measure (BIS). Using formal single case statistics (see Supplementary Table S4a), a notable finding is that 40% (8/20) of DPs showed significant impairment at the very earliest stages of face perception (detection, face categorization, age and gender categorization)

5.2. Global face perception scores were significantly lower in the DP group

Different labs have different cut-offs for "impairment". When only one or two perceptual tasks are used, any observed perceptual deficit may arguably be at least partially accounted for by the relatively poor psychometric properties of a single test, by measurement error, or by an impairment on one sub process of face perception. To address these potential sources of bias, we administered a range of perceptual tasks, tapping different sub processes of face perception and averaging the six upright face perception measures to create a global upright face perception score. On this global measure, the DP group performed significantly less well than controls, confirming a broad perceptual deficit. Importantly, this group difference was not driven by one or two outliers, perceptual impairments were widespread in the DP group; 8/20 DPs (40%) showed a global face perception BIS impairment of at least 1 SD alongside their face memory deficit. By contrast, only 5/60 (8.3%) of controls showed impairment on this measure, a similar proportion to the 9% of controls with mild perceptual impairments recently reported by DeGutis et al. (2024).

5.3. Perceptual deficits appear to be face-specific in DP

The DP group's perception deficit appears to be face-specific since the DP group was significantly less accurate than controls at discriminating changes to faces but not houses. This difference was driven by the DP group being less accurate at discriminating changes to face parts (but not face spacing). This finding supports a domain-specific account of DP (Yovel & Duchaine, 2006) and argues against a process-specific (configural) processing account. We note, however, that only one of our tests had a non-face equivalent and, given the heterogenous face perception performance by the DPs, it is plausible the same could be true for objects if other object tests were included in the battery. The specificity of face perception impairments in DPs warrants further study (c.f. Geskin & Behrmann, 2018).

5.4. Subtypes of DP

5.4.1. Apperceptive subtype (impaired face perception and face memory)

The largest subgroup of DPs ($n = 13$) showed widespread face perception deficits, affecting early-stage non-identity sub processes, in addition to face memory impairments. A notable finding was that, despite showing widespread impairment on multiple sub processes of face perception, almost half of this subgroup (6/13) was *not* impaired on the CFPT. This highlights the importance of not relying on a single task to assess face perception ability more broadly.

Our results extend previous literature which suggested an apperceptive subtype of DP (Biotti et al., 2019; Stollhoff et al., 2011; Ulrich et al., 2017) since they confirm the existence of a subset of DPs with relatively normal, though below average, face perception that cannot be attributed to performance on a single face perception task (CFPT only, e.g., Bate et al., 2019; Dalrymple et al., 2014) and/or low internal reliability on certain tasks or measures (Bennetts et al., 2024; DeGutis et al., 2024).

5.4.2. Mnemonic subtype (unimpaired perception but impaired face memory)

We also found robust evidence for a mnemonic subtype of DP characterized by a selective face *memory* impairment alongside typical face perception. Using the DSM-5 approach to classifying mild neuropsychological deficits (Sachdev et al., 2014), individual level data show that 7/20 DPs demonstrated objective face *memory* difficulties but *typical face perception*. The finding that around one third of the DP sample had relatively unimpaired face perception is consistent with recent results from a much larger sample of DPs ($n = 110$) using a different perceptual battery which reported that 33% of DPs had *unimpaired* face perception (DeGutis et al., 2024). However, our results contrast with those of Ulrich and colleagues (2017) who reported finding no individuals in their sample of 11 DPs with a purely mnemonic form of DP, possibly due to their somewhat smaller sample size.

One criticism levelled at some other studies estimating the prevalence of face perception difficulties in DPs is that of "double dipping". Importantly, we did not use perceptual scores for classifying DP therefore we can be confident that DPs with a perceptual impairment were not over-represented in the present sample because of the chosen inclusion criteria. To further ensure that we did not introduce an artificial distinction between the perceptual performance of controls and DPs, we adopted a conservative approach and retained control participants who showed some perceptual difficulties, provided their face memory was not impaired. This approach means the present control group mean scores are lower, and between-group effects are thus *smaller* (more conservative) than would have been the case had we excluded control participants who produced low perceptual scores yet reported no subjective face recognition difficulty.

5.4.3. Tentative third subtype

Results showed some very tentative evidence for a subgroup ($n = 4$) of DPs whose perceptual difficulties affect *only* face identity perception, but not earlier non-identity stages of face perception. Further work would be needed to confirm this finding in a larger sample. However, the individual level data clearly show two important things. First that some DPs can show severe perceptual impairment ONLY on the CFPT and, second, that many show major impairments on several other aspects of face perception and yet not the CFPT. One possible explanation is that the CFPT is the task within this battery that most selectively measures face perception, whereas the other perceptual tasks draw, at least in part, on additional cognitive processes. Taken together with the ICC analyses, the results show that it is not possible to generalize performance from one face perception task to another.

