

Objective and subjective experiences of developmental prosopagnosia

Judith Lowes

University of Stirling

Faculty of Natural Sciences, Psychology

Thesis submitted for the degree of Doctor of Philosophy

August 2024

Declaration

I wish to submit the thesis detailed above in accordance with University of Stirling research degree regulations. I declare that thesis embodies the results of my own research and was composed by me. Where appropriate I have acknowledged the nature and extent of the work carried out in collaboration with others included in this thesis.

The ethical approaches of the studies reported in this thesis were approved by the University of Stirling General University Ethics Panel.

Signature:  **Date:** 9 August 2024

Wordcount: Thesis 67,601

This work is licensed under [CC BY-NC-ND 4.0](https://creativecommons.org/licenses/by-nc-nd/4.0/)

Acknowledgements

I would like to express my sincere thanks to my fantastic supervisors Peter Hancock and Anna Bobak. You really have been the dream team, I have learned so much from you about faces, science and what excellent mentorship looks like. Without exception, every single piece of advice and feedback has been constructive, and I have greatly appreciated the freedom you both allowed me to pursue the questions that interested me whilst always ensuring I felt supported. I have genuinely enjoyed the journey so much, largely I think because you have been such amazing supervisors.

Thanks too to past and present members of the Face Lab, for your assistance with so many things. In particular to students Benji Anderson who assisted with the single case analysis for Chapter 5 and Mhairi Webster who help to develop the famous faces test in Chapter 4. I am grateful to Gabriele Chierchia, Kirsten Dalrymple, Brad Duchaine, and Galit Yovel and colleagues for kindly sharing their materials and experimental set up on Gorilla and Testable.

I would particularly like to acknowledge colleagues in the wider Psychology team at Stirling, particularly Lesley McGregor for her wise advice and collaboration on the study on chapter 6, Rachel Crocket for some very helpful early discussions about this study, and the PhD student cohort for your friendship and support.

During my PhD I was fortunate and privileged to visit some leading researchers of developmental prosopagnosia whose brilliant work first sparked my interest in the topic. Thanks to Joe Degutis and his group at Boston VA Healthcare/Harvard, Brad Duchaine and the Social Perception Lab at Dartmouth College and Randi Starrfelt and her team at the University of Copenhagen for generously hosting my visits, for their interest in my work, and for providing the inspiration to pursue this research area. I would also like to thank Meike Ramon for her enthusiastic encouragement to include a qualitative study when we met at ECVP in Nijmegen. Being able to attend many of your excellent online research seminars during Covid times and beyond were a wonderful way to connect with a wider group of face researchers – thank you for making this opportunity available to so many of us.

None of the research would have been possible without the participants and their parents. To everyone who took part, a huge thank you. I know some of the tests were hard! I would also like to thank the Communications Team at University of Stirling and Faceblind UK for helping to advertise the study and recruit participants.

During my PhD I, to my surprise, also became an avid year-round outdoor swimmer in the beautiful rivers and lochs that surround my village. Huge thanks to my swimming buddies Mo, Fiona, Max, Vivienne, Felicity, and Jen for the welcome work life balance and regular energising (freezing) dips and the warm cuppas and chats afterwards. These helped keep me sane (despite the ungodly hour at which you all prefer to swim!) during these last three and a half years.

Finally enormous thanks to my family, Pete, Meg, and Ewan for their endless support and encouragement and for listening patiently to many, many tales about the fascinating (to me) world of developmental prosopagnosia research. Pete, I know that my one year “sabbatical” from work to do a masters in psychology in 2018 turned into rather more than that (!) - I could not have done this without you. Thank you.

Abstract

In this thesis, I investigated objective and subjective experiences of developmental prosopagnosia (DP), a neurodevelopmental condition that results in severe face recognition difficulties.

To assess objective face processing, I first screened for general cognitive deficits that could explain poor face test performance. Participants completed a fluid reasoning task using abstract shapes. Potential DPs were more accurate than controls but significantly *slower*, suggesting *speed-accuracy trade-off* in non-face tasks. To address this, I calculated the Balanced Integration Score (BIS). Results showed no group differences, highlighting that DPs clearly adopted a *different strategy* from controls.

I next investigated which face memory measures best classified DP. Participants completed two face memory tests, the Cambridge Face Memory Test (CFMT) and Old New Faces, a bicycle memory task and a famous faces recognition test. I directly compared how well accuracy and (BIS) classified DP. BIS was a more sensitive measure than accuracy alone on these tasks, classifying 83% vs 58% of DPs respectively and produced larger group effect sizes.

I next systematically tested which subprocesses of face perception were impaired in DP. Single case analysis, group comparisons and intraclass correlations showed that face perception ability was highly heterogeneous. Data showed evidence for two sub types of DP; an *apperceptive* subtype, in which both face perception and face memory are impaired and a *mnemonic* subtype (in which only face memory is impaired).

Finally, I explored subjective experiences of living with DP using a mixed methods survey. Responses provided further evidence of the heterogeneity of the condition and highlighted the types of difficulties that DPs experienced in everyday life and their priorities for future research. Being unable to recognise close friends and family was a common experience and lower CFMT scores were not associated with the ability to recognise friends and family.

Key words: Developmental prosopagnosia, face processing, speed-accuracy trade off, face recognition, face perception, balanced integration score, face recognition impairment

Table of contents

Declaration	II
Acknowledgements	III
Abstract	V
Table of contents	VI
List of figures	XI
List of tables	XII
Glossary of terms.....	XIII
Impact of Covid 19 statement.....	XV
Author contributions and dissemination.....	XVI
1 Chapter 1: Thesis overview and aims.....	1
1.2 Aims of the thesis.....	1
1.2.1 Research questions	1
1.2 Thesis structure	2
2. Chapter 2: General introduction	3
2.1 What is developmental prosopagnosia?	3
2.2 Subjective experiences of living with DP	4
2.3 How do humans recognise faces? Overview of theories of face recognition	5
2.3.1 The Bruce and Young model	5
2.3.2 Neural models	8
2.4 How and when does face recognition typically develop?	11
2.4.1 The two process theory of face recognition	12
2.4.2 Perceptual narrowing hypothesis	13
2.4.3 How does face recognition develop through childhood?	13
2.5 Aetiology and proposed accounts of DP	16
2.6 Which aspects of face processing are impaired in DP?	19
2.6.1 Holistic face processing	19
2.7 Thesis overview and aims	22
3. Chapter 3: Evidence for different visual processing strategy for non-face stimuli in developmental prosopagnosia.....	24
Abstract.....	24
3.1 Background	24
3.1.1 Evidence for different processing strategies for faces	24
3.2 Method	26
3.2.1 Participants.....	26
3.2.2 Materials and procedure.....	29

3.2.3 Analysis and pre-processing	31
3.3 Results	32
3.3.1 Response time	34
3.3.2 Balanced Integration Score (BIS)	35
3.4 Discussion	40
4. Chapter 4: A new way of classifying Developmental Prosopagnosia: Balanced Integration Score (BIS)	43
Recap and present chapter	43
Abstract	43
4.1 Introduction	44
4.1.1 Different approaches to classifying DP	44
4.1.2 Different approaches to the selection of tests for classification	45
4.1.3 Different approaches to the selection of outcome measures to quantify test performance	46
4.1.4 Different approaches to defining inclusion and exclusion criteria	47
4.1.5 The present study	49
4.2 Research transparency and openness	49
4.3 Methods	50
4.3.1 Participants	50
4.3.2 Materials	54
4.3.3 Procedure	60
4.3.4 Analysis plan	61
4.3.5 Statistical analysis	63
4.4 Results	64
4.4.1 Reliability	64
4.4.2 Old New Faces	65
4.4.3 Cambridge Face Memory Test (group and age group)	67
4.4.4 Difference between face (CFMT) and object (CBMT) memory accuracy	69
4.4.5 Famous Faces Test	71
4.4.6. Global face memory measures	74
4.4.7 Which measures best predict group membership?	77
4.4.8 How does using BIS change classification of DP?	79
4.5 Discussion	83
4.5.1 Is self-report a useful initial classification measure for DP?	83
4.5.2 DP performance on Old New Faces	85
4.5.3 DP performance on Famous Faces Test	86
4.5.4 Do accuracy or BIS measures best predict group membership?	86
4.5.5 Use of global scores and choice of cut off	89

4.5.6 Online vs lab-based data	90
4.6 Conclusion	91
5. Chapter 5: An in-depth investigation of face perception in developmental prosopagnosia .	92
5.1 Introduction	92
5.1.1 What is the nature of the impairment in DP? A memory impairment or perceptual impairment, or both?	92
5.1.2 Limitations of the literature to date	95
5.1.3 The present study	96
5.2 Data transparency statement	98
5.3 General methods	99
5.3.1 Participants.....	99
5.3.2 Overview of tasks.....	101
5.3.3 Procedure	102
5.3.4 Statistical analyses	103
5.4 Research questions and results.....	105
5.4.1 Research question 1 Are all, or some, DPs impaired at face detection?	106
5.4.2 Research question 2: Are all, or some, DPs impaired at categorising facial gender?	115
5.4.3 Research question 3: Are some or all DPs impaired at judging facial age?.....	119
5.4.4 Research question 5: Do participants with DP show deficits in configural or featural face processing? Are any deficits observed domain specific or process specific?	122
5.4.5 Research question 6a: Is face perception ability related to face identity recognition? ..	131
5.4.6 Summary of group level results	139
5.4.7 Research question 6b: Do all DPs show face perception impairments?	141
5.6 General discussion	144
5.6.1 DP showed impairments in nonidentity face perception.....	145
5.6.2 Perceptual deficits appear to be face specific in DP	145
5.6.3 Global face perception scores were significantly lower in the DP group	146
5.6.4 Sub types of DP.....	146
5.6.5 Theoretical implications.....	148
5.6.6 Practical implications	148
5.6.7 Limitations	148
5.6.8 Directions for future research.....	149
5.7 Conclusion	149
6. Chapter 6: This condition impacts every aspect of my life: A survey to understand the experience of living with developmental prosopagnosia.....	151
6.1 Introduction.....	151
6.1.1 The present study	154

6.1.2 Research questions	154
6.2 Methods.....	155
6.2.1 Context and sampling strategy	155
6.2.2 Ethical issues pertaining to human subjects.....	156
6.2.3 Data collection methods and instruments	156
6.2.4 Participant characteristics.....	157
6.2.5 Data processing.....	157
6.2.6 Data analysis qualitative approach.....	159
6.2.7 Techniques to enhance trustworthiness.....	160
6.2.8 Researcher characteristics and reflexivity.....	161
6.3 Quantitative results	161
6.3.1 How many individual faces can people with DP typically recognise?	161
6.3.2 Can DPs recognise close friends and family?	161
6.3.3 Relationship between real life experience and test scores	164
6.3.4 What types of face recognition errors do DPs make?	164
6.4 Qualitative analysis of the written narratives of living with DP	166
Theme 1: Seeking explanation, validation, and diagnosis	166
Theme 2: “You don’t know how difficult it is navigating life as a prosopagnosic”	169
Theme 3: Widespread negative psychosocial impact	181
Theme 4: Worrying what others think	185
Theme 5: Appropriate support helps, but is often unavailable	190
6.5 Research priorities.....	193
6.6 Discussion	195
6.6.1 Lived experiences of DP are extremely varied	196
6.6.2 Face perception can be affected	196
6.6.3 Theoretical implications.....	197
6.6.4 Quality of life and workplace experience – a case for DP to be included in the definition of neurodivergence.	198
6.6.5 Limitations and directions for future research	199
7. Chapter 7: General discussion	200
7.1 Summary of chapters	200
7.2 The present chapter	202
7.3 Individuals with DP show metacognitive insight.....	202
7.3.1 DPs have insight into their difficulties, even when unaware of DP as a condition. .. 202	
7.3.2 PI-20 scores predicted objective face memory scores	203
7.4 DP is characterised by the frequency and type of recognition errors.....	206
7.4.1 Inability to recognise immediate family and close friends is common in DP.....	207

7.5 The Underpinnings of DP are heterogeneous	208
7.5.1 Inter and intra individual differences in behavioural patterns of face perception and face memory.....	208
7.5.2 Caution needed when generalising from the CFPT.	209
7.5.3 Inter individual differences in subjective experiences of face perception and face memory.....	210
7.6 The use of group level statistics alone to characterise DP can be misleading	211
7.7 Other neurodevelopmental conditions also show spiky cognitive profiles	212
7.8 Limitations	213
7.9 Future directions	214
7.9.1 Where in development does DP occur?	214
7.9.2 Individualised training	214
7.9.3 Awareness and changes to public policy (priority for this cohort of DPs) has potential to attenuate the difficulties DPs experience	215
7.9.4 Other future directions	215
8.0 Conclusion	215
References	217
8. Appendices	259
Appendix 1: Chapter 3 supplementary materials	259
Appendix 2: Chapter 4 supplementary materials	263
Appendix 3: Chapter 5 supplementary materials	273
Appendix 4: Chapter 6 supplementary materials	283

List of figures

Figure 2.1 The Bruce and Young model of face recognition.....	6
Figure 2.2 Burton, Bruce and Johnston's (1990) Interactive Activation and Competition Model8	
Figure 2.3 The Gobbini and Haxby (2007) Neural Model of Familiar Face Recognition.....	9
Figure 2.4 Duchaine and Yovel's Revised Neural Framework for Face Processing (2015)	11
Figure 2.5 Summary of research design Chapters 4 and 5.....	23
Figure 3.1 Example items from the Matrix Reasoning Item Bank (MaRs-IB).....	29
Figure 3.2 MaRs-IB score distribution by group (DP or control).....	33
Figure 3.3 Comparison of RT difference between the DP and control groups across 20 easier trials and 20 more difficult MaRs-IB trials.....	35
Figure 3.4 MaRs-IB group difference.....	37
Figure 3.5 Group difference CBMT.....	39
Figure 4.1 Overview of recruitment and classification process	55
Figure 4.2 Old New Faces test design.....	57
Figure 4.3 Famous Faces Test design	60
Figure 4.4 Overview of the method used to calculate z scores and BIS	62
Figure 4.5 Group difference (self-reported DPs vs controls) Old New Faces	66
Figure 4.6 CFMT group difference in performance.....	68
Figure 4.7 CMT group difference	71
Figure 4.8 Group difference Famous Faces Test	73
Figure 4.9 Global face memory scores showing group differences (self-reported DPs vs controls) in mean face memory scores using (A) accuracy and (B)	76
Figure 5.1 Overview of tasks	102
Figure 5.2 Group differences Faces among non faces	109
Figure 5.3 Face inversion effect: Faces among non faces.....	112
Figure 5.4 Gender categorisation task.....	116
Figure 5.5 Comparing group performance on the gender categorisation task	117
Figure 5.6 Age categorisation task.....	120
Figure 5.7 Comparing group performance spacing versus features.....	129
Figure 5.8 Comparing CFPT inversion effects DP versus control groups.....	134
Figure 6.1 Ability to recognise immediate family and close friends if meeting them unexpectedly out of context	163
Figure 6.2 Number of encounters DPs needed to be able to recognise a person	163
Figure 6.3 Frequency of error types.....	165
Figure 6.4 Who had participants disclosed their face recognition difficulties to?	165
Figure 6.5 Themes and sub themes.....	166
Figure 6.6 Compensatory strategies rated for usefulness by participants	177
Figure 6.7 Participants' priorities for future research.	195

List of tables

Table 3.1 Participant demographics	26
Table 3.2 Descriptive statistics Matrix Reasoning Item Bank	32
Table 3.3: Correlation tables matrix reasoning and bicycle memory tasks.....	40
Table 4.1 Individual case scores of self-reported DP participants.....	52
Table 4.2 Test reliabilities.....	64
Table 4.3 Old New Faces accuracy and BIS by age group	67
Table 4.4 CFMT accuracy and BIS by age group.....	69
Table 4.5 Cambridge Bicycle Memory Test accuracy and BIS by age group	70
Table 4.6 FFT identification accuracy by age group	73
Table 4.7 Global face memory BIS by age group.....	75
Table 4.8 Comparing effect sizes for the difference between potential DP and control group scores before (accuracy) and after (BIS) accounting for RT	77
Table 4.9 Logistic regression models predicting self-reported group membership (DP or control).....	78
Table 4.10 Global memory accuracy and BIS classification for all self-reported DP participants	81
Table 4.11 Individual cases compared using accuracy and BIS	82
Table 5.1 Participant demographics.....	99
Table 5.2: DP Z scores on tests used for classification.....	101
Table 5.3 Test reliability	106
Table 5.4 Faces among non faces descriptive statistics	110
Table 5.5 Descriptive statistics Mooney faces.....	114
Table 5.6 Descriptive statistics gender categorisation task.....	117
Table 5.7 Descriptive statistics age categorisation task	121
Table 5.8 Descriptive accuracy statistics configural featural task	125
Table 5.9 Configural featural task: Comparing DP and control group performance	126
Table 5.10 Descriptive statistics: Configural and featural task spacing and features accuracy	128
Table 5.11 Descriptive statistics Cambridge Face Perception Test	133
Table 5.12 Intraclass correlation estimates	138
Table 5.13 Summary of hypothesis testing results at group level.....	140
Table 5.14 DP z scores.....	142
Table 6.1 Participant characteristics	158
Table 6.2 Compensatory strategies (spontaneous report)	172

Glossary of terms

Acquired prosopagnosia (AP): A face recognition disorder with sudden onset e.g. following stroke or head injury

Ambient images: Images that capture a natural range of variation in facial appearance.