Although we propose that these results provide evidence supporting the idea of at least two sub types of DP, consistent with previous ERP studies (e.g., Eimer et al., 2012) and previous behavioural work, we acknowledge that the subgroups identified are small and that behavioural tests are less sensitive than neural markers. As the DSM_5 acknowledges, the distinction between major and mild neurocognitive disorders is "inherently arbitrary, and the disorders exist along a continuum. Precise thresholds are therefore difficult to determine" (American Psychiatric Association, 2013, p. 608). Further research with larger samples is required to confirm these findings.

5.5. Discussion of individual level results

Findings were that 10/20 or 50% of DPs (vs 10% of controls) showed major face perception difficulties scoring < −1.7 SD on at least *three* face perception tasks.

Three striking findings immediately stand out from inspection of the single case data. First, the *inconsistency* of within-participant performance that was previously demonstrated by the low ICC estimates (see Section 4.5) can be clearly seen. Across the six face perception tasks, the median number of tasks on which individual DPs showed impairment (≤ −1 SD) was 3 (range 0–5), indicating a fairly common, though not universal, difficulty encoding a visual percept of a face (control median for comparison = 1). However, the *patterns* of impairment were highly variable between participants. For example, 50% of the DP group showed face detection impairment and 50%, though not the same 50%, were impaired on the CFPT – a quite different task which requires participants to rank the similarity of facial identities to a target. Heterogeneity of face perception abilities in DP has been previously reported in smaller samples (e.g., Dalrymple et al., 2014; Ulrich et al., 2017), and within families of DPs (Lee et al., 2010; Schmalzl, Palermo, & Coltheart, 2008), but the extent to which these findings may generalize to the wider DP population is potentially limited by sampling bias resulting from small sample sizes, or measurement error when a single perceptual task (typically the CFPT) was used to assess the presence, or absence, of face perception difficulties (e.g., Dalrymple et al., 2014). Our findings using both a slightly larger sample size (n = 20) and a greater number of face perception tasks (n = 6) are an advancement on previous work.

Second, although DP is usually thought of as a condition that affects face *identity recognition*, 13/20 DPs showed mild or major impairment at very early stages of face perception – *face detection* and *gender categorization*, indicating that the perceptual difficulties in DP commonly, though not always, arise *before* facial identity is encoded. Face identity perception deficits have previously been reported in DP (Bate et al., 2019; Biotti et al., 2019; Dalrymple et al., 2014; Duchaine & Nakayama, 2004; Stumps et al., 2020) and we extend these findings by showing that DP also commonly affects non-identity aspects of face perception.

Our behavioural findings are consistent with ERP research showing that many individuals with DP exhibit atypical early, pre-identity perceptual encoding of faces. Towler and Eimer (2012), for instance, found that, although the face-sensitive N170 component is present, the usual N170 inversion effect is reduced or absent in DP. This suggests an abnormal processing of upright facial structure rather than a purely mnemonic deficit. Additional work by Fisher et al. (2016) demonstrated decreased sensitivity to contrast information from the eye region, and Towler et al. (2016) reported that people with DP are less responsive to the typical spatial arrangement of facial features. Together, these findings point to region- and feature-specific perceptual impairments at early perceptual processing stages prior to individuation.

Third, a small minority (3/20) of DPs showed typical early-stage face perception and were impaired *only* on the CFPT, suggesting relatively spared non-identity face perception but difficulty with *identity-based* face perception tasks (specifically, the CFPT). An important related finding is that 6/12 DPs who met the classification for face perception impairment described above would *not* typically be considered to be impaired on the CFPT since their scores were within 1 SD of typical controls. This finding has practical relevance for studies using CFPT as the sole face perception task (e.g., Bate et al., 2019; Biotti et al., 2019; Dalrymple et al., 2014) since our results clearly demonstrate it is possible to show impairment on *multiple* aspects of face perception and yet score within typical norms (within 1 SD) on the CFPT.

5.6. Theoretical implications

We found no evidence for general face processing ability (Verhallen et al., 2017) in either the DP or control groups. At the individual level, performance on one task was not associated with a similar rank level of performance on others. The DP and control groups both showed clear intra- and inter-individual differences across tests.

Results highlight that results from a single face perception task (e.g., CFPT) cannot be taken as evidence of broader face perception ability in DP.

Several studies of DP (e.g., Bate et al., 2019; Dalrymple et al., 2014; Nørkær et al., 2023) used the CFPT as the sole face perception measure; from CFPT performance the authors made wider inferences about face perception performance in DP. Yet it remains uncertain whether face perception ability is unidimensional or multi-dimensional. For example, it has been proposed that a single face ability factor *f* may underlie much face processing ability although interestingly not perception of Mooney faces (Verhallen et al., 2017). By contrast, a recent study of face matching ability among typical perceivers (Bobak et al., 2023) found weak correlations and low consistency in performance across different face matching tests, suggesting that reliable generalizations cannot be made even between perceptual tasks that use an identical paradigm (in Bobak et al., 2023, this was simultaneous identity matching). An important implication of this finding is that it may not be possible to generalize performance on a single face perception task to infer ability across other face perception tasks.

A large scale review of 66 imaging studies of the neural correlates of DP (Manippa et al., 2023, p. 22) concluded that "DP is characterised by abnormal face presentations that differ qualitatively from healthy controls". The authors highlighted widespread evidence of reduced grey matter in DP, especially in the posterior Superior Temporal Sulcus (pSTS), middle temporal gyrus (MTG) and fusiform gyrus (but not the occipital face area (OFA)). White matter abnormalities have also been reported in the core face networks, especially around the right Fusiform Face Area (rFFA) and they suggest that reduced connectivity between visual cortex and temporal lobes (OFA and FFA) might be responsible for DPs' reduced activation to faces in the right FFA and OFA that has been observed in functional magnetic resonance imaging (fMRI) studies. Of particular relevance to the questions addressed in the present study are, first, the finding that connectivity between primary and secondary visual cortices and the core face network are significantly reduced in DPs compared with controls (Lohse et al., 2016). This could indicate that faces are not being rapidly detected (something that is thought to be necessary for the face selective processing system to be recruited). Second, Manippa and colleagues point out that different types of alterations have been observed in DPs' brains which could account for perceptual deficits or memory deficits in DP. For example, they conclude that reduced volume of white and grey matter in r-FFA and short-range functional connectively would account for perceptual deficits in DP whereas disrupted functional connectivity between r-FFA and r-pSTS (Zhao et al., 2018) would explain deficits in face memory. However, an important point made in the Manippa et al. review is that results from brain imaging studies are not consistent. The authors suggest that this may be due to small sample sizes – especially in early studies – and the lack of consistent classification criteria across studies, or because of the heterogenous nature of DP. They recommend that future work should test a homogenous sample of DPs who show similar behavioural performance in order to exclude these alternative explanations.

5.7. Practical implications

Recent studies have argued for the inclusion of DPs with milder face memory impairments in DP research since these participants can still show atypical face perception (Burns et al., 2023; DeGutis et al., 2023). We agree since, consistent with their results, our findings demonstrate that the severity of individuals' face memory impairment was not always related to the severity of their face perception impairment. For example, two of the three individuals with the mnemonic subtype of DP (showing *no face perception impairment*) nevertheless showed a *major face memory impairment*. Equally, two individuals with the apperceptive form of DP showed consistent impairment on *multiple* aspects of face perception yet had only "mild" face memory difficulties and thus may not have been classified as DP by some research groups (see Lowes et al., 2024a). We would therefore also argue for the inclusion of DPs with "mild" impairment in research to better characterize the condition since the degree of face perception deficits did not predict the degree of face memory deficit. Indeed, some DPs with unimpaired face perception (on these measures) in fact showed the most substantial face memory deficits.

5.8. Limitations and directions for future research

Accuracy ceiling effects in the age/gender tasks resulted in low Cronbach's alpha scores and likely insufficient sensitivity to detect group differences. However, RT on these tasks showed good internal reliability, and we therefore used BIS as an integrated measure. However, as previously noted, when accuracy is close to ceiling with low SDs – as they were in the age and gender categorization tasks – BIS differences mainly reflect differences in RT and the underlying scores should be interpreted with caution as they may be unstable. As noted above (Section 2), general confidence or motor issues may contribute to slower RT (and thus affect BIS) but do not seem to explain observed group differences since DP and control BIS do not differ on non-face tasks (Lowes et al., 2024b)

We included the CFPT as a measure of face identity perception since this is the perceptual measure most widely used in DP research. However, this task has only moderate internal reliability and was the only *identity* perception task in the battery so the finding of a possible third subtype of DP affecting face memory and face *identity* processing, but not earlier sub processes of face perception, should be interpreted with caution. It is perhaps more accurate at this stage to say that a subset of DP participants showed face memory impairment but the only perceptual task they were impaired on was the CFPT – all other sub processes of face perception tested were within the typical range.

Given the high observed prevalence of early-stage face processing impairments in this cohort, for example face detection and gender categorization of naturalistic faces, it would be useful to attempt to replicate this finding in a larger sample of DPs using more difficult tasks to avoid ceiling effects observed on the gender task used here.

The clear evidence for subtypes of DP alongside more a detailed understanding of the perceptual sub processes that may be impaired, or relatively preserved, at individual DP level presented here can inform future rehabilitation attempts. The heterogeneous results suggest that a one size fits all approach to remedial training may not be optimal and future research could test a more individualized approach to rehabilitation of DP.

5.9. Conclusion

We show that the perceptual difficulties in DP extend well beyond perception of face identity; half of the DP group were impaired at face detection and 40% were impaired at categorizing facial gender, suggesting early-stage perceptual difficulties.

Two subtypes of DP were identified. The largest subgroup showed face perception *and* face memory deficits (the *apperceptive* subtype). A second subgroup showed typical face perception yet a distinct face memory deficit (the *mnemonic* subtype). For the first time our findings provide tentative evidence of a possible third subtype of DP: These DPs showed face perception deficits only on the CPFT, a *face identity perception* task, yet their performance on five non-identity face perception tasks was within the typical range. If we had used CFPT as the sole perpetual task, these participants could, incorrectly, have been classified as having a general face perception impairment.

Additionally, we found no evidence for "general face processing ability" in DP. At the individual level, performance on one task was not associated with a similar rank level of performance on others. The DP and control groups both showed clear intra- and inter-individual differences across tests. Results clearly show that results from a single face perception task (e.g., CFPT) cannot be taken as evidence of broader face perception ability in DP. We therefore recommend the use of multiple tasks assessing different stages of face perception.