Apperceptive subtype of DP: widespread deficits in face perception in addition to impairment in face memory

BIS: Balanced Integration Score ($Z_{\text{accuracy}} \text{ minus } Z_{\text{RT}}$), an integrated measure that adjusts accuracy to take account of response time this controlling for differential speed-accuracy trade offs

CBMT: Cambridge Bicycle Memory Test

CFMT: Cambridge Face Memory Test

Configural processing: encoding of the spacing between internal face parts, for example the distance between the eyes

Developmental prosopagnosia (DP): A lifelong face recognition condition that cannot be explained by brain injury or other neurodevelopmental or psychiatric condition

DPs: Individuals with DP

EEG: Electroencephalography

Face categorisation: Classifying visual input as a face

Face detection: Finding face among other stimuli e.g. objects, or within a scene

Face identification: The ability to identify a face, either by name or other biographical detail (e.g., Actor who plays Mr. Bean).

Face inversion effect (FIE): The observation that presentation face stimuli upside down impairs face processing performance compared with upright presentation

Face perception: Encoding of a face

Face processing: The overall process involved in recognising a human face

Face recognition/face memory: The ability to decide if a face has been seen before. Face *memory* tasks do not require participants to be able to name the facial exemplar or to recall any biographical details about an individual.

Featural processing: encoding of individual face parts such as nose or eyes

FFA: Fusiform Face Area

FFT: Famous Faces Test

FG: Fusiform gyrus

fMRI: Functional magnetic resonance imaging

Holistic processing: the processing of a face as an integrated whole, rather than as a collection of parts

MaRS-IB: Matrix reasoning item bank

MEG: Magnetoencephalography

Mnemonic subtype of DP: Affects only face memory, face perception is within the typical range

MTG: Middle temporal gyrus

PI-20: Twenty item Prosopagnosia Index – a validated self-report questionnaire measuring prosopagnosic traits

STS: Superior temporal sulcus

Impact of Covid 19 statement

I began my PhD in October 2020. Covid-19 restrictions in place in the University of Stirling, and more generally in the UK in 2020-2022 combined with guidance from my funder ESRC, had the result that all data collection for the studies reported in chapters 3,4, and 5 had to be moved online. I had originally intended to test participants in person, and to use eye tracking in addition to behavioural methods but this was not possible. Instead, I took the opportunity to develop a more extensive test battery that provided richer data and was able to recruit participants from across the UK, Ireland and the US. As a result, I was able to recruit a larger sample size of DPs than the 12 participants originally envisaged in my research proposal.

Author contributions and dissemination

Chapter 3 is published as Lowes, J., Hancock, P. J. B., & Bobak, A. K. (2024). Evidence for different visual processing strategy for non-face stimuli in developmental prosopagnosia. *Visual Cognition*, 1–12. <https://doi.org/10.1080/13506285.2024.2359743>

Author contributions: **Judith Lowes**: Conceptualisation, Data curation, Formal analysis, Writing - Original draft preparation, Visualisation, Investigation, Funding acquisition, Project administration, **Anna Bobak**: Conceptualisation, Supervision, Writing - Reviewing and Editing, Funding Acquisition, **Peter Hancock**: Conceptualisation, Supervision; Writing - Reviewing and Editing.

Chapter 4 is published as Lowes, J., Hancock, P. J. B., & Bobak, A. K. (2024). A new way of classifying developmental prosopagnosia: Balanced Integration Score. *Cortex; a journal devoted to the study of the nervous system and behavior*, 172, 159–184. <https://doi.org/10.1016/j.cortex.2023.12.011>

Author contributions: **Judith Lowes**: Conceptualisation, Data curation, Formal analysis, Writing - Original draft preparation, Visualisation, Investigation, Funding acquisition, Project administration, **Anna Bobak**: Conceptualisation, Supervision, Writing - Reviewing and Editing, Funding Acquisition, **Peter Hancock**: Conceptualisation, Supervision; Writing - Reviewing and Editing,

Chapter 5 is in preparation. **Author contributions**: **Judith Lowes**: Conceptualisation, Data curation, Formal analysis, Writing - Original draft preparation, Visualisation, Investigation, Funding acquisition, Project administration; **Peter Hancock**: Conceptualisation, Supervision; Writing - Reviewing and Editing, **Benjamin Armstrong** Formal analysis (single cases), **Anna Bobak**: Conceptualisation, Supervision, Writing - Reviewing and Editing, Funding Acquisition.

Chapter 6 is in preparation. **Author contributions**: **Judith Lowes**: Conceptualisation, Data curation, Formal analysis, Writing - Original draft preparation, Visualisation, Investigation, Funding acquisition, Project administration, **Lesley McGregor**: Formal analysis, Supervision, Writing - reviewing and editing, **Peter Hancock**: Conceptualisation, Supervision; Writing - Reviewing and Editing, **Bradley C. Duchaine**,

Resources, **Anna Bobak**: Conceptualisation, Supervision, Writing - Reviewing and Editing, Funding Acquisition.

Conference presentations:

Neurodevelopmental Disorders Annual Seminar June 2022 (poster)

Experimental Psychology Society meeting July 2022, Stirling (poster)

European Conference on Visual Perception August 2022, Nijmegen (poster)

Lowes, J., Hancock, P., & Bobak, A. (2022, December). Investigating developmental prosopagnosia across the lifespan. *PERCEPTION* (Vol. 51, pp. 144-144)

Experimental Psychology Society meeting January 2023, London (talk)

Vision Sciences Society Conference May 2023 Florida (poster)

Lowes, J., Hancock, P.J.B., Bobak, A.K. Accounting for speed-accuracy trade-offs in developmental prosopagnosia. *Journal of Vision* 2023;23(9):4688.

<https://doi.org/10.1167/jov.23.9.4688>.

Experimental Psychology Society meeting July 2023, York (talk)

European Conference on Visual Perception August 2024, Aberdeen (accepted)

Chapter 1: Thesis overview and aims

1.2 Aims of the thesis

The overall aim of this thesis is to understand at a fine-grained level the nature of the face processing deficits that occur in developmental prosopagnosia, and whether these are the same in all cases of DP, or whether different sub types of the condition might exist. The thesis has three research objectives:

- To understand which processes and subprocesses of face perception are objectively impaired in DP
- To assess which measure, or combination of measures, best classifies DP
- To understand the experiences of living with developmental prosopagnosia

1.2.1 Research questions

Chapter 4:

RQ1: Across four tests of face memory, does accuracy or balanced integration scores (BIS) yield an objective measure that better accounts for self-reported face recognition ability?

Chapter 5:

RQ2: Are all, or some, DPs impaired at face detection?

RQ3: Are all or some DPs impaired at facial gender categorisation?

RQ4: Are some or all DPs impaired at judging facial age?

RQ5: Do children and/or adults with DP show deficits in configural face processing?

RQ6: Is face perception ability related to face identity recognition?

Chapter 6:

RQ7: How many individual faces can people with DP typically recognise?

RQ8: Can DPs recognise close friends and family?

RQ9: Is there a relationship between real life reported and objective measures such as Twenty Item Prosopagnosia Index (Shah, Gaule, Sowden, et al., 2015) and Cambridge Face Memory Test (Duchaine & Nakayama, 2006) scores?

RQ10: What aspects of DP do people find most difficult – and conversely are there any positives?

RQ11: What is DPs' experience of seeking a diagnosis?

RQ12: What is DP's experience of disclosing their difficulties to others and requesting adaptations?

RQ13: What do DPs believe the priorities for future research into DP should be?

Specific hypotheses relating to research questions 2 – 6 were preregistered and are outlined within the relevant chapters.

1.2 Thesis structure

The thesis is structured as follows. Chapter 2 provides a general overview of DP. Chapters 3-6 are empirical chapters; Chapters 3, 4, and 5 present the experimental research and Chapter 6 a mixed methods survey. Each empirical chapter is presented in the form of a manuscript, two of which have already been published (see Dissemination in the preface). Each empirical chapter begins with a more specific review of the literature relevant to the research questions for each chapter and concludes with a discussion. Chapter 7 contains a general discussion that considers additional insights and implications for future research and policy. For ease of reference, the preface contains a glossary of some commonly used terms.

Chapter 2: General introduction

2.1 What is developmental prosopagnosia?

Developmental prosopagnosia (DP) is a neurodevelopmental condition that manifests in severe face recognition problems due to the visual mechanisms for face processing having failed to develop (Duchaine & Nakayama, 2006). The word prosopagnosia comes from the Greek, “prosopon” meaning face and “agnosia” meaning without knowledge. DP is characterised by severe difficulty recognising faces, including those of family, friends, colleagues and teachers and is believed to be present from birth (Cook & Biotti, 2016).

One of the first descriptions and definitions of developmental prosopagnosia was provided by Jones and Tranel (2001). Their definition highlights that DP is a face selective difficulty that cannot be explained by factors such as other neurodevelopmental conditions (e.g. Autism, Turner Syndrome), IQ or low-level visual impairment. Throughout this thesis I use the term “DP”, but other literature has used different terminology, notably “congenital” (e.g., Ariel & Sadeh, 1996; Avidan & Behrmann, 2008; Bate et al., 2008; Carbon et al., 2007; Esins et al., 2015) or “hereditary” (Kennerknecht et al., 2006, 2008) for this condition. There has been some discussion of the inconsistent use of the terms “developmental”, “congenital” or “hereditary” prosopagnosia in the literature (Bate, 2013; Nørkær et al., 2024; Susilo & Duchaine, 2013). Bate argues that the terms should not be used interchangeably since “hereditary” implies a genetic link, yet many studies do not test for this. She notes that DP is the broader term and defines DP as a face recognition disorder that is present from birth and cannot be explained by known brain injury pre or post birth, or other neurodevelopmental or social communication disorder (Bate, 2013, p. 123). Bate’s definition is similar to the original definition proposed by Jones and Tranel (2001) but, unlike theirs, does *not* specify that the recognition difficulties are face specific. Face selectivity is discussed further in the introductions to Chapters 4 and 5. I use Bate’s definition.

Since the first case of DP was reported (McConachie, 1976), there have been more than 220 published studies of DP (for reviews see Albonico & Barton, 2019; Bate, 2011; Bate & Bennetts, 2014; Corrow et al., 2016; Dalrymple, 2012; DeGutis et al., 2014; Manipa et al., 2023; Nørkær et al., 2024; Susilo & Duchaine, 2013) and the heterogeneity of DP, both in terms of presentation and severity, is a prevalent theme in

behavioural and neuroimaging studies (e.g., DeGutis et al., 2024; Minnebusch et al., 2007; Rosenthal & Avidan, 2018; Schmalzl et al., 2008). Another important point to note is that DP is not an all-or-nothing condition. Individuals with DP can usually recognise at least *some* faces at least *some* of the time and many use cues including voice, gait and extra facial information to identify others - albeit with varying degrees of success (Fine, 2012; Murray et al., 2018; Portch et al., 2023; Yardley et al., 2008). Because DP does not usually involve a *complete* absence of facial recognition ability, a few authors (Rossion, 2018; Sorensen & Overgaard, 2018) argue that a more accurate term for the condition would be “*prosopodysgnosia*”, “*Dys*” meaning difficulty as in “dyslexia”.

A second, acquired form of prosopagnosia (AP) also exists. Unlike DP, which is assumed to be a lifelong difficulty, AP occurs suddenly following brain damage, for example due to stroke or head injury. In cases of AP brain lesions can be observed, usually in the right hemisphere, or bilaterally. DP research studies therefore typically exclude individuals with a history of head injury or brain damage, and any instances of face recognition difficulties that developed suddenly. For a fuller discussion of classification criteria see Chapter 4.

2.2 Subjective experiences of living with DP

Please see Chapter 6 for a fuller introduction and discussion of this topic, as well as Chapter 4 for a discussion of whether DPs have metacognitive awareness of their difficulties.

Guidelines recommend that diagnosis, or classification, of DP is made on the basis of subjective difficulties in everyday life alongside impaired performance on two objective measures of face memory (Dalrymple & Palermo, 2016). More recently it has been suggested that subjective report alone may be sufficient (Burns, 2023a, 2023b; Burns et al., 2022), although there is disagreement about this approach (DeGutis, Yosef, et al., 2023).

Individuals with DP experience difficulties in interpersonal relationships (Adams et al., 2020a; Diaz, 2008; Yardley et al., 2008) because they can find it hard to form and maintain friendships due to difficulties recognising people they have previously met and interacted with and thus build relationships. Their challenges extend to school and workplaces (Dalrymple, Fletcher, et al., 2014; Fine, 2012; Yardley et al., 2008) and potentially limit the career options of those affected. For example the mother reported by Diaz (2008) found particular difficulties working in retail since this required interacting

with customers and she found difficulty handing items to customers as she was unable to recognise the customer who had requested it. As a result of their face recognition impairment, many DPs experience stress and anxiety in social situations (Yardley et al., 2008) and some parents of children with the condition restrict their social opportunities such as meeting up with friends because of a fear that their child will be unable to recognise their friends and get lost (Dalrymple, Fletcher, et al., 2014; Diaz, 2008).

2.3 How do humans recognise faces? Overview of theories of face recognition

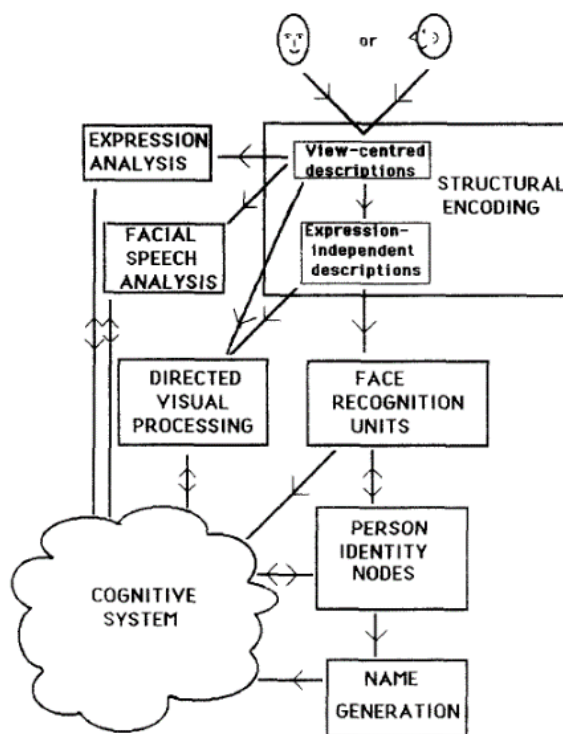
This thesis investigates the objective and subjective experience of living with developmental prosopagnosia (DP) and having lifelong severe difficulties recognising faces. However, in order to comprehend the various ways in which face recognition can go wrong, it is helpful to first have an understanding of the main theories of how face processing and face recognition occur in typical perceivers - people who have no difficulty recognising faces. I therefore next provide a brief overview of the main theories of human face recognition and how this typically develops.

2.3.1 The Bruce and Young model

The dominant cognitive model of face recognition developed by Bruce and Young (Bruce & Young, 1986; Young & Bruce, 2011) is shown below in Figure 2.1. This hierarchical model used evidence from neuropsychological case studies of acquired prosopagnosia, experimental research and diary studies with patient and typical populations. It proposes that seven different types of information are used to process faces. These information types are as follows: *Pictorial codes* which allow the recognition of images only, for example the recognition of a photograph of a face rather than of the face itself in real life; *expression codes* which are processed independently of face familiarity; *facial speech codes* which provide visual information about mouth and lip movements and help viewers to interpret the expression and meaning of speech (note that these facial speech codes may also help identification when such movements are characteristic or idiosyncratic); *visually derived semantic codes* which convey information about a person not related to identity recognition, for example gender, estimated age, or personality traits; *structural codes* which are required to recognise a face as familiar from a novel image that that differs in viewpoint or expression [note that Bruce and Young the term “recognition” here to mean familiarity rather than

identification]. Next are *identity specific semantic codes* which are activated only once recognition has occurred. Examples of the type of semantic information provided by these codes include a person's hobbies and occupation. These *identity specific semantic codes* resolve the sense of familiarity or 'feeling of knowing' that can be generated from structural codes and result in person identity recognition. The final codes are *name codes* which are thought to be stored separately from identity specific semantic codes since it is known that it is possible to recognise a person and recall semantic information about them, such as where they live or and the type of car they drive, without being able to recollect their name. In this cognitive functional model, there is a distinction between *encoding* of faces, that is forming an accurate facial percept, and the later stages that lead to *recognition*. This topic is discussed further in Chapter 5.