Notes

1. Also known as available-case analysis. To aid statistical power, group analyses were run using all available data, even if a participant had missing data on another task.
2. The perceptual test order was Faces among non faces, Cambridge Face Perception Test, age categorization, gender categorization, configural and featural faces and houses, and finally Mooney faces.

Acknowledgments

The authors would like to thank Dr Kirstin Dalrymple, Professor Brad Duchaine, and Professor Galit Yovel for sharing tests and stimuli.

Author contributions

CRediT: **Judith Lowes:** Conceptualization, Data curation, Formal analysis, Funding acquisition, Investigation, Methodology, Project administration, Visualization, Writing – original draft, Writing – review & editing; **Peter J. B. Hancock:** Conceptualization, Supervision, Writing – review & editing; **Benjamin L. Armstrong:** Formal analysis, Writing – review & editing; **Anna K. Bobak:** Conceptualization, Funding acquisition, Supervision, Writing – review & editing.

CRediT author statement

Judith Lowes: Conceptualization, Methodology, Investigation, Data curation, Formal analysis, Writing – Original draft preparation, Writing Reviewing and Editing, Visualization, Funding acquisition, Project administration: **Peter Hancock:** Conceptualization, Supervision; Writing – Reviewing and Editing; **Benjamin Armstrong**: Formal analysis: Single Cases, Writing – Reviewing and Editing **Anna Bobak:** Conceptualization, Supervision, Writing – Reviewing and Editing, Funding Acquisition.

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

This work was supported by an Economic and Social Research Council studentship under grant number ES/P000681/1 and grant from the British Psychological Society Cognitive Section to Judith Lowes and a Leverhulme Trust Early Career Fellowship under grant number ECF-2019-416 to Anna K. Bobak.

Data availability statement

Data are available at https://osf.io/s57zq/

ORCID

Judith Lowes http://orcid.org/0000-0002-1151-5072
Peter J. B. Hancock http://orcid.org/0000-0001-6025-7068
Benjamin L. Armstrong http://orcid.org/0009-0006-4533-4495
Anna K. Bobak http://orcid.org/0000-0002-4100-5807

References

American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders* (5th ed.). https://doi.org/10.1176/appi.books.9780890425596

Ariel, R., & Sadeh, M. (1996). Congenital visual agnosia and prosopagnosia in a child: A case report. *Cortex*, *32*(2), 221–240. https://doi.org/10.1016/S0010-9452(96)80048-7

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

Bate, S., Bennetts, R., Mole, J. A., Ainge, J. A., Gregory, N. J., Bobak, A. K., & Bussunt, A. (2015). Rehabilitation of face-processing skills in an adolescent with prosopagnosia: Evaluation of an online perceptual training programme. *Neuropsychological Rehabilitation*, *25*(5), 733–762. https://doi.org/10.1080/09602011.2014.973886

Bate, S., Bennetts, R. J., Gregory, N., Tree, J. J., Murray, E., Adams, A., Bobak, A. K., Penton, T., Yang, T., & Banissy, M. J. (2019). Objective patterns of face recognition deficits in 165 adults with self-reported developmental prosopagnosia. *Brain Sciences*, *9*(6), 133–133. https://doi.org/10.3390/brainsci9060133

Bate, S., Frowd, C., Bennetts, R., Hasshim, N., Murray, E., Bobak, A. K., Wills, H., & Richards, S. (2018). Applied screening tests for the detection of superior face recognition. *Cognitive Research: Principles and Implications*, *3*(1), 22. https://doi.org/10.1186/s41235-018-0116-5

Bate, S., & Tree, J. J. (2017). The definition and diagnosis of developmental prosopagnosia. *Quarterly Journal of Experimental Psychology*, *70*(2), 193–200. https://doi.org/10.1080/17470218.2016.1195414

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

Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, *57*(1), 289–300. https://doi.org/10.1111/j.2517-6161.1995.tb02031.x

Bennetts, R. J., Gregory, N. J., & Bate, S. (2024). Both identity and non-identity face perception tasks predict developmental prosopagnosia and face recognition ability. *Scientific Reports*, *14*(1), 6626. https://doi.org/10.1038/s41598-024-57176-x

Bennetts, R. J., Gregory, N. J., Tree, J., Di Bernardi Luft, C., Banissy, M. J., Murray, E., Penton, T., & Bate, S. (2022). Face specific inversion effects provide evidence for two subtypes of developmental prosopagnosia. *Neuropsychologia*, *174*, 108332. https://doi.org/10.1016/j.neuropsychologia.2022.108332

Bennetts, R. J., Mole, J., & Bate, S. (2017). Super-recognition in development: A case study of an adolescent with extraordinary face recognition skills. *Cognitive Neuropsychology*, *34*(6), 357–376. https://doi.org/10.1080/02643294.2017.1402755

Bennetts, R. J., Murray, E., Boyce, T., & Bate, S. (2017). Prevalence of face recognition deficits in middle childhood. *Quarterly Journal of Experimental Psychology*, *70*(2), 234–258. https://doi.org/10.1080/17470218.2016.1167924

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

Bobak, A. K., Jones, A. L., Hilker, Z., Mestry, N., Bate, S., & Hancock, P. J. B. (2023). Data-driven studies in face identity processing rely on the quality of the tests and data sets. *Cortex*, *166*, 348–364. https://doi.org/10.1016/j.cortex.2023.05.018

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. *Cognitive Neuropsychology*, *26*(5), 423–455. https://doi.org/10.1080/02643290903343149

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

Brunsdon, R., Coltheart, M., Nickels, L., & Joy, P. (2006). Developmental prosopagnosia: A case analysis and treatment study. *Cognitive Neuropsychology*, *23*(6), 822–840. https://doi.org/10.1080/02643290500441841

Burns, E. J., Gaunt, E., Kidane, B., Hunter, L., & Pulford, J. (2023). A new approach to diagnosing and researching developmental prosopagnosia: Excluded cases are impaired too. *Behavior Research Methods*, *55*(8), 4291–4314. https://doi.org/10.3758/s13428-022-02017-w

Chatterjee, G. (2012). *Perceptual and cognitive processes in developmental prosopagnosia* (Issues 4-B) [ProQuest Information & Learning]. http://ezproxy.stir.ac.uk/login?url=https://search.ebscohost.com/login.aspx?direct=true&db=psyh&AN=2012-99200-568&site=ehost-live

Chatterjee, G., & Nakayama, K. (2012). Normal facial age and gender perception in developmental prosopagnosia. *Cognitive Neuropsychology*, *29*(5-6), 482–502. https://doi.org/10.1080/02643294.2012.756809

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

Crawford, J., & Garthwaite, P. (2005). Testing for suspected impairments and dissociations in single-case studies in neuropsychology: Evaluation of alternatives using Monte Carlo simulations and revised tests for dissociations. *Neuropsychology*, *19*(3), 318–331. https://doi.org/10.1037/0894-4105.19.3.318

Crawford, J. R., & Garthwaite, P. H. (2007). Comparison of a single case to a control or normative sample in neuropsychology: Development of a Bayesian approach. *Cognitive Neuropsychology*, *24*(4), 343–372. https://doi.org/10.1080/02643290701290146

Crawford, J. R., Garthwaite, P. H., & Porter, S. (2010). Point and interval estimates of effect sizes for the case-controls design in neuropsychology: Rationale, methods, implementations, and proposed reporting standards. *Cognitive Neuropsychology*, *27*(3), 245–260. https://doi.org/10.1080/02643294.2010.513967

Dalrymple, K. A., & Duchaine, B. (2016). Impaired face detection may explain some but not all cases of developmental prosopagnosia. *Developmental Science*, *19*(3), 440–451. https://doi.org/10.1111/desc.12311

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., Pancaroglu, R., & Barton, J. (2014). Acquired prosopagnosia: Structural basis and processing impairments. *Frontiers in Bioscience-Elite*, *6*(1), 159–174. https://doi.org/10.2741/E699

DeGutis, J., Bahierathan, K., Barahona, K., Lee, E., Evans, T. C., Shin, H. M., Mishra, M., Likitlersuang, J., & Wilmer, J. B. (2023). What is the prevalence of developmental prosopagnosia? An empirical assessment of different diagnostic cutoffs. *Cortex*, *161*, 51–64. https://doi.org/10.1016/j.cortex.2022.12.014

DeGutis, J., Chatterjee, G., Mercado, R. J., & Nakayama, K. (2012). Face gender recognition in developmental prosopagnosia: Evidence for holistic processing and use of configural information. *Visual Cognition*, *20*(10), 1242–1253. https://doi.org/10.1080/13506285.2012.744788

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. https://doi.org/10.1093/brain/awu062

DeGutis, J., Kirsch, L., Evans, T. C., Fry, R., Lee, D., Mishra, M., & Campbell, A. (2024). Perceptual heterogeneity in developmental prosopagnosia is continuous, not categorical. Cortex. 2024 Jul;176:37-52. https://doi.org/10.1016/j.cortex.2024.03.011

DeGutis, J., Wilmer, J., Mercado, R. J., & Cohan, S. (2013). Using regression to measure holistic face processing reveals a strong link with face recognition ability. *Cognition*, *126*(1), 87–100. https://doi.org/10.1016/j.cognition.2012.09.004

DeGutis, J. M., 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. https://doi.org/10.1162/jocn.2007.19.11.1790

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

Dobel, C., Bölte, J., Aicher, M., & Schweinberger, S. R. (2007). Prosopagnosia without apparent cause: Overview and diagnosis of six cases. *Cortex: A Journal Devoted to the Study of the Nervous System and Behavior*, *43*(6), 718–733. https://doi.org/10.1016/S0010-9452(08)70501-X

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

Duchaine, B., & Nakayama, K. (2006). The Cambridge face memory test: Results for neurologically intact individuals and an investigation of its validity using inverted face stimuli and prosopagnosic participants. *Neuropsychologia*, *44*(4), 576–585. https://doi.org/10.1016/j.neuropsychologia.2005.07.001