Figure 2.1 The Bruce and Young model of face recognition

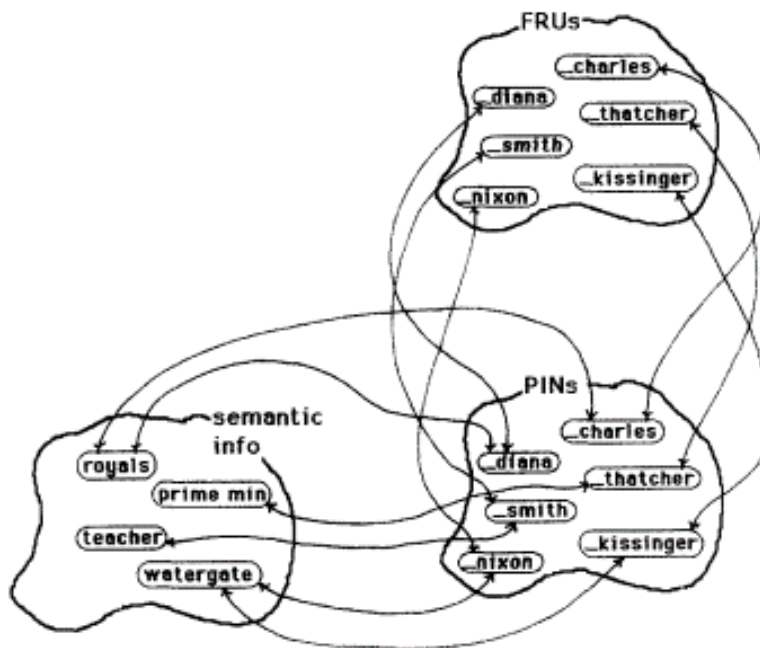


Note: Figure reproduced with permission¹ from Bruce and Young (1986) showing the functional model. The structural encoding stage is involved in face perception, the later stages are involved in recognition

¹ Bruce & Young, Understanding face recognition, British Journal of Psychology (1986), 77, 305-327 © 1986 The British Psychological Society

Further elaborations of the original model include Burton and colleagues' Interactive Activation and Competition model (IAC, Burton et al., 1990). The IAC model (Figure 2.2) proposes there is an additional stage of processing represented by a pool of semantic identification units (SIUs) which form part of Bruce and Young's cognitive cloud. The Bruce and Young model (1986) mentions, but does not talk in detail about, the activation threshold required to move to the next stage and the IAC model makes an assumption that once face recognition units (FRUs) and person identity nodes (PINs) have been activated to a particular threshold this triggers a familiarity judgement and the biographical information within the SIUs can be accessed. The model further posits that individual identities within each of the pools are linked by bi-directional *inhibitory connections* and that units which are linked to the same face in different pools, for example FRU and PIN, have bi-directional *excitatory connections*. Thus, the activation of a particular FRU and its associated PIN occurs in parallel with the *inhibition* of all the other PINs within the pool. The IAC model can therefore account for repetition priming effects (e.g. Ellis et al., 1987) and semantic priming of familiar faces (e.g. Bruce & Valentine, 1986; Rhodes & Tremewan, 1993), as well as covert recognition in prosopagnosics (Barton et al., 2001; De Haan et al., 1991; Rivolta et al., 2010). Covert recognition is recognition without awareness of recognition (Young, 1998), in other words it is unconscious recognition and is measured physiologically though measures such as galvanic skin response and behaviourally using paradigms such as repetition priming in which names and/or faces of identities that are associated in some way (e.g., after hearing after John Lennon's name, Paul McCartney's face would be recognised faster than if hearing the name Emmeline Pankhurst due to the bidirectional activation of person identity nodes (name) and semantic information (Beatles, pop star) and the face recognition units). The IAC model helps explain why contextual and situation cues as well as voice, assist face recognition. Covert recognition in prosopagnosia demonstrates that face recognition is not always a binary (yes/no) phenomenon.

Figure 2.2 Burton, Bruce and Johnston's (1990) Interactive Activation and Competition Model



Note: Figure reproduced² from Burton et al. (1990) showing the interactive and bi-directional elements of the face processing system

2.3.2 Neural models

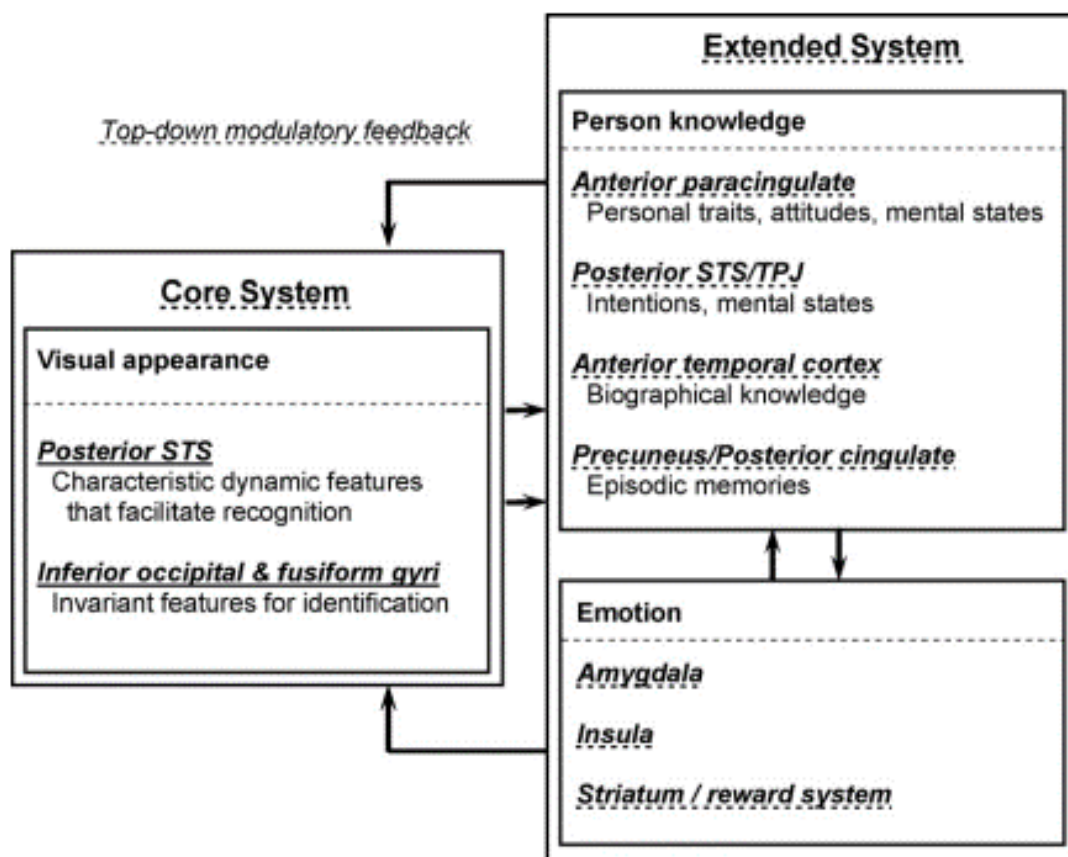
The introduction and widespread adoption of neuroimaging methods in the 1990s led to development of neural models of face recognition. A key discovery underpinning these models was the existence of the Fusiform Face Area (FFA), a region of the brain that becomes more activated when looking at faces than at other classes of stimuli (Kanwisher et al., 1997).

The influential Gobbini and Haxby model (2007) draws on previous neural models (Haxby et al., 2000) and proposes that human face recognition comprises a core and an extended system (Figure 2.3). The core system is made of three brain regions with different roles. The fusiform face area (FFA) within the fusiform gyrus processes *non-changeable* aspects of a face such as identity and gender, the posterior superior temporal sulcus (pSTS) processes *changeable* facial aspects, such as expression and gaze direction. Finally, the occipital face area (OFA) encodes basic visual aspects and information about

² Used with permission of John Wiley & Sons from [Understanding face recognition with an interactive activation model, Burton, Bruce, and Johnston, 81: 361-380, (1990) permission conveyed through Copyright Clearance Center, Inc.

face parts. The extended system processes information about the person, for example mental state, episodic memories and biographical (semantic) information such as occupation. Two important differences between the Gobbini and Haxby neural model and the Bruce and Young functional model (1986) are first that the neural model assumes a more general system compared with the independent systems posited by Bruce and Young (1986). Second, the Gobbini and Haxby model assumes that face processing is interactive (like the IAC model), integrating information from regions that process physical information about faces as well as the *meaning* of that physical (configural) information.

Figure 2.3 The Gobbini and Haxby (2007) neural model of familiar face recognition



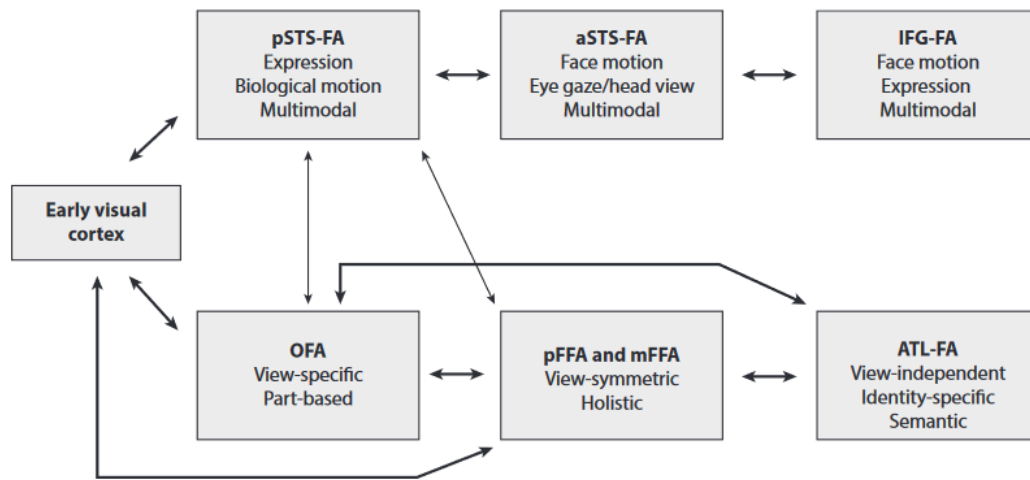
Note: Showing the core and extended face processing systems³. STS = superior temporal sulcus, TPJ = temporoparietal junction.

³ Figure reproduced with permission from *Neuropsychologia* (45) pages 32-41, Gobbini and Haxby (2007) © Elsevier

More recently, Duchaine and Yovel (2015) reviewed the neurological evidence for face selective regions of the brain to date and proposed an updated neural framework (Figure 2.4) for face processing to take account of newer neurobiological findings. For example that lesions to a third face selective area in the fusiform gyrus, the anterior temporal lobe face area (ATL-FA) can cause prosopagnosia (Barton et al., 2002; Busigny et al., 2014). They began by summarising the evidence for face selective regions of the brain, separating these into those found in the ventral stream that are thought to be responsible for the perception of form and those in the dorsal stream responsible for processing dynamic social information and facilitation of social interaction. They concluded that there are multiple pathways into the face recognition system, in contrast to previous models (Gobbini & Haxby, 2007; Haxby et al., 2000, 2002) which suggested the OFA was the single entry point. For example, they cite evidence from PS, an acquired prosopagnosic whose OFA was destroyed by brain damage but who nevertheless exhibits a right FFA and right posterior STS-FA similar to those in healthy patients in relation to magnitude and face selectivity (Sorger et al., 2007) and from patient DF (Milner and Goodale 1995) who lacked an OFA in either hemisphere as evidence that the OFA is not required. Duchaine and Yovel's revised model (2015) therefore proposes the early visual cortex and pSTS as additional routes into the FFA and thus the core face network. According to their model, the FRUs needed for person identity recognition are found in the anterior superior temporal sulcus face area (aSTS-FA). Additionally, this revised framework suggests that the aSTS-FA might also play a role in voice processing.

Although the three leading models of face recognition described have some key differences, for example in the extent to which they suggest that the various parts of the face processing system function independently or interactively, they nonetheless share many common elements. Of particular relevance to the study of the DP are the many brain areas and connectivity routes that are putatively thought to be disrupted in DP and these are considered below in the section 2.5.3 Neurobiological accounts and neural correlates of DP.

Figure 2.4 Duchaine and Yovel’s Revised Neural Framework for Face Processing (2015)



Note: The ventral face-processing pathway (bottom) comprises the occipital face area (OFA), the fusiform face area (FFA), and the anterior temporal lobe face area (ATL-FA). The dorsal face-processing pathway (top) consists of the posterior superior temporal sulcus face area (pSTS-FA), the anterior superior temporal sulcus face area (aSTS-FA), and the inferior frontal gyrus face area (IFG-FA)⁴.

2.4 How and when does face recognition typically develop?

Although DP is widely accepted to be a neurodevelopmental condition (but see Volfart & Rossion, 2024), little is known about precisely *when* disruption of the face processing system occurs. DP is widely described as lifelong (by which it is typically meant that DP is incurable). It *may* be present from birth (congenital) but at present there is no strong empirical evidence to support this assumption (see the section above 2.1 What is developmental prosopagnosia?). Alternatively, DP could be the result of disruption to the face processing system at some point during development. It is also possible that both a congenital and developmental form of the conditions exist (Nørkær et al., 2024). Therefore, the questions of *how* and *when* face processing abilities typically develop into the specialist, rapid, and expert face processing system seen in most healthy

⁴ Figure used with permission of Annual Reviews from The Annual Review of Vision Science (1: 393-416), Duchaine and Yovel (2015).

adults are relevant both to the understanding of the possible causes of DP, and precisely which aspects of face processing (see Chapter 5) may be impaired. The models of face processing outlined above attempt to theorise how face processing occurs in adults. But how and when does this particular expertise for faces develop through childhood and adolescence?

To first address the *how*. Several hypotheses have been advanced to explain *how* the specialism for faces develops and I introduce four key hypotheses below. Although the accounts differ, a common theme is the important role that expertise plays in the development of a mature, rapid and highly expert face identity recognition system.

2.4.1 The two process theory of face recognition

The two process theory of face recognition (Johnson et al., 2015; Morton & Johnson, 1991) is informed by work on filial imprinting in chicks (Bateson & Horn, 1994). It proposes that human infants are born with some knowledge of faces and show early preference for faces and face-like stimuli (CONSPEC, which is controlled by a subcortical process and is primarily for face detection and threat detection (i.e. fearful faces)). This early predisposition to faces, which dissociates later in development, then biases the visual input to prioritise faces, resulting in the development of a highly specialised cortical face processing system (CONLEARN). *Experience* therefore drives this specialisation. More recent work has extended the theory by highlighting that the subcortical route continues to play an important role in face perception into adulthood, including for detecting eye contact (and thus may help explain the difficulties autistic people have in interpreting social cues such as eye contact (Sasson, 2006)). It is possible the interaction between cortical and subcortical routes may be a fruitful area for further research in DP since CONSPEC (detection) appears to activate CONLEARN, the cortical face processing route.

However, there is much evidence against the CONSPEC theory. For example, there is evidence that face-like stimuli also activate the FFA (Sagiv & Bentin, 2001) and it has been reported that newborns prefer their mothers' faces to those of strangers within 24 hours suggesting very early development of face recognition (Bushnell, 2001). As such Bate (2013, p. 114) suggests that CONSPEC might be used both for detecting faces and also directing attention to faces that are receptive to social interaction.

2.4.2 Perceptual narrowing hypothesis

A second account that also emphasises the key role of experience in the development of a special face processing system is the *perceptual narrowing hypothesis* (Nelson, 2001). This hypothesis suggests that the plasticity of the brain in early months and years of life allows the development of a highly specialised face processing system from the relatively crude one present at birth through experience with faces. This theory was refined (Scott et al., 2007) to suggest that regular exposure to familiar faces causes the neural circuits to strengthen, whilst those for infrequently seen faces are weakened. Effects such as the well-documented other race effect (ORE, for a review and meta-analysis see Meissner & Brigham, 2001) and the fact six month old babies can discriminate monkey faces, but 10 month old babies and adults could not (Pascalis et al., 2002) are argued to provide support for this hypothesis.