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., & Nakayama, K. (2006). Developmental prosopagnosia: A window to content-specific face processing. *Current Opinion in Neurobiology*, *16*(2), 166–173. https://doi.org/10.1016/j.conb.2006.03.003

Duchaine, B. C., Yovel, G., Butterworth, E. J., & Nakayama, K. (2006). Prosopagnosia as an impairment to face-specific mechanisms: Elimination of the alternative hypotheses in a developmental case. *Cognitive Neuropsychology*, *23*(5), 714–747. https://doi.org/10.1080/02643290500441296

Eimer, M., Gosling, A., & Duchaine, B. (2012). Electrophysiological markers of covert face recognition in developmental prosopagnosia. *Brain*, *135*(2), 542–554. https://doi.org/10.1093/brain/awr347

Esins, J., Schultz, J., Stemper, C., Kennerknecht, I., & Bulthoff, I. (2016). Face perception and test reliabilities in congenital prosopagnosia in seven tests. *I-Perception*, *7*(1), 1–37. https://doi.org/10.1177/2041669515625797

Fisher, K., Towler, J., & Eimer, M. (2016). Reduced sensitivity to contrast signals from the eye region in developmental prosopagnosia. *Cortex: A Journal Devoted to the Study of the Nervous System and Behavior*, *81*, 64–78. https://doi.org/10.1016/j.cortex.2016.04.005

Gamer, M., Lemon, J., Fellows, I., & Singh, P. (2019). *Various coefficients of interrater reliability and agreement* [Computer software]. https://www.r-project.org

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

Gerlach, C., & Mogensen, E. (2024). The face inversion effect does not provide a pure measure of holistic face processing. *Behav Res*, *56*, 330–341. https://doi.org/10.3758/s13428-022-02054-5

Gerlach, C., Nørkær, E., & Starrfelt, R. (2024). Class A, class B. Is that the only chemistry? *Cortex*, *177*, 385–388. https://doi.org/10.1016/j.cortex.2024.03.001

Gerlach, C., & Starrfelt, R. (2021). Patterns of perceptual performance in developmental prosopagnosia: An in-depth case series. *Cognitive Neuropsychology*, *38*(1) 27–49. https://doi.org/10.1080/02643294.2020.1869709

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

Grant, J. G., Wisinger, A. M., Abel, H. F., Hunter, J. M., Smith, G. E., & the Alzheimer's Disease Neuroimaging Initiative. (2024). Quick-reference criteria for identifying multivariate cognitive change in older adults with mild cognitive impairment and dementia: An ADNI study. *Journal of the International Neuropsychological Society*, *30*(10), 944–953. https://doi.org/10.1017/S1355617724000407

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

JASP Team. (2023). *JASP* (Version 0.18.0) [Computer software]. https://jasp-stats.org

Jones, R. D., & Tranel, D. (2001). Severe developmental prosopagnosia in a child with superior intellect. *Journal of Clinical and Experimental Neuropsychology*, *23*(3), 265–273. https://doi.org/10.1076/jcen.23.3.265.1183

Kennerknecht, I., Ho, N., & Wong, V. (2008). Prevalence of hereditary prosopagnosia (HPA) in Hong Kong Chinese population. *American Journal of Medical Genetics Part A*, *146A*(22), 2863–2870. https://doi.org/10.1002/ajmg.a.32552

Klargaard, S. K., Starrfelt, R., & Gerlach, C. (2018). Inversion effects for faces and objects in developmental prosopagnosia: A case series analysis. *Neuropsychologia*, *113*, 52–60. https://doi.org/10.1016/j.neuropsychologia.2018.03.026

Koo, T. K., & Li, M. Y. (2016). A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *Journal of Chiropractic Medicine*, *15*(2), 155–163. https://doi.org/10.1016/j.jcm.2016.02.012

Kriegeskorte, N., Simmons, W. K., Bellgowan, P. S., & Baker, C. I. (2009). Circular analysis in systems neuroscience: The dangers of double dipping. *Nature Neuroscience*, *12*(5), 535–540. https://doi.org/10.1038/nn.2303

Lee, Y., Duchaine, B., Wilson, H. R., & Nakayama, K. (2010). Three cases of developmental prosopagnosia from one family: Detailed neuropsychological and psychophysical investigation of face processing. *Cortex*, *46*(8), 949–964. https://doi.org/10.1016/j.cortex.2009.07.012

Le Grand, R., Cooper, P. A., Mondloch, C. J., Lewis, T. L., Sagiv, N., de Gelder, B., & Maurer, D. (2006). What aspects of face processing are impaired in developmental prosopagnosia? *Brain and Cognition*, *61*(2), 139–158. https://doi.org/10.1016/j.bandc.2005.11.005

Leong, B. Q. Z., Hussain Ismail, A. M., Wong, H. K., & Estudillo, A. J. (2024). The heterogeneity of holistic processing profiles in developmental prosopagnosia: Holistic processing is impaired but not absent. *Cognitive Neuropsychology*, *41*(3-4), 129–147. https://doi.org/10.1080/02643294.2024.2371384

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

Liesefeld, H. R., & Janczyk, M. (2022). Same same but different: Subtle but consequential differences between two measures to linearly integrate speed and accuracy (LISAS vs. BIS). *Behavior Research Methods*, *55*(3), 1175–1192. https://doi.org/10.3758/s13428-022-01843-2

Lohse, M., Garrido, L., Driver, J., Dolan, R. J., Duchaine, B. C., & Furl, N. (2016). Effective connectivity from early visual cortex to posterior occipitotemporal face areas supports face selectivity and predicts developmental prosopagnosia. *The Journal of Neuroscience*, *36*(13), 3821–3828. https://doi.org/10.1523/JNEUROSCI.3621-15.2016

Lowes, J., Hancock, P. J. B., & Bobak, A. K. (2024a). A new way of classifying developmental prosopagnosia: Balanced integration score. *Cortex*, *172*, 159–184. https://doi.org/10.1016/j.cortex.2023.12.011

Lowes, J., Hancock, P. J. B., & Bobak, A. K. (2024b). Evidence for different visual processing strategy for non-face stimuli in developmental prosopagnosia. *Visual Cognition*, *32*(1), 1–12. https://doi.org/10.1080/13506285.2024.2359743

Lowes, J., McGregor, L. M., Hancock, P. J. B., Duchaine, B., & Bobak, A. K. (2025). This condition impacts every aspect of my life: A survey to understand the experience of living with developmental prosopagnosia. *PLoS ONE*, *20*(4), e0322469. https://doi.org/10.1371/journal.pone.0322469

Manippa, V., Palmisano, A., Ventura, M., & Rivolta, D. (2023). The neural correlates of developmental prosopagnosia: Twenty-five years on. *Brain Sciences*, *13*(10), 1399. https://doi.org/10.3390/brainsci13101399

Marsh, J. E., Biotti, F., Cook, R., & Gray, K. L. H. (2019). The discrimination of facial sex in developmental prosopagnosia. *Scientific Reports*, *9*(1), 19079. https://doi.org/10.1038/s41598-019-55569-x

Maurer, D., Grand, R. L., & Mondloch, C. J. (2002). The many faces of configural processing. *Trends in Cognitive Sciences*, *6*(6), 255–260. https://doi.org/10.1016/s1364-6613(02)01903-4

McIntosh, R. D., & Rittmo, J. Ö. (2021). Power calculations in single-case neuropsychology: A practical primer. *Cortex*, *135*, 146–158. https://doi.org/10.1016/j.cortex.2020.11.005

Minnebusch, D. A., Suchan, B., Ramon, M., & Daum, I. (2007). Event-related potentials reflect heterogeneity of developmental prosopagnosia. *European Journal of Neuroscience*, *25*(7), 2234–2247. https://doi.org/10.1111/j.1460-9568.2007.05451.x

Nørkær, E., Guðbjörnsdóttir, E., Roest, S. B., Shah, P., Gerlach, C., & Starrfelt, R. (2023). The Danish version of the 20-item prosopagnosia index (PI20): Translation, validation and a link to face perception. *Brain Sciences*, *13*(2), 337. https://doi.org/10.3390/brainsci13020337

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

Ramon, M. (2021). Super-recognizers – A novel diagnostic framework, 70 cases, and guidelines for future work. *Neuropsychologia*, *158*, 107809. https://doi.org/10.1016/j.neuropsychologia.2021.107809

R Core Team. (2021). *R: The R project for statistical computing* [Computer software]. R Foundation for Statistical Computing. https://www.r-project.org/

Rezlescu, C., Danaila, I., Miron, A., & Amariei, C. (2020). Chapter 13 – More time for science: Using Testable to create and share behavioral experiments faster, recruit better participants, and engage students in hands-on research. In B. L. Parkin (Ed.), *Progress in brain research* (Vol. 253, pp. 243–262). Elsevier. https://doi.org/10.1016/bs.pbr.2020.06.005

Rezlescu, C., Pitcher, D., & Duchaine, B. (2012). Acquired prosopagnosia with spared within-class object recognition but impaired recognition of degraded basic-level objects.

Cognitive Neuropsychology, *29*(4), 325–347. https://doi.org/10.1080/02643294.2012.749223

Riddoch, M. J., & Humphreys, G. W. (1993). *Birmingham object recognition battery*. Lawrence Erlbaum Associates.

Robotham, R. J., & Starrfelt, R. (2018). Tests of whole upright face processing in prosopagnosia: A literature review. *Neuropsychologia*, *121*, 106–121. https://doi.org/10.1016/j.neuropsychologia.2018.10.018

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

Ruxton, G. D. (2006). The unequal variance *t*-test is an underused alternative to student's *t*-test and the Mann–Whitney *U* test. *Behavioral Ecology*, *17*(4), 688–690. https://doi.org/10.1093/beheco/ark016

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. *Nature Reviews Neurology*, *10*(11), 634–642. https://doi.org/10.1038/nrneurol.2014.181

Schmalzl, L., Palermo, R., & Coltheart, M. (2008). Cognitive heterogeneity in genetically based prosopagnosia: A family study. *Journal of Neuropsychology*, *2*(1), 99–117. https://doi.org/10.1348/174866407X256554