2.4.3 How does face recognition develop through childhood?

Second, the question of precisely *when* face processing ability typically matures, and indeed whether all elements of the face processing mature at the same time is also important for understanding DP. First, from a purely practical perspective, understanding how and when particular abilities develop is needed to inform the choice of tests to be used to assess face processing ability in children (see Dalrymple & Palermo, 2016). Second, from an aetiological perspective, although the condition is lifelong, it is possible that the face recognition difficulties that characterise DP may arise from disrupted development of the face processing system rather than being present from birth. However, relatively few studies have investigated DP in children with only 14 identified in the recent scoping review by Nørkær and colleagues (2024). Further, Epihova and Astle (2024) point out that very few studies have so far attempted to integrate neural and developmental accounts of DP and suggest that future research should investigate the neural correlates of DP across childhood in order to better understand the developmental nature of the condition.

Within the relatively sparse literature on DP in children, there is some evidence to suggest that DP in childhood may be both qualitatively and quantitatively different from DP in adulthood. Dalrymple et al. (2014) reported performance differences between children and adults with DP; all the children but only half of the adults, showed impaired face *perception* difficulties in addition to their face memory difficulties. Further support comes from work investigating the prevalence of DP in children (Bennetts, Murray, et

al., 2017) which estimated that between 1.2% and 4% of children aged between five and 11 years had face-specific processing deficits. This reported prevalence is higher than the commonly-proposed estimate for adults of around 2% (Bowles et al., 2009; Kennerknecht et al., 2006, 2008), however see DeGutis (2023) and a fuller discussion of the prevalence of DP in Chapter 4. Together, these studies suggest that it might be possible to “grow out of” DP – or at least some aspects of the condition. Alternatively, the reported differences between children and adults could be due to the methodological challenges that arise from testing young children, or because the compensatory strategies that many individuals with DP adopt can mask their difficulties on certain laboratory tests.

Face recognition ability has been argued to continue to develop up to one’s early thirties peaking between 30 and 34 years of age (Germine et al., 2011). Several accounts have been proposed to explain the development of face processing over time and those that are most relevant to the research questions I address in this thesis are discussed below.

2.4.3.1 Encoding switch hypothesis

Carey and Diamond (1977, 1994) tested children and adults on face recognition tasks to estimate the age at which face perception becomes adult like. A key measure used by Carey and Diamond was the face inversion effect (FIE, Yin, 1969); the finding that upright faces are processed more quickly and accurately than inverted faces is one of the most robust in the face processing literature. Their hypothesis was that young children cannot encode *unfamiliar* facial configurations from a single photo and instead encode faces on a feature-by-feature basis until the age of approximately 10 years when they switch to the use of configural processing – that is the processing of the spacing between features. Note that Carey and Diamond classed facial expression and paraphernalia (i.e. hats) to be “features” but more typically “features” in the context of face research would be features of the face itself, such as eyes and nose. Results were that 6-year-olds showed no FIE, but FIEs were observed in eight-year-olds and more strongly at age 10. When asked to rate which two of four faces were most similar, children under 10 years typically chose the faces wearing hats, regardless of the similarity of the faces themselves. The authors argue this means young children relied on “features” (in this case paraphernalia) more than configuration, despite instructions to look at faces. The more recent version of the hypothesis (1994) suggests that the switch from featural to configural face processing occurs around six years of age. Note that paraphernalia would not usually be considered

to be a face feature and the claims made here cannot be generalised to facial features per se. Furthermore, other work has failed to replicate these findings and found evidence of face inversion effects in children as young as three to four years (Pellicano & Rhodes, 2003; Picozzi et al., 2009) as well as in infants (Turati et al., 2004).

2.4.3.2 Expertise hypothesis

Similar to the encoding switch hypothesis, the expertise hypothesis (Diamond & Carey, 1986; Gauthier & Tarr, 2002) posits that improvements in face recognition ability are associated with increased use of configural processing, and thus developing expertise, through childhood exposure. Diamond and Carey (1986) tested undergraduates and dog experts and found that dog experts showed inversion effects for dogs similar to faces and thus argue that expertise, rather than any face-specific mechanisms, drives inversion effects.

Other research supporting the role of face expertise in the development of face recognition includes a study which found that people living in sparsely populated areas performed less well on task of face identification than those who had experienced a rich variety of faces, sometimes called “face diet” (Balas & Saville, 2015, though see Sunday et al., 2019).

It is known that people find familiar faces much easier to discriminate and recognise than unfamiliar faces (Hancock, 2012; Megreya & Burton, 2006). Further, it is known that exposure to a high level of within-person variability (for example seeing faces of the same individual under different lighting conditions and from different viewpoints) assists with face learning and recognition (Baker et al., 2017; Bindemann & Hole, 2020; Ritchie & Burton, 2017) perhaps because this permits the observer to learn particular identity characteristics that do not change over time and thus gain insight into ways in which the appearance of an individual face can naturally vary (Bruce et al., 1994). However, until fairly recently it was not known whether variability was also important for face learning in children. Laurence and Mondloch (2016) conducted one of the first studies to consider the role of within-person variability in children’s face recognition. They investigated the developmental trajectory of children’s recognition ability for both familiar and unfamiliar faces and found evidence that recognition ability for familiar faces improved between four and six years of age but that by age six, children’s recognition ability for *familiar* faces is adult-like and very accurate. By contrast, children’s ability to recognise *unfamiliar* faces improved across the age range (up to age

12 years in this study). Using a simplified face sorting task and, importantly, ambient (i.e. naturalistic, unstandardised) images of familiar and unfamiliar faces that contained a high degree of within-person variability, the authors report that children - like adults - were very good at matching highly variable images of personally familiar faces but made more errors with unfamiliar faces. Additionally, children – like adults (Jenkins et al., 2011) - found telling faces together (i.e. being able to match highly variable images of the same person) more difficult than telling faces apart (i.e. knowing when two images were different people).

To refocus on the question of whether children’s face recognition is similar to that of adults, it has been argued (Bate, 2013, p. 121) that, overall, there is evidence to support the idea that children possess the mechanisms needed for adult-like levels of efficient face recognising at a young age (see Bennetts, Mole, et al., 2017 for a case of 14 year old ‘super’ face recogniser), but that it is likely that wider cognitive factors e.g. attention or concentration result in reduced performance vs adults. In other words, differences in face processing between young children and adults are likely quantitative rather than qualitative.

2.5 Aetiology and proposed accounts of DP

Having considered how human face processing typically develops and functions, how do aberrations result in DP? The cause(s) of DP are currently unclear. Several accounts have been proposed and these are briefly discussed below.

2.5.1 A genetic basis?

There are numerous reports of a familial factor in DP (Bate et al., 2024; De Haan, 1999; Duchaine, Germine, et al., 2007; Duchaine, Yovel, et al., 2007; Grueter et al., 2007; Johnen et al., 2014; Kennerknecht et al., 2006; Lee et al., 2010; Schmalzl et al., 2006, 2008), and in face processing ability in the population more generally (Wilmer et al., 2010; Zhu et al., 2010). One study investigated face processing ability among the large extended family of a Bengali student with DP (Kennerknecht et al., 2006). The pattern of impairment was suggestive of an autosomal-dominant inheritance, in others word a gene that only has to be carried by one parent to be passed on. However, in a commentary on the Kennerknecht et al., (2006) study, Duchaine (2008) argued that the assumption that all cases had a genetic component overstated the contribution of genetics and suggested that DP may have some hereditary element in only around half of cases (although the

penetrance rate of genetic conditions, i.e. the proportion of individuals with a particular genotype who go on to develop the disease varies from condition to condition so it is possible that individuals carry, but do not express, the genetic traits). To date, no single gene for DP has been identified and since most neurodevelopmental disorders result from a combination of multiple genetic variants, or alleles, that are also found in the general population (Albonico & Barton, 2019), it is possible that a combination of genes rather than a single gene is responsible for DP.

The family studies of DP have also revealed other interesting findings, for example that even within the same family the pattern of face processing impairments can differ not only in severity (Johnen et al., 2014, but also type (Lee et al., 2010; Schmalzl, Palermo, & Coltheart, 2008). It has therefore been suggested that DP may not be a single condition but instead a “cluster of related subtypes” (Schmalzl, Palermo, & Coltheart, 2008).

2.5.2 Does DP result from early visual deprivation?

Given the importance of experience in the development of typical face processing discussed above (see 2.4 How and when does face recognition typically develop?), it has been proposed that DP may be explained by early visual deprivation. Support for this hypothesis comes from studies of infants born with severe bilateral cataracts. For example, Geldart and colleagues (2002) investigated face identity and expression recognition in 18 patients who were born with dense, central bilateral cataracts that had been removed after 51 – 586 days. Cataract patients’ expression recognition ability was similar to healthy controls’, but face identity recognition ability with viewpoint change (but not expression change) was significantly lower in the patient vs control group with no effect of duration of cataracts. The authors concluded that early visual input is essential for the development of configural face processing. Consistent with this, other work with patients whose cataracts were removed at a young age have reported lifelong impairment in face detection (Mondloch et al., 2013), face matching and lip reading (Putzar et al., 2010). Children whose early visual input was disrupted by visual impairment have also been reported to show deficits in face recognition ability vs healthy controls. Freire et al., (2022) used the Cambridge Face Memory Test-Children (CFMT-C, Croydon et al., 2014) to compare the face recognition ability of 40 children with binocular and monocular visual impairment and healthy age matched controls. Children with binocular impairment were significantly less accurate than those with monocular visual impairment (strabismic

amblyopia) and also healthy controls. The finding that strabismic amblyopia did not affect face memory should be interpreted with some caution since other research (Cattaneo et al., 2013) reported that strabismic amblyopia is associated with face perception difficulties, specifically the ability to discriminate faces based on configural relations (i.e. the spacing between face features) and thus may cause face recognition deficits.

However, many DPs have no history of visual impairment or previous cataracts and DP research typically excludes participants with a history of atypical or impaired visual processing so early visual deprivation is unlikely to be the primary cause of DP.

2.5.3 Neurobiological accounts and neural correlates of DP

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 several common findings within the neurobiological literature that might explain DP. In particular, there is 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 (r-FFA) 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. 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.

Of specific interest to the research questions I address in my thesis 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, the authors 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 connectivity 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*.

One possible neurobiological explanation for heterogeneity across neuroimaging as well as behavioural studies of DP (for fuller discussion of other factors that might explain the heterogeneity of DP see Chapter 4) is the neuronal migration error account (Ramus, 2004). This hypothesis posits that neurons in the developing brain migrate abnormally and can therefore end up in the wrong place. Crucially, the neurons can end up in slightly different places in different individuals and therefore the structural abnormalities can also differ between individuals, resulting in heterogeneous neurodevelopmental disorders such as dyslexia (Ramus, 2004) and dyscalculia (Butterworth et al., 2011). Although no direct evidence to support the neuronal migration account of DP has yet been found, other findings that are consistent with this hypothesis includes widespread evidence that DP often, though not always, co-occurs with other neurodevelopmental conditions including difficulties recognising bodies and scenes (Jiahui et al., 2018) and objects (for a review see (Geskin & Behrmann, 2018). This frequent, though not universal, co-occurrence with other forms of visual agnosia led to the development of the independent disorders hypothesis (IDH, Gray & Cook, 2018) which predicts that cases of ‘pure’ face and ‘pure’ object recognition can occur, but that the two conditions can also co-occur.

2.6 Which aspects of face processing are impaired in DP?

Precisely which aspects of face processing are impaired in DP, beyond face identity recognition, remains an open question and the literature contains many mixed findings. For a fuller discussion of this question see Chapter 5. Here I present only a short overview as context for my research questions.

2.6.1 Holistic face processing

Briefly, the most investigated question in the DP literature is whether holistic processing is disrupted (Nørkær et al., 2024). Holistic processing refers to the idea that faces are processed as an integrated whole, or “gestalt” rather than as a series of independent features and impaired holistic processing has been proposed as a cause of

DP (Avidan et al., 2011). However other studies have reported normal strength holistic processing (e.g. Susilo et al., 2010) or normal holistic processing of the mouth but not the eyes in DP (DeGutis et al., 2012). Such conflicting results (for an overview see DeGutis et al., 2012) may be partly due to methodological factors.

For example, different measures have been used within the literature to measure holistic processing, most commonly the face inversion effect (FIE Yin, 1969), the composite face effect (Young et al., 1987), and the part whole effect (Tanaka & Farah, 1993) - for a review of holistic processing see (Richler & Gauthier, 2014). Importantly for the study of DP, these different holistic face processing measures may in fact tap rather different abilities (Richler & Gauthier, 2014). Supporting evidence for this view comes from Boutet et al (2021) who used factor analysis to assess whether commonly used tasks thought to index holistic processing do in fact tap the same underlying construct. Results indicated a two-factor solution with the part-whole effect and configural-featural detection task loading onto one factor and the composite face effect loading onto a second factor. Thus, these findings provide evidence *against* a common holistic mechanism for face processing. Further, the authors suggest that the observed individual differences in holistic processing might be accounted for by a differential ability to process the featural aspects of faces.

Consistent with Boutet et al., (2021), a recent study (Leong et al., 2023) investigated the role of featural and holistic processing on face recognition ability among 87 Malaysian Chinese adults with typical face processing. The study used a fixed trajectory aperture paradigm which constrains how participants scan faces. Results showed that participants' ability to use *holistic* processing was related to better face recognition ability as measured by the Cambridge Face Recognition Test-Chinese (CFMT-Chinese, McKone et al., 2012) and, conversely, that better *featural* processing was related to better performance on the old new task. The authors posit that holistic and featural processing play distinct roles at different stages of face identity recognition. A limitation of this study is that the CFMT-Chinese version was used as the basis for estimating objective face recognition ability and later work (Kho et al., 2023) has shown that Malaysian Chinese do not perform as well on the CFMT-Chinese as they do on the Malaysian version of the test (CFMT MY, Kho et al., 2023) and therefore the version of the CFMT by Leong and colleagues used may not be a precise measure of face recognition ability in that Malaysian Chinese population.

Work using a similar fixed aperture paradigm to test holistic processing in DPs rather than among typical perceivers found evidence for holistic processing in DP (Tsantani et al., 2020). One benefit of the fixed aperture paradigm is that each participant views each facial region for the same amount of time, unlike in gaze contingent paradigms (e.g. (Verfaillie et al., 2014) where DPs may be looking at the “wrong” areas of the face. Results were that DPs performed below controls in both conditions (whole face and aperture viewing) and were affected to a similar degree to controls by the aperture condition. The authors argued that DPs *do* process faces holistically, since if they relied only on piecemeal region by region inspection it would be expected that DPs would be *less* affected by the aperture condition than controls. The findings suggest that DPs are less efficient at extracting vital information that can support identity recognition from the facial regions they view and thus their encoded representations are *insufficiently precise* to allow reliable recognition. This is regardless of whether the representations are generated through viewing the full face (whole face condition) or via a serial inspection of facial regions (aperture condition). The authors conclude that DPs appear to be able to process faces holistically (see also Biotti et al., 2017; Le Grand et al., 2006; Susilo et al., 2010 for similar findings).

Additionally, many studies have investigated specific subprocesses in DP (for example, face detection or gender categorisation) again with mixed results. These studies are discussed in detail in Chapter 5. However, two important overall conclusions emerge from this review of the literature. First, face processing comprises many separate subprocesses, only some of which may be impaired in DP and the precise underpinnings of DP may therefore vary between individual cases. If this is the case then group level analysis of small sample sizes, which has traditionally been the norm in DP research, could fail to detect important and theoretically interesting impairments or differences that may be observed in only a minority of DPs. Second, there is converging evidence that different measures of face processing in fact tap different sub-processes (e.g. Bate, Bennetts, et al., 2019) – or even sub-sub-processes in the case of holistic processing (see Chapter 5). Together these suggest that using multiple measures of face perception and face memory is important to fully understand DP.