Schmalzl, L., Palermo, R., Green, M., Brunsdon, R., & Coltheart, M. (2008). Training of familiar face recognition and visual scan paths for faces in a child with congenital prosopagnosia. *Cognitive Neuropsychology*, *25*(5), 704–729. https://doi.org/10.1080/02643290802299350

Schwartz, L., & Yovel, G. (2016). The roles of perceptual and conceptual information in face recognition. *Journal of Experimental Psychology: General*, *145*(11), 1493–1511. https://doi.org/10.1037/xge0000220

Selker, R., Love, J., Dropmann, D., & Moreno, V. (2022). *jmv: The 'jamovi' analyses* (Version 2.3.4) [Computer software]. https://CRAN.R-project.org/package=jmv

Shah, P., Gaule, A., Gaigg, S. B., Bird, G., & Cook, R. (2015). Probing short-term face memory in developmental prosopagnosia. *Cortex*, *64*, 115–122. https://doi.org/10.1016/j.cortex.2014.10.006

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), 140343–140343. https://doi.org/10.1098/rsos.140343

Signorelli, A. (2025). *DescTools: Tools for descriptive statistics* [R]. https://andrisignorell.github.io/DescTools/

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

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

Towler, J., & Eimer, M. (2012). Electrophysiological studies of face processing in developmental prosopagnosia: Neuropsychological and neurodevelopmental perspectives. *Cognitive Neuropsychology*, *29*(5-6), 503–529. https://doi.org/10.1080/02643294.2012.716757

Towler, J., Fisher, K., & Eimer, M. (2017). The cognitive and neural basis of developmental prosopagnosia. *Quarterly Journal of Experimental Psychology*, *70*(2), 316–344. https://doi.org/10.1080/17470218.2016.1165263

Towler, J., Fisher, K., & Eimer, M. (2018). Holistic face perception is impaired in developmental prosopagnosia. *Cortex*, *108*, 112–126. https://doi.org/10.1016/j.cortex.2018.07.019

Towler, J., Gosling, A., Duchaine, B., & Eimer, M. (2012). The face-sensitive N170 component in developmental prosopagnosia. *Neuropsychologia*, *50*(14), 3588–3599. https://doi.org/10.1016/j.neuropsychologia.2012.10.017

Towler, J., Gosling, A., Duchaine, B., & Eimer, M. (2016). Normal perception of Mooney faces in developmental prosopagnosia: Evidence from the N170 component and rapid neural adaptation. *Journal of Neuropsychology*, *10*(1), 15–32. https://doi.org/10.1111/jnp.12054

Ulrich, P. I. N., Wilkinson, D. T., Ferguson, H. J., Smith, L. J., Bindemann, M., Johnston, R. A., & Schmalzl, L. (2017). Perceptual and memorial contributions to developmental prosopagnosia. *Quarterly Journal of Experimental Psychology*, *70*(2), 298–315. https://doi.org/10.1080/17470218.2016.1177101

Verhallen, R. J., Bosten, J. M., Goodbourn, P. T., Lawrance-Owen, A. J., Bargary, G., & Mollon, J. D. (2017). General and specific factors in the processing of faces. *Vision Research, Individual Differences as a Window Into the Structure and Function of the Visual System*, *141*, 217–227. https://doi.org/10.1016/j.visres.2016.12.014

Verhallen, R. J., & Mollon, J. D. (2016). A new Mooney test. *Behavior Research Methods*, *48*(4), 1546–1559. https://doi.org/10.3758/s13428-015-0666-0

Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L., François, R., Grolemund, G., Hayes, A., Henry, L., Hester, J., Kuhn, M., Pedersen, T., Miller, E., Bache, S., Müller, K., Ooms, J., Robinson, D., Seidel, D., Spinu, V., … Yutani, H. (2019). Welcome to the Tidyverse. *Journal of Open Source Software*, *4*(43), 1686. https://doi.org/10.21105/joss.01686

Wilcockson, T. D. W., Burns, E. J., Xia, B., Tree, J., & Crawford, T. J. (2020). Atypically heterogeneous vertical first fixations to faces in a case series of people with developmental prosopagnosia. *Visual Cognition*, *28*(4), 311–323. https://doi.org/10.1080/13506285.2020.1797968

Yin, R. K. (1969). Looking at upside-down faces. *Journal of Experimental Psychology*, *81*(1), 141–145. https://doi.org/10.1037/h0027474

Young, A. W., & Bruce, V. (2011). Understanding person perception. *British Journal of Psychology (London, England : 1953)*, *102*(4), 959–974. https://doi.org/10.1111/j.2044-8295.2011.02045.x

Yovel, G., & Duchaine, B. (2006). Specialized face perception mechanisms extract both part and spacing information: Evidence from developmental prosopagnosia. *Journal of Cognitive Neuroscience*, *18*(4), 580–593. https://doi.org/10.1162/jocn.2006.18.4.580

Zhao, Y., Zhen, Z., Liu, X., Song, Y., & Liu, J. (2018). The neural network for face recognition: Insights from an fMRI study on developmental prosopagnosia. *NeuroImage*, *169*, 151–161. https://doi.org/10.1016/j.neuroimage.2017.12.023