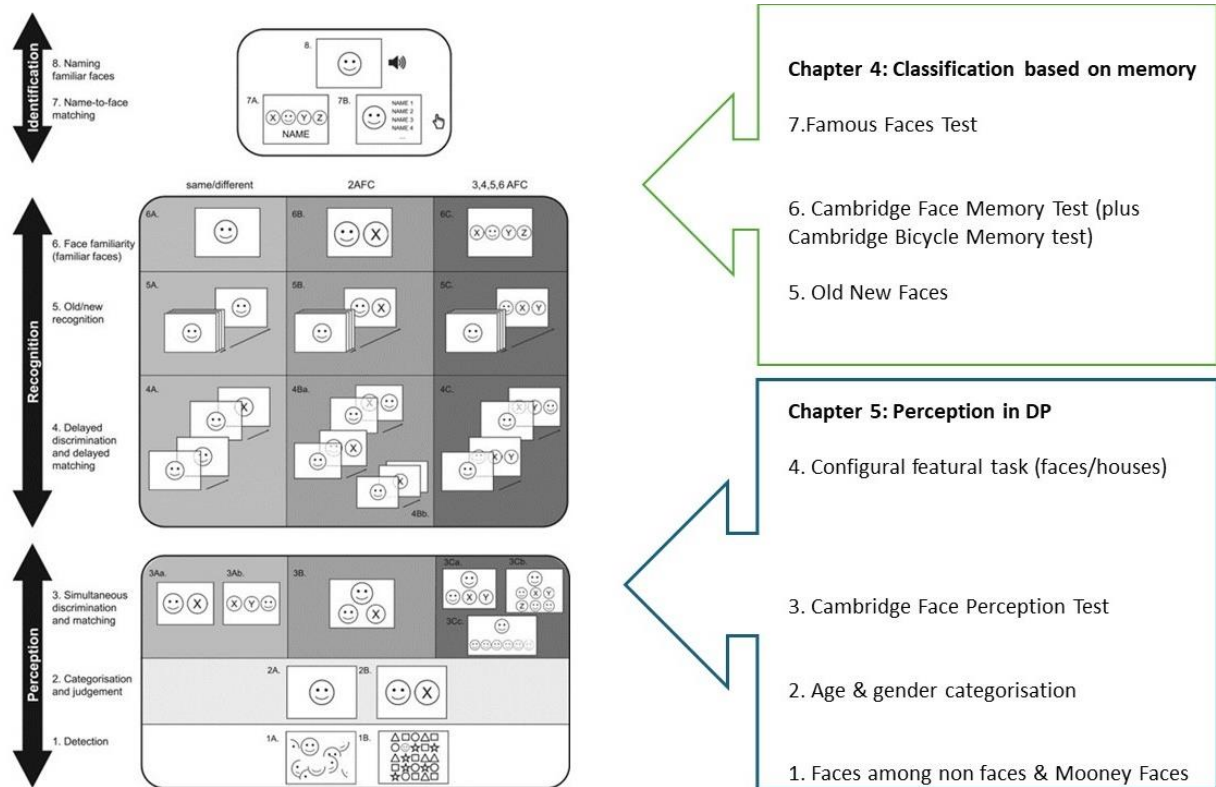
The DP literature contains many mixed findings (see also Chapters 4 and 5). For example, at group level face gender categorisation has been reported both as impaired (Esins et al., 2016; Marsh et al., 2019) and unimpaired (Chatterjee & Nakayama, 2012; Le Grand et al., 2006). Pertinently, due to small sample sizes, null effects could be due

to insufficient statistical power. Therefore, the choice of analysis methods should include group level analysis *alongside* individual case analysis to detect patterns of impairment at the individual level. The use of Bayesian statistics also permits an assessment of the *strength* of the evidence for and against null hypotheses. Together these three analytical methods can help resolve some of the open questions that exist about the exact nature of the impairments in of DP.

2.7 Thesis overview and aims

To recap, the overall aim of this thesis (see Chapter 1) was to understand at a more fine-grained level the nature of the deficits in DP, and crucially whether these are the same in all cases of DP. I addressed limitations of previous studies in three key ways. First, my research design (see Figure 2.5) followed the framework developed by Robotham and Starrfelt (2018) and systematically tested seven of the eight subprocesses of face recognition they identified in the same participant group. Second, I used frequentist group and individual case analysis alongside Bayesian analysis to seek converging evidence for, or against, the preregistered research hypotheses to mitigate the effects of low statistical power. Third, I extended previous research by also investigating qualitative experiences of DP (Chapter 6), and where possible, I analysed qualitative responses in the light of the participants' performance on experimental tasks of face memory and face perception.

Figure 2.5 Summary of research design Chapters 4 and 5



Note: Boxes on the right indicate the tasks used to test subprocesses 1-7. Reprinted with permission⁵ from Robotham and Starrfelt (2018).

⁵ from Elsevier from Neuropsychologia (121), Robotham and Starrfelt, Tests of whole upright face processing in prosopagnosia: A literature review, Pages 106-121, Copyright (2018),

Chapter 3: Evidence for different visual processing strategy for non-face stimuli in developmental prosopagnosia

Abstract

We assembled a test battery to investigate developmental prosopagnosia (DP), a neurodevelopmental syndrome resulting in severe face recognition difficulties. To screen for general cognitive deficits that could explain poor face test performance, participants completed a fluid reasoning task using abstract shapes. This initial screening showed that DPs ($n = 21$) were more accurate than neurotypical controls ($n = 90$) but significantly *slower*, suggesting *speed-accuracy trade-off*. To address this, we calculated the Balanced Integration Score and found no group differences, highlighting that DPs clearly adopted a *different strategy* from controls. DPs' longer response times (RT) on face tasks vs controls are commonly interpreted as evidence of impairment and of lengthy, atypical featural face processing. Our data suggest this interpretation may be unreliable since clear RT differences were also observed in two non-face tasks where DPs showed no accuracy impairment. Instead, slower RTs appear to reflect a strategy shift in DPs.

3.1 Background

Developmental prosopagnosia (DP) is a neurodevelopmental syndrome resulting in severe face recognition problems due to the visual mechanisms for face processing having failed to develop (Duchaine & Nakayama, 2006). The prevalence of DP is estimated at around 2 - 2.9% in adults (Bowles et al., 2009) and between 1.2% and 4% in children (Bennetts, Murray, et al., 2017).

3.1.1 Evidence for different processing strategies for faces

The literature suggests that DP is a heterogenous condition both in presentation and severity. It is currently unclear whether DP simply represents the bottom end of the face processing ability spectrum or whether there are *qualitative* differences in how individuals with DP (DPs) and typical perceivers encode and recall faces.

Although recognition accuracy is usually the main outcome measure of interest when investigating DP, differences in response time (RT) between DPs and typical perceivers are increasingly considered to be informative. For example, studies using the standardised Benton Face Recognition Test (BFRT; Benton et al., 1983) reported that

DPs can often perform within typical accuracy limits when the task has extended or unlimited presentation time (B. C. Duchaine & Nakayama, 2004; Nunn et al., 2001). By contrast, in other studies using a revised BFRT with a shortened presentation time, (Rossion & Michel, 2018) DPs performed less accurately than typical perceivers (Mishra et al., 2021; Murray et al., 2022). It is posited that a longer presentation time allows DPs to process faces in a laborious feature-by-feature manner, resulting in comparable accuracy to typical perceivers (Fysh & Ramon, 2022; Mishra et al., 2021; Murray et al., 2022). Conversely, short presentation times ‘force’ fast, holistic processing that is challenging for DPs, resulting in low recognition accuracy and making RT a useful additional measure of face processing in DP.

3.1.2 Do these different strategies extend beyond processing of faces?

The specificity of DP as a face selective condition independent of other object processing impairments remains a key question. Geskin and Behrmann (2018) reviewed published data on 238 cases of DP and found that most DPs (80% or 62% depending on the method used) showed evidence of concurrent object recognition impairment from accuracy and/or RT data. They argued that this suggests a shared underlying mechanism for face and object processing. However, the many responses to the review (summarised and responded to in Behrmann and Geskin, 2018) illustrate the current lack of consensus on this topic. Several researchers argue that there may be a subset of DPs with face-specific impairments (Dalrymple et al., 2017; Eimer, 2018); others that prosopagnosia should be considered an independent disorder that may co-occur with other forms of agnosia (Gray & Cook, 2018).

It is also not clear whether these processing differences extend beyond faces, objects with clear semantic information, and tests using within-class exemplars (e.g., cars, bicycles) to *objects without associated semantic information* (e.g., abstract shapes). Indeed, abstract shapes are widely used in fluid intelligence tests (e.g., matrix reasoning). It is widespread practice to administer some form of intelligence test when investigating face recognition ability despite there being no evidence that poor face recognition is associated with lower IQ (Connolly et al., 2019, Davis et al., 2011; Jones and Tranel, 2001; Wilmer, 2017) for two reasons. Firstly, to rule out a more general cognitive deficit as an explanation for impaired face processing test performance and/or, secondly, to match typical perceivers and DPs on IQ.

Here, we briefly report results from a fluid reasoning task, the Matrix Reasoning Item Bank (MaRs-IB, Chierchia et al., 2019) which formed a part of a comprehensive neuropsychological test battery administered to 111 participants to examine the underpinnings of DP across the lifespan. Full analysis of the test battery is underway and will be reported in another publication [Chapters 4 and 5]. The main objective of this brief report is to highlight evidence of a strategy difference in DP that extends to non-face stimuli. This finding could stimulate the DP research in new directions and offer a new perspective and new training regimes to help DPs navigate everyday life.

3.2 Method

3.2.1 Participants

Participants were 111 individuals: 21 DPs (15 women and 5 men; range = 8 - 71 years) and 90 age-matched neurotypical controls (51 women and 39 men; range 6 - 74 years, see Table 1). All reported normal or corrected-to-normal vision. Normal low-level vision was confirmed by performance on five subtests of the Birmingham Object Recognition Battery (Riddoch & Humphreys, 1993). We excluded participants who experienced major head injury at any time, a mild head injury or a concussion within 12 months or had a neurodevelopmental or major psychiatric condition diagnosis. Participants from the UK, Ireland and the USA were recruited through social media, media coverage, personal networks, and prosopagnosia support groups and received a £10 gift voucher upon completion.

Table 3.1 Participant demographics

Age Group	DP		Control	
	<i>n</i>	Mean age (<i>SD</i>)	<i>n</i>	Mean age (<i>SD</i>)
6-9 years	1 (F)	8	14 (7F; 7M)	7.9 (1)
10-13 years	1 (M)	10	19 (9F; 10M)	11.5 (0.96)
14-35 years	7 (5F, 2M)	24.6 (6.3)	20 (11F; 9M)	23.3 (6.9)
35-59 years	7 (6F, 1M)	48.6 (5.1)	21 (15F; 6M)	50.6 (6.0)
60-74 years	5 (3F; 2M)	66.6 (4.2)	16 (9F; 7M)	65.7 (4.8)
Total	21 (15F; 6M)	41.1 (20.1)	90 (51F; 39M)	32.3 (22.5)

For this initial screening, participants were classified as DP if they reported lifelong difficulty recognising familiar faces and scored ≥ 61 on the PI-20 (Shah, Gaule, Gaigg, et al., 2015), a cut off recommended by Tsantani et al., (2021). For children and young people aged 6-17 years we required parental report of face recognition difficulties classified as a score $< -1.7 SD$ below the control group mean on a questionnaire comprising 32 items: 17 items from the PI-20 (Shah et al., 2015), rephrased where necessary to make it suitable for reporting a third party's ability, and 15 additional items identified as hallmarks of prosopagnosia for non-experts (Murray et al., 2018). Inclusion and exclusion criteria and case histories were confirmed through interviews. DP classification using objective face memory measures is reported by Lowes et al 2024⁶.

For completeness, the test scores later used to classify participants [Chapter 4] are reported below. At the time of writing this paper/chapter, three participants (AF010, AF019 and AF020) reported in Chapters 4 and 5 had not yet participated so their data are not included here.

⁶ As reported in Lowes et al 2024a [Chapter 4], despite subjective face recognition difficulties, four of the 21 self-reported DPs in this chapter showed no consistent objective difficulties on four measures of face memory. Despite this, participants CF007 and CF008 completed an additional personally-familiar face test and failed to recognise friends and family suggesting that both had impaired familiar face recognition, despite typical unfamiliar face recognition.

Participant scores used for classification in Chapter 4

ID	Age	Self report	Old New Faces	FFT	CFMT	CBMT	CMT Difference	Global face memory	Face memory measures	
									No. impaired (of 4)	Severe impairment on ≥ 2 measures?
Accuracy										
AF019	68	-7.75	-3.88	-5.22	-1.64	0.84	-2.48	-3.31	4	yes
AF002	25	-7.28	-0.96	0.41	-1.53	0.11	-1.64	-0.93	2	no
CF008	10	-7.18	0.11	-3.33	-0.78	0.50	-1.27	-1.32	2	no
AF016	23	-7.01	-1.35	-2.63	-0.83	-0.35	-0.48	-1.32	2	no
AF017	29	-6.60	-1.74	-0.04	-1.53	1.36	-2.89	-1.55	3	yes
AF006	54	-6.16	-2.70	-1.58	-2.09	-0.41	-1.67	-2.01	4	yes
AF003	55	-6.05	0.88	-1.52	-1.55	0.87	-2.42	-1.15	3	no
AF018	49	-5.72	0.88	-4.30	-0.91	0.08	-0.99	-1.33	1	no
CF059	16	-5.59	NA	-0.58	-3.45	-2.84	-0.60	-1.54	1 of 3	no*
AF009	42	-5.17	0.88	-0.63	-1.23	0.77	-2.00	-0.74	2	no
AF010	53	-5.17	-3.84	-2.63	-3.05	-0.12	-2.93	-4.62	4	yes
CF007	8	-5.15	0.30	-1.74	0.90	1.06	-0.16	-0.18	1	no
AF007	31	-4.95	-0.56	0.34	-0.53	1.02	-1.54	-0.57	1	no
AF008	71	-4.78	-0.17	-1.37	-0.52	0.84	-1.36	-0.85	2	no
AF013	44	-4.73	0.88	0.92	0.92	0.67	0.24	0.74	0	no
AF022	64	-4.65	-0.17	0.72	0.04	0.75	-0.70	-0.03	0	no
AF060	61	-4.65	-0.92	-1.92	-0.42	0.84	-1.27	-1.13	2	no
AF001	51	-4.62	-0.91	-0.24	-1.23	0.38	-1.60	-1.00	2	no
AF004	31	-4.54	0.62	-1.87	-0.93	1.58	-2.51	-1.17	2	yes
AF099	70	-4.38	-3.14	NA	0.33	0.09	0.23	-0.86	1 of 3	no*
AF075	45	-3.64	0.28	-1.40	-0.80	-0.02	-0.78	-0.67	1	no
AF021	50	-3.09	0.88	-2.03	-1.55	0.97	-2.52	-1.31	3	yes
AF098	67	-2.76	-3.14	-1.45	-1.74	0.00	-1.74	-2.01	4	yes
CF005	17	-0.66	-1.35	-1.98	-2.34	1.36	-3.70	-2.34	4	yes
Proportion impaired			0.35	0.65	0.50	0.04	0.67	0.58		0.33
Balanced Integration Scores										
AF019	68	-7.75	-6.92	-5.22	-1.72	0.02	-1.74	-3.90	4	yes
AF002	25	-7.28	-4.35	0.41	-2.70	-0.37	-2.33	-2.24	3	yes
CF008	10	-7.18	0.37	-3.33	-0.47	-0.48	0.01	-0.86	1	no
AF016	23	-7.01	-2.11	-2.63	-1.52	0.35	-1.87	-2.03	4	yes
AF017	29	-6.60	-3.17	-0.04	-0.17	2.50	-2.67	-1.51	2	yes
AF006	54	-6.16	-2.42	-1.58	-1.16	0.18	-1.34	-1.63	4	no
AF003	55	-6.05	-0.60	-1.52	-2.25	0.04	-2.29	-1.66	3	yes
AF018	49	-5.72	-1.17	-4.30	-2.88	-1.65	-1.23	-2.40	4	yes
CF059	16	-5.59	NA	-0.58	-5.53	-3.07	-2.46	-2.85	2 of 3	yes
AF009	42	-5.17	0.03	-0.63	-5.56	-0.85	-4.71	-2.72	2	yes
AF010	53	-5.17	-12.13	-2.63	-3.48	0.89	-4.37	-5.65	4	yes
CF007	8	-5.15	0.23	-1.74	1.71	-0.91	2.63	0.71	1	no
AF007	31	-4.95	-2.88	0.34	-4.67	-0.85	-3.82	-2.76	3	yes
AF008	71	-4.78	-5.72	-1.37	-1.23	-2.19	0.96	-1.84	3	yes
AF013	44	-4.73	-0.80	0.92	0.79	0.49	0.29	0.30	0	no
AF022	64	-4.65	-0.28	0.72	-0.51	-0.83	0.32	0.06	0	no
AF060	61	-4.65	-2.94	-1.92	-0.07	0.77	-0.85	-1.44	2	yes
AF001	51	-4.62	-2.45	-0.24	-0.35	1.92	-2.26	-1.33	2	yes
AF004	31	-4.54	-0.67	-1.87	-2.20	0.85	-3.05	-1.95	3	yes
AF099	70	-4.38	-7.53	NA	0.22	0.11	0.11	-2.40	1 of 3	no*
AF075	45	-3.64	-3.07	-1.40	-3.83	-2.37	-1.46	-2.44	4	yes
AF021	50	-3.09	-1.38	-2.03	-2.03	1.08	-3.11	-2.14	4	yes
AF098	67	-2.76	-3.84	-1.45	-1.02	-0.05	-0.97	-1.82	3	no
CF005	17	-0.66	-2.27	-1.98	-2.73	2.18	-4.90	-2.97	4	yes
Proportion impaired			0.70	0.65	0.67	0.17	0.67	0.83		0.71

Note. FFT = Famous Faces Test, CFMT = Cambridge Face Memory Tests, CBMT = Cambridge Bicycle memory Test, CMT difference = CFMT – CBMT, Global face memory = standardised mean average. BIS was not calculated for the FFT so the results are the same as for accuracy alone but are shown again in the second table for ease of comparison. Major

impairment $z \leq -1.7$ shown in red, mild impairment $-1 \geq z \geq -1.69$ shown in yellow. All scores are standardised and centred on age-matched control means. The two farthest right columns show firstly the number of independent face processing measures on which a participant was impaired (mild or major) and secondly whether the participant showed an impairment of at least 2 *SD* below control means on at least two of the four face memory measures, * indicates that participant completed three rather than four tasks.

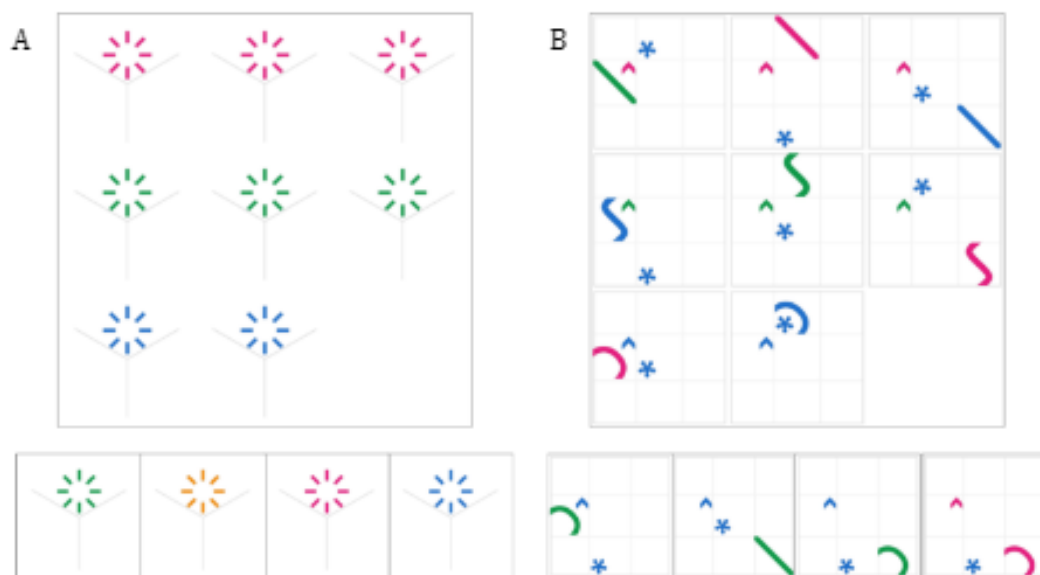
Approval for the study was granted by the University of Stirling General University Ethics Panel (project code 204). Informed consent and child assent were obtained from all participants and research was conducted in accordance with the principles of the Declaration of Helsinki.

3.2.2 Materials and procedure

Experiment 1

Participants completed the MaRs-IB (Chierchia et al., 2019) - a computerised non-verbal reasoning task suitable for children and adults. Chierchia and colleagues' (2019) results correlated well with another test of fluid intelligence, the International Cognitive Ability Resource (Condon & Revelle, 2014) and were broadly replicated in an online study (Nussenbaum et al., 2020). Participants chose one of four possible abstract shapes that will correctly complete a 9 x 9 grid of shapes (Figure 3.1).

Figure 3.1 Example items from the Matrix Reasoning Item Bank (MaRs-IB)



Note: Example items from the MaRs-IB. Panel A shows a simple item containing a one-relational change (i.e. only the colour changes) and answer options. The fourth option is the correct option. Panel B shows a harder item containing a three-relational change (i.e.

shape, colour and position change. The third option is the correct option. Figure reproduced with permission.

Participants have up to 30 s to complete each trial. After 25 s, a countdown clock appears, indicating that 5 s remains until the next trial. Participants are instructed to “*try to be as fast and accurate as you can be*” and, if they are unable to solve the puzzle, to guess. They are informed that they “*will have 30 seconds to complete each puzzle*”. The full task instructions are shown in Appendix 1. We administered the colour-blind friendly version of the test to all participants. The test begins with three practice trials that must be completed correctly. Feedback is provided after each trial. All participants see trials in the same order and the test ends *either* after 8 minutes *or* when the maximum possible 80 trials have been completed, whichever occurs first. Trials broadly increase in difficulty as the task progresses. Crucially, the design means that participants may complete a different number of trials (the range in our study was 21-80 and 24-79 among controls and DP respectively, see Table 3.1). Therefore, one participant score of 48/80 and another of 24/40 would both result in the *same accuracy* score of 60%. Similarly, in the maximum allowable time of 8 minutes one participant who completes the maximum possible 80 trials with 40% accuracy and another who adopts a different speed-accuracy trade off strategy and completes only 40 trials but with 80% accuracy would both achieve the *same number correct* score of 32. For a full description of the task and validation data see Chierchia et al. (2019). Following Chierchia et al. (2019) we calculated and report the average of each participant’s *median* item RT to account for the fact that participants who complete more trials will attempt a higher proportion of relatively more difficult items.

Cambridge Bicycle Memory Test (CBMT, Dalrymple, Garrido, et al., 2014)

Participants study six target bicycles each from three different viewpoints. Cropped and greyscale bicycles from which logos are removed are presented for 3 s. The test stage uses a three alternate forced choice format, comprising one target and two distractor bicycles. In the introduction test phase, each target bicycle is tested with three identical viewpoints. In the second test phase, after review, each target bicycle is shown in a novel viewpoint from that learned in study and, finally, the noise test section introduces novel views of the target bicycle with added Gaussian noise. There are 72 trials in total and chance = 33%. RT is unlimited.

Participants and parents of children under 18 years old received written instructions in advance of testing in addition to the on-screen instructions. Parents supervised their children and could assist with explaining the task but were instructed not to help with responses. The test was administered using the online testing platform Gorilla (www.gorilla.sc) with Google Chrome as the recommended browser.

3.2.3 Analysis and pre-processing

The exploratory analysis for the full study was preregistered at <https://osf.io/qne8d>. We centred z scores on control age group means because accuracy and RT may vary on some tasks as a function of age. This age group standardisation allowed us to classify DP participants' performance relative to controls across the full age range since here the z scores quantify participants' performance *relative to their own age group*. Following previous DP literature, age groups were predefined as 6-9, 10-13, 14-35⁷, 36-59, and 60-74 years and descriptive statistics for each age group are shown in Appendix 1 Table S1 in Supplementary Material. The DP and control groups showed no significant difference in age ($X^2 = 50.05$, $p = .316$).

Chierchia and colleagues (2019) recommend four dependent variables (DVs): i) accuracy (proportion of trials completed in the allowable time that were correct (8 mins or when the maximum 80 items were completed, whichever came first)); ii) median response time (RT) on correct trials; iii) number of trials attempted (correct and incorrect), and iv) inverse efficiency (IE; a measure that considers both RT and accuracy). However, the use of IE is not recommended when mean accuracy is less than around 90%, or when there is a speed-accuracy trade off (Bruyer & Brysbaert, 2011), and both are true in the case of the MaRs-IB (slower participants were more accurate whether considering the sample as a whole ($r(109) = .772$, $p < .001$ or just controls ($r(88) = .735$, $p < .001$)). We thus elected to use the Balanced Integration Score (Liesefeld & Janczyk, 2019, 2022)). BIS is an integrated measure that adjusts accuracy to account for RT whilst controlling for speed-accuracy trade-offs and is calculated as $Z_{\text{accuracy}} \text{ minus } Z_{\text{RT}}$, (RT for correct trials only). Data were analysed in R (R Core Team, 2021) using R Studio 2021.09.1. We conducted Welch's t tests which are recommended in independent subject designs with different experimental group sizes and/or unequal group variance (Ruxton,

⁷ Bowles et al. 2009 used 18-35 years for young adults but because there are ceiling effects above age 13 years on the CFMT Kids (Dalrymple et al., 2012) and other studies have successfully administered adult versions of face processing to adolescents age 14-15 years (Bate et al., 2015) we selected 14-35 years as our young adult age band.

2006), and also Bayesian independent samples *t* tests using the default Cauchy prior with a scale of 0.707. Data and code are available at <https://osf.io/btyud>.

Data from two child controls were excluded from the study prior to analysis because they performed at chance on this test *and* well below chance on a second test suggesting either an inability to follow instructions or suboptimal effort, or both.

3.3 Results

Accuracy and RTs in typical perceivers' data were normally distributed. There were no gender differences in the data ($p > .05$) or in the gender distribution between the DP and control groups ($X^2 = 1.54$, $p = .215$) and so we collapsed the data across gender to increase statistical power (Table 3.2).

Table 3.2 Descriptive statistics Matrix Reasoning Item Bank

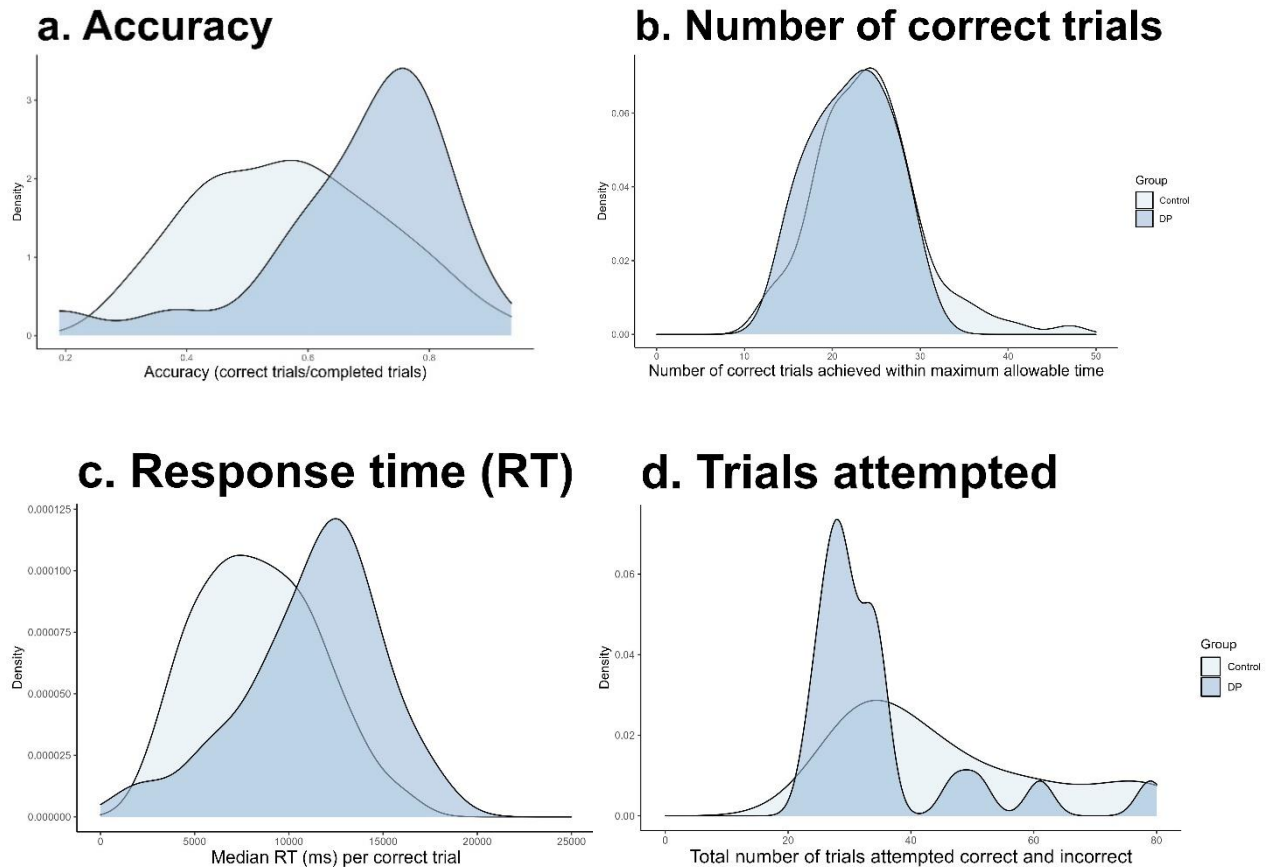
	Group	<i>M</i>	<i>SD</i>	Minimum	Maximum
Trials attempted	Control	45.18	16.71	21	80
	DP	35.24	13.66	24	79
Number correct trials	Control	24.04	5.98	13	47
	DP	22.24	4.49	15	30
Accuracy <i>z</i> score	Control	0	0.98	-1.99	2.72
	DP	0.54	1.11	-2.97	1.71
RT <i>z</i> score	Control	0	0.98	-1.91	2.28
	DP	0.73	1.12	-2.23	2.30
Balanced Integration Score	Control	0	0.71	-1.68	1.32
	DP	-0.2	0.63	-1.38	0.77

Note: *Z* scores are calculated centred on the age group means to ensure that the performance of potential DPs was always compared with age-matched typical controls. Accuracy = number of correct trials/number of trials completed and RT = average of each participant's median RT on correct trials).

DPs were, on average, more accurate in the MaRs-IB ($M = .680$, $SD = .162$, [95% CI .611 - .749]) than controls ($M = .572$, $SD = .154$ [95% CI .540 - .603]). Independent samples *t* tests on the standardised scores showed that this difference approached

significance Welch's $t(27.6) = 2.03$, $p = .052$, Cohen's $d = 0.51$ (see Figure 3.2a). However, the Bayesian t test indicated only weak support for this difference ($BF_{10} = 1.94$), in other words, the alternative hypothesis (a true difference exists between groups) is around twice as likely as the null hypothesis given the data. Trials increase in difficulty as the test progresses, so to exclude the possibility that the DP group's higher numerical accuracy was simply due to the fact this group, on average, completed a lower number of trials vs controls and thus an easier overall set of trials, we repeated the analysis using only trials 6-25 i.e. excluding the more difficult later trials that not all DPs attempted. Results were again that the DP group was more accurate ($M = .68$, $SD = .16$) than controls ($M = .57$, $SD = .15$); this time the group difference was significant, $t(28.8) = 2.16$, $p = .040$, Cohen's $d = 0.53$, although Bayesian analysis provided only anecdotal support for a true group difference, $BF_{10} = 2.09$. The observed accuracy difference is therefore not due to DPs completing a higher proportion of easier trials. The important finding, confirmed by this additional analysis, is that DPs were not impaired on the MaRs-IB and had similar or even slightly better accuracy than controls, meaning that any group RT differences on this task are not due to impaired accuracy.

Figure 3.2 MaRs-IB score distribution by group (DP or control).



Note: Panel a shows accuracy scores (i.e. proportion of trials completed in the maximum allowable time that were correct. Chance = 0.25, range was 0.19 - 0.94). Panel b shows number of correct trials completed in the maximum time allowed which was 30000 ms per trial, overall maximum test duration was 8 minutes or when all 80 trials were completed - whichever occurred first. Maximum possible score = 80, observed range was 13 – 47. Panel c shows average median RT in ms (correct trials only). Panel d shows the total number of trials attempted within the allowable 8 minutes.

3.3.1 Response time

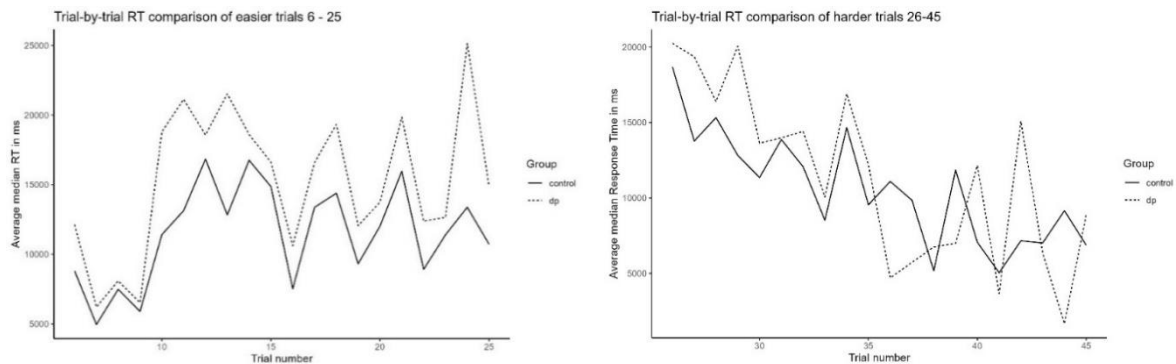
In this task there was a strong positive relationship between accuracy and response time, whether considering the sample as a whole ($r(111) = .762, p < .001$), or just controls $r(90) = .730, p < .001$), that is on average participants who took longer to respond were also more accurate. As shown in Figure 3.2 above, the different pattern of results observed between accuracy (proportion correct) and number correct (raw score) appears to be driven by DPs adopting a different strategy to controls. Namely, as shown in Figure 3.2c, on average DPs took significantly longer per correct trial (median item RT 11219 ms) than controls (8380 ms). An independent samples t test comparing group standardised RTs showed this difference was significant Welch's $t(27.6) = 2.77, p = .010$, Cohen's $d = 0.70$ (Figure 3.2a) and DPs were likely therefore more accurate due to the speed-accuracy trade off on this task. Importantly, this difference was not due to differences in age distribution between the groups since the z scores were centred on age group control means. Further information on age group level performance is provided in Appendix 1 Figure S1. An independent samples Bayesian t test ($BF_{10} = 11.3$) provided strong support for the alternative hypothesis. In other words, the alternative hypothesis (a true difference in RT exists between the DP and control groups) is more than 11 times as likely as the null hypothesis given the data.

In the MaRs-IB, trials are presented in a fixed order and, as with Raven's Progressive Matrices, difficulty broadly increases as trials progress. Unfortunately, there is no item level difficulty data for the colour-blind version of the task that we used. To check that RT differences were not due to DPs simply completing a higher proportion of easier trials, we therefore separately compared mean group performance on 20 easier trials (trials 6-25, we excluded trials 1-5 which are very easy and mainly used for familiarisation) vs the following 20 trials (numbers 26-45) which are more difficult. As can be clearly seen from Figure 3.3, even on the easier trials, the DP group's mean item

RT ($M = 15732$ ms, $SD = 4841$) was consistently and significantly slower than controls' ($M = 12403$ ms, $SD = 3276$), $t(33.4) = 2.55$, $p = .016$, Cohen's $d = 0.81$, $BF_{10} = 3.63$. The observed group RT differences cannot therefore be accounted for by DPs completing a higher proportion of relatively easier trials. On the 20 harder trials, although numerically DPs were again slower than controls, this difference was smaller and NS, $p = .666$, Cohen's $d = .14$, $BF_{10} = .34$. Similar analysis comparing standardised accuracy on the 20 easier trials found that DPs ($Mz = 0.67$, $SD = 1.05$) were significantly *more accurate* than controls ($Mz = 0.13$, $SD = 0.99$), $t(28.8) = 2.156$, $p = .040$, Cohen's $d = .53$, $BF_{10} = 2.09$.

Taken together, these results show that the DP group's higher accuracy and longer RTs on the matrix reasoning task is not simply explained by the task design whereby faster participants complete a higher proportion of more difficult trials. This additional exploratory analysis was suggested during peer review and was not pre-registered. An ANOVA was also suggested but was not feasible due to small numbers of DP correctly completing some of the later trials, and a large variation of overall number of correct responses across trials.

Figure 3.3 Comparison of RT difference between the DP and control groups across 20 easier trials and 20 more difficult MaRs=IB trials.

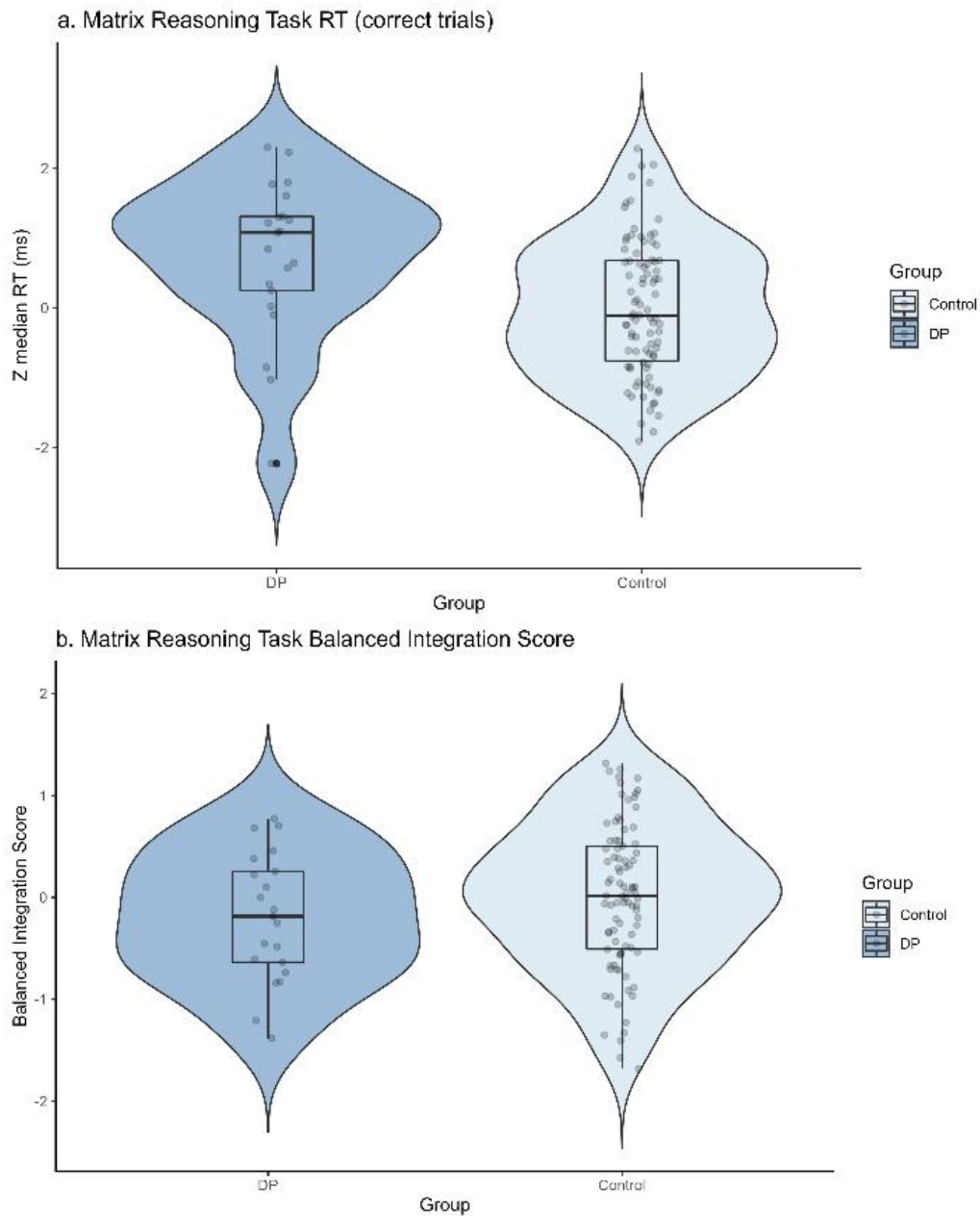


3.3.2 Balanced Integration Score (BIS)

To confirm whether DPs' higher group-level accuracy was indeed driven by differences in the speed accuracy trade-off strategies adopted by DPs and controls, we calculated the BIS (Liesefeld and Janczyk, 2019), a measure that can be interpreted as accuracy adjusted for RT thereby controlling for differential speed-accuracy trade-offs. The authors reported that drift diffusion modelling using simulated data showed that BIS was suitable for tasks with varying levels of mean accuracy and decision thresholds

(Liesefeld and Janczyk, 2019). As a reminder, the z scores used to calculate the BIS were centred on control age group means thus *all DPs are compared to typical controls of their own approximate age* which controls for any differences that might be observed in speed-accuracy trade-offs in different age groups. As shown in Figure 3.4b below, no significant difference in BIS was observed between DPs and controls, Welch's $t(33.2) = -1.27$, $p = .212$. Cohen's $d = -0.30$. An independent samples Bayesian t test ($BF_{10} = 0.45$) provided moderate support for the null hypothesis. In other words, as expected, once RT and *differential speed accuracy trade-offs* had been accounted for, there was no difference in accuracy between the DP and control groups. The absence of group difference in Balanced Integration Scores strongly suggests that the unexpected group accuracy differences were driven by different speed-accuracy trade-offs between the groups. To confirm this we also inspected and report performance on an object recognition task that was part of our test battery, the Cambridge Bicycle Memory Task (Dalrymple, Garrido, et al., 2014).

Figure 3.4 MaRs-IB group difference



Note: Panel a shows the difference in average standardised median item RT between groups and panel b the difference in balanced integration score (calculated as Z_{accuracy} minus Z_{RT}) between the control and DP groups. Each dot represents a single data point, the box shows the interquartile range (IQR) and the mid line indicates the group median RT in ms. The end of each whisker line represents 1.5 x the IQR. Maximum allowable time per trial = 30000 ms.

Our test battery also included an object recognition task, the Cambridge Bicycle Memory Test, so to check whether these apparent strategy differences extended to objects, we next used independent t tests to compare DP and control group accuracy, RT and BIS. As before, we computed standardised scores centred on age-group specific

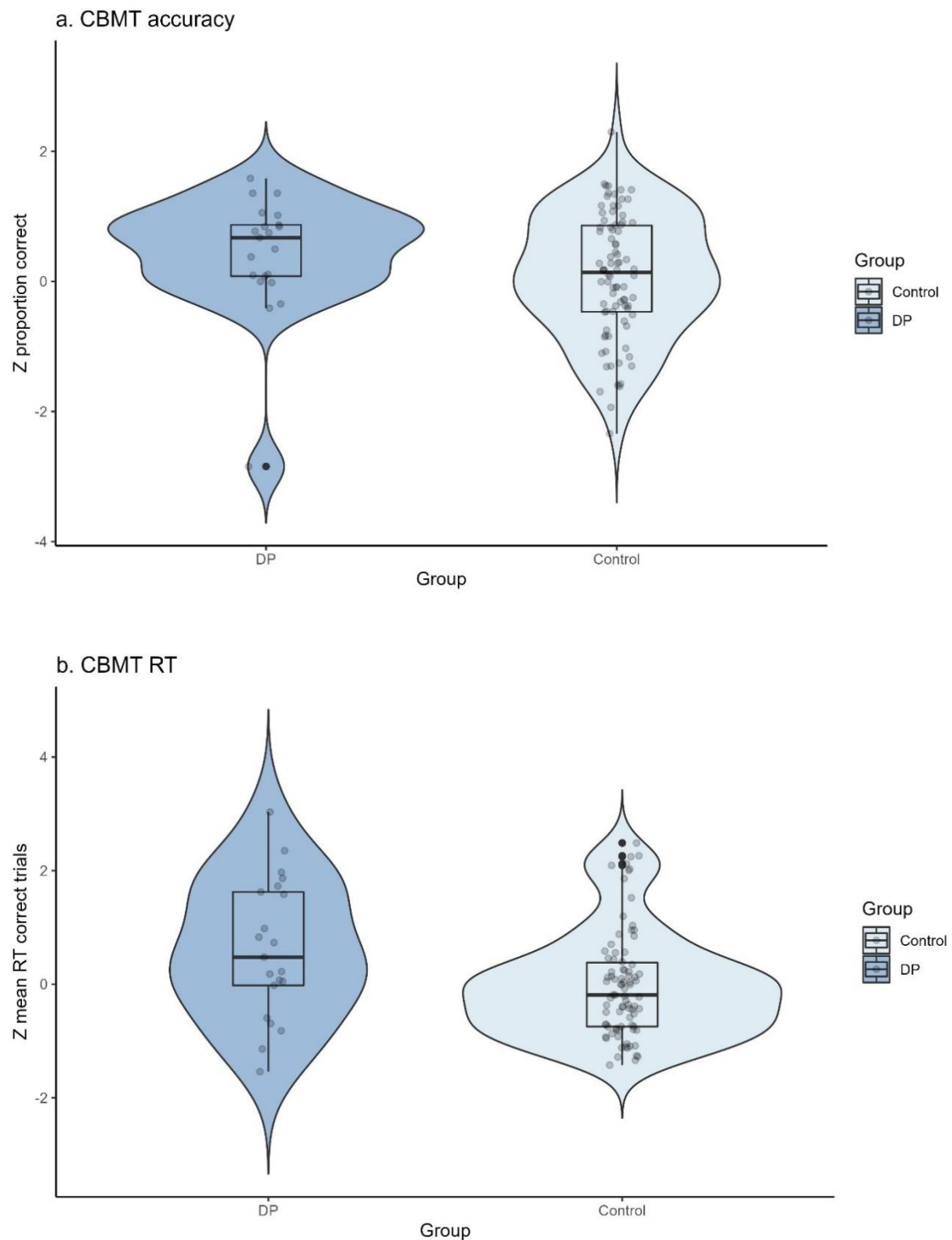
control group means to account for age differences. Four control participants did not complete the CBMT. To maximise power, participants with missing data were retained, and analyses were conducted using pairwise deletion⁸.

As shown in Figure 3.5, the DP group was, on average, slightly more accurate ($Z_{\text{accuracy}} M = 0.41, SD = 0.93$) than controls ($Z_{\text{accuracy}} M = 0.08, SD = 0.95$) but the difference was ns ($t(31.1) = 1.47, p = .151$, Cohen's $d = 0.36$, $BF_{10} = 0.61$). As with the MaRs-IB, the DP group was significantly slower ($Z_{RT} = 0.61, SD = 1.22$) than age-matched controls ($Z_{RT} = -0.02, SD = 0.94$) with moderate evidence for a true group difference, $t(26.1) = 2.20, p = .037$, Cohen's $d = 0.58$, $BF_{10} = 4.11$. When BIS was calculated to control for differences in speed-accuracy trade-offs (DP = -0.20, $SD = 1.44$; controls = 0.09, $SD = 1.26$), group performance did not differ, $t(27.9) = -0.86, p = .398$, Cohen's $d = -0.22$, $BF_{10} = 0.36$ indicating that the previously observed group RT differences on the CBMT were again a result of strategy differences between DPs and controls.

Finally, after removing one outlier (DP) with unusually fast response time and low accuracy, we conducted Pearson product-moment correlations to assess the relationship between standardised RT and accuracy performance on the MaRs-IB and CBMT (Table 3.3). Results showed a significant moderate positive correlation for RT when examining the sample as a whole and the control group only. In the DP group we observed a stronger relationship which trended toward significance $p = .057$, likely due to reduced power given the much smaller sample size. Nevertheless, the strength of the RT correlations on these quite different paradigms supports the idea that RT differences are related across tasks - particularly since accuracy scores on these same tasks were unrelated in both the DP and control groups, and also the sample as a whole.

⁸ Also known as available-case analysis. To aid statistical power, group analyses were run using all available data, even if a participants had missing data on another task.

Figure 3.5 Group difference CBMT



Note. CBMT = Cambridge Bicycle Memory Test. Panel a shows the difference in average standardised accuracy and panel b the group difference in standardised RT between the control and DP groups. Each dot represents a single data point, the box shows the interquartile range (IQR) and the mid line indicates the group median RT in ms. The end of each whisker line represents 1.5 x the IQR.

Table 3.3: Correlation tables matrix reasoning and bicycle memory tasks

Full sample df =104				
Variable	RT MaRs-IB	RT CBMT	MaRs-IB accuracy	CBMT accuracy
RT MaRs-IB	-			
RT CBMT	.39***	-		
MaRs-IB accuracy	.77***		-	
CBMT accuracy		.16	.14	-
DP group df =18				
Variable	RT MaRs-IB	RT CBMT	MaRs-IB accuracy	CBMT accuracy
RT MaRs-IB	-			
RT CBMT	.43	-		
MaRs-IB accuracy	.74***		-	
CBMT accuracy		.12	-.06	-
Control group df = 84				
Variable	RT MaRs-IB	RT CBMT	MaRs-IB accuracy	CBMT accuracy
RT MaRs-IB	-			
RT CBMT	.31**	-		
MaRs-IB accuracy	.75***		-	
CBMT accuracy		.12	.11	-

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

MaRs-IB = Matrix Reasoning Item Bank, CBMT = Cambridge Bicycle Memory Test.
RT = standardised response time, accuracy = standardised proportion correct.

3.4 Discussion

Participants who reported severe face recognition difficulties (DPs) and age-matched typical perceivers completed an abstract reasoning test to screen for wider cognitive difficulties that might account for atypical face recognition performance e.g., inability to follow instructions. DPs were, on average, *more accurate* than controls (that is, of the trials they attempted, a higher proportion were correct, (importantly this was also the case when considering only a subset of 20 easier trials) but, overall, both groups achieved a *similar number* of correct trials in the maximum allowable time. This difference in accuracy was driven by DPs spending longer per trial, despite the test instructions stressing both speed and accuracy and thus leaving participants to decide their own approach to how to attempt to balance these. The DPs clearly adopted a

different strategy to typical perceivers in this visual processing task that did not involve faces or objects with clear semantic information, and one that did not involve long-term memory. Once accuracy was adjusted to account for differential speed-accuracy trade-offs, no group difference was observable.

In short, DPs were, on average, more accurate owing to the speed-accuracy trade-off in this visual processing task. Importantly, typical perceivers' accuracy broadly replicated previous published data from an online sample (Nussenbaum et al., 2020) suggesting that our results cannot be accounted for by an unusual control sample.

Results from a bicycle recognition task provided converging evidence for a strategy difference. A very similar pattern of results was observed on both tasks; DPs were more accurate yet slower to make correct responses than controls. The significant positive relationship between RTs on the MaRs-IB and CBMT provides further support that RT differences are indeed related across very different paradigms.

The fact that the DP and control groups appear to have taken different approaches to speed-accuracy trade-offs is perhaps not unexpected. In their review of reaction time and individual differences, Draheim and colleagues (2019) point out that participants with different levels of ability are likely to respond differently to the same instructions, as are participants of different ages. There may also be within-participant differences observed. For example, participants may prioritise speed on easier trials and accuracy on harder trials (Liesefeld and Janczyk, 2019). Accounting for differential speed accuracy trade-offs between and within participants is therefore important more generally (Meyer et al., 2018) and is of critical importance in syndromes such as DP, but also other syndromes for which there is no genetic or clinical diagnostic test and diagnosis and/or classification is made on the basis of impaired performance on behavioural tests.

We have reported elsewhere (Lowes et al., 2024b) that many more individuals who reported face recognition difficulties were classified as DP once accuracy was adjusted for RT through use of Balanced Integration Scores (BIS) on four objective measures of face memory compared with accuracy scores alone (83% vs 58% of the potential DP group respectively). In other words, many of the potential DP group appeared initially to have below average but nevertheless typical face memory accuracy, but this was only achieved at the expense of abnormally prolonged RT. Once accuracy was adjusted to account for RT (using BIS), thus controlling for these differential speed-accuracy trade-offs, a far greater proportion showed a face memory deficit. Additionally, data from the Cambridge Face Memory Test (Duchaine & Nakayama, 2006), a test that uses the same

format as the CBMT and is considered the ‘gold standard’ for detecting prosopagnosia, showed clear evidence of different speed-accuracy trade-offs among DPs and controls: The most *accurate* controls were also *faster* but, conversely, among DPs the more *accurate* DPs were *slower*.

Our finding here is important because a common interpretation of abnormally prolonged RT in prosopagnosia (both acquired and developmental) is that it indicates a deficit in holistic face processing and that extended or unlimited presentation time allows individuals to process faces in a serial feature-by-feature manner (Fysh & Ramon, 2022; Mishra et al., 2021; Murray et al., 2022). This feature-by-feature strategy is atypical relative to neurotypical, rapid holistic face processing. While this may be true in some cases, a more parsimonious explanation may be that differences in strategy preference exists between DPs and controls since our results suggest that different strategies in DPs are *also* present for non-face stimuli that do not rely on holistic processing in the way faces are widely argued to.

We therefore caution against interpreting RT differences in DP as evidence of a deficit in holistic face processing because while these longer processing times may *facilitate* feature-by-feature processing of faces, they may not be driven by this. The long, although not unlimited, stimulus presentation times in the MaRs-IB and the unlimited RT allowed in the CBMT also produced a shift in strategy amongst DPs that crucially did not reflect a lack of ability on either of these tasks, quite the opposite since DPs were numerically (though not always statistically) *more* accurate than controls. For clarity, we do not claim that these interesting findings suggest any relationship between performance on these tasks and face processing more generally, instead we show that DPs demonstrated a preference to proceed more carefully and cautiously than controls across a range of cognitive tasks and that this strategy preference is unrelated to ability, or face processing. Our findings strengthen the case for RTs to be routinely inspected in face and other visual stimulus categories tests and, crucially, for speed accuracy trade-offs that may differ between participants and between groups, to be accounted for using an appropriate measure such as the Balanced Integration Score (Liesefeld and Janczyk, 2019).

Chapter 4: A new way of classifying Developmental Prosopagnosia: Balanced Integration Score (BIS)

Recap and present chapter

Due to the evidence for different speed-accuracy trade-offs in the DP and typical controls groups identified in Chapter 3, it was clear that such differences, should they also be observed on face processing tasks, could confound results and must therefore be controlled for. Chapter 3 showed that BIS appeared to a useful measure for this purpose. I had pre-registered a series of measures that I would use to classify participants as DP or controls and therefore I decided to directly compare how well the original accuracy measures, and the BIS measures, classified the same participants on the same task. The online experimental platform Testable automatically collects RT data, in addition to accuracy scores, making it easy to calculate BIS. In the present chapter I compare whether accuracy (proportion correct) or BIS yields an objective measure across four tests of face memory that better accounts for self-reported face recognition ability.

Abstract

Despite severe everyday problems recognising faces, some individuals with developmental prosopagnosia (DP) can achieve typical accuracy scores on laboratory face recognition tests. To address this, studies sometimes also examine response times (RTs), which tend to be longer in DPs relative to control participants. In the present study, 24 potential (according to self-report) DPs and 110 age-matched controls completed the Cambridge Face and Bicycle Memory Tests, old new faces task, and a famous faces test. We used accuracy and the Balanced Integration Score (BIS), a measure that adjusts accuracy for RTs, to classify our sample at the group and individual levels. Subjective face recognition ability was assessed using the PI-20 questionnaire and semi structured interviews. Fifteen DPs showed a major impairment using BIS compared with only five using accuracy alone. Logistic regression showed that a model incorporating the BIS measures was the most sensitive for classifying DP and showed highest area under the curve (AUC). Furthermore, larger between-group effect sizes were observed for a derived global (averaged) memory measure calculated using BIS vs accuracy alone. BIS is thus

an extremely sensitive novel measure for attenuating speed-accuracy trade-offs that can otherwise mask impairment measured only by accuracy in DP.

4.1 Introduction

Developmental prosopagnosia (DP) is a neurodevelopmental syndrome that manifests in severe face recognition problems due to the visual mechanisms for face processing having failed to develop (Duchaine & Nakayama, 2006). DP occurs despite normal vision and IQ and lack of obvious brain damage. Prevalence of DP is estimated at around 2% - 2.9% in adults (Bowles et al., 2009; Kennerknecht et al., 2006, 2008) and between 1.2% and 4% in children (Bennetts et al., 2017). DP is thus as common as severe dyslexia (~ 2%-4%, European Dyslexia Association, n.d.) and more common than autism (~0.6%, World Health Organisation, 2018) despite being relatively unknown. However, recent work by DeGutis and colleagues (2023) demonstrates that prevalence estimates vary considerably depending on the measures and cut-offs used.

4.1.1 Different approaches to classifying DP

No clinical definition of DP exists, instead it is usually classified by poor performance relative to neurotypical controls (usually 1.7 or 2 standard deviations (*SD*)) below the control mean) on at least one laboratory (i.e. objective) test of face processing alongside personal (i.e. subjective) report. Some studies argue that DP is an identifiable disorder with qualitative, as well as quantitative differences between DPs and controls (Behrmann et al., 2007; Bobak et al., 2017; Burns et al., 2014; Towler et al., 2018), others that it may simply represent the lower end of the face recognition (FR) ability spectrum (Bowles et al., 2009; Johnen et al., 2014). More generally, many studies have noted that DP is a heterogeneous condition, both in presentation and severity (Bobak et al., 2017; Corrow et al., 2016; Dobel et al., 2007; Duchaine & Nakayama, 2005; Wilcockson et al., 2020). This heterogeneity has even been observed among members of the same family (De Haan, 1999; Duchaine et al., 2007; Lee et al., 2010; Schmalzl et al., 2006). For example, Lee et al. (2010) reported in their study three members of the same family with impaired face memory, but only one additionally with impaired object recognition.

Overall, the literature to date contains many mixed findings. These could be due to genuine heterogeneity, perhaps accentuated by the small sample sizes typical in neuropsychology. However, it is possible that the competing findings may, in part, also be explained by the varied approaches taken by different research groups to classifying

DP (Bate & Tree, 2017) and by the range of different tests and measures used for classification and assessment of face processing in DP (for an overview see DeGutis et al., 2023; Robotham & Starrfelt, 2018). Broadly, the literature shows that approaches to categorisation differ across studies in three main, yet inter-related, ways. Firstly, in the selection of test(s) and associated cut off levels to be used; secondly, in the selection of the outcome measure(s) used to quantify test performance and, thirdly, in the inclusion and exclusion criteria used for both the DP and typical controls groups. As we show here, the DP heterogeneity might also arise from study participants adopting different strategies in approaching the tasks, i.e. by taking longer to complete it and thus potentially masking impairments when only accuracy is taken into consideration.

4.1.2 Different approaches to the selection of tests for classification

The criteria to classify DP vary between research groups, which is problematic for study and case comparison (Corrow et al., 2016). To address this, Dalrymple and Palermo (2016) recommended that, in addition to subjective reports of face recognition difficulties, individuals should exhibit impairments on at least *two* objective tests of face processing (cf., Burns et al., 2022 who argue that subjective report alone is sufficient for classification). Notably these guidelines did not specify that these should test face recognition (as opposed to face perception) specifically. This may be problematic because several studies report that as many 50% of DPs show no face perception impairment (e.g., Dalrymple et al., 2014), meaning that perceptual tests may therefore not always be suitable for classifying DP. Indeed, DeGutis and colleagues (2023) have recently proposed that *two tests of face recognition* specifically should be used for classification, but this has yet to become the norm. Since the 2016 guidelines were published, other researchers have gone further, arguing that converging evidence of impairment across *multiple* tests of face processing would offer stronger evidence for classifying potential DPs (Bate & Tree, 2017), and that participants scoring more than 1 *SD* below the mean on *two or more tests* should be classified as impaired (Mishra et al., 2021; Stumps et al., 2020) since this approach is common in other areas of neuropsychology such as mild cognitive impairment (Sachdev et al., 2014). At the other end of the face recognition ability spectrum, it has also been argued that super recognizers should be assessed using converging evidence from *multiple* tests (Bobak et al., 2023). The guidelines for the best combination of tests predicting real-life face recognition are

constantly evolving in both superior (Mayer & Ramon, 2023) and typical face recognition (Bobak et al., 2023).

4.1.3 Different approaches to the selection of outcome measures to quantify test performance

Impaired accuracy is the outcome measure traditionally used to classify DP. However, it has been shown that DPs can sometimes perform within typical accuracy limits when the task has extended or unlimited presentation time (Albonico et al., 2017; Dobel et al., 2007; Duchaine & Nakayama, 2004) leading to recommendations for both accuracy and response time (RT) to be considered (Fysh & Ramon, 2022; Stacchi et al., 2020), typically using Inverse Efficiency Scores (IES, Townsend et al., 1983). Unfortunately, the IES is *only suitable* when mean accuracy is above around 85%, a level not commonly found in DP research, and/or when there are no accuracy vs RT trade-offs (Bruyer & Brysbaert, 2011). Pertinently, such speed-accuracy trade-offs are observed in common face perception tests (Stacchi et al., 2020) and have been shown to differ between not only DPs and controls but also DPs and participants with acquired prosopagnosia (Behrmann et al., 2005; Fysh & Ramon, 2022).

A better single measure that combines accuracy and RT is the Balanced Integration score (BIS, Liesefeld & Janczyk, 2019, 2022). Using drift diffusion modelling, the authors demonstrated that BIS is relatively unaffected by speed-accuracy trade-offs (SATs) and is appropriate for between-subjects designs and tasks with varying levels of mean accuracy and decision thresholds thus making it suitable for individual differences face processing research. Face processing tasks often instruct participants to respond as accurately and as fast as possible. This instruction is not neutral as it requires participants to decide whether to prioritise speed or accuracy since one may affect the other. Even when instructions are silent on this matter, or the time allowed for responses is not constrained, participants must still decide how to balance speed and accuracy. Controlling for differential speed-accuracy trade-offs is essential in a lifespan study of DP. Draheim and colleagues (2019) point out in their review of RT and individual differences that multiple studies have shown speed accuracy trade-offs differing across ability level (participants with *lower ability* are more likely to *sacrifice speed for accuracy* vs those with higher ability) and age (older adults tend to proceed more slowly and carefully than young adults regardless of instructions).

As Liesefeld and Janczyk (2019) point out, the relative weightings applied to speed or accuracy may not only differ *between* participants, but also *within* participants e.g., an individual might prioritise speed on easier trials and accuracy on harder trials. BIS addresses these challenges by equally weighting accuracy and RT and can be thought of as accuracy adjusted for RT, thereby *controlling for differential speed-accuracy trade-offs*. BIS is calculated by subtracting a participant's standardised RT score on correct trials from their standardised accuracy score [BIS = (Z accuracy – Z RT)], see Figure 4.4 below. So, for example, the BIS for a hypothetical DP participant whose accuracy z score is -1.2 (less accurate than average) and RT z score is 1 (slower than average) would be -2.2 (-1.2 *minus* 1), whereas BIS for a hypothetical control participant with an accuracy Z score of -1.2 (also below average) and a RT z score of -1.2 (faster than average) would be 0 (-1.2 *minus* -1.2).

4.1.4 Different approaches to defining inclusion and exclusion criteria

How researchers choose to define study inclusion and exclusion for the DP and control groups has important implications for our ability to fully understand DP. A key question is why so many individuals who report severe problems recognising familiar faces in everyday life do not appear to meet the classification for DP (for a discussion see Burns et al., 2022; DeGutis et al., 2023). For example, a recent large-scale study of 165 adults who reported severe everyday face recognition problems (Bate et al., 2019) showed that 61.8% of the 165 suspected DPs did not meet the commonly used diagnostic threshold of at least 2 SDs below control means on a minimum of two of the three commonly used diagnostic tests. These were the Cambridge Face Memory Test (CFMT; Duchaine & Nakayama, 2006), the Cambridge Face Perception Test (CFPT; Duchaine et al., 2007) and a Famous Faces Test (FFT). The CFMT is the gold standard test for detecting DP, but even on this test 107/165 (65%) participants did not meet strict cut off criteria (- 2 SD). Burns et al (2022) also report that a similar proportion of DPs (56%) did not score more than 2 SD below controls means on the CFMT.

The studies by Bate and colleagues (2019) and Burns et al., (2022) are relatively unusual in assessing the performance of all the self-reported DPs. More commonly in the literature only those participants who meet strict objective criteria for DP are included in studies (for a review see DeGutis et al., 2023). Unnecessary exclusion of a high proportion of potential participants who report this rare condition is problematic, both for practical and, arguably, ethical reasons (Burns et al., 2022).

A final consideration is exclusion criteria based on face test performance. It is common practice to exclude control participants whose performance falls more than 1.7 or 2 *SDs* below control mean. However, although this approach avoids the unwelcome possibility of including potentially undiagnosed DPs in the control group, it also could lead to an overestimation of the FR abilities of the remaining control group by inflating the control group mean. This is problematic because excluding individuals who produce low test scores but report no day-to-day face recognition problems - even when prompted through a comprehensive and ecologically valid PI-20 questionnaire (Shah et al., 2015) or similar instrument could, in turn, create two artificially-distinct groups. After some initial uncertainty, the PI-20 has been shown to be a reliable and valid means of classifying DP (Burns et al., 2022; Gray et al., 2017; Tsantani et al., 2021).

Subjective face recognition difficulty is widely considered a pre-requisite for classification as a DP. Arguably therefore, it logically follows that an *absence* of subjective difficulty – provided subjective experience has been interrogated using a questionnaire - rules out the possibility of undiagnosed DP. Some low scoring individuals may not experience face recognition difficulties in day-to-day life because they have developed effective compensatory techniques or because their poor test performance simply represents natural variability in face processing or other cognitive processes such as attention - or both. In reality, there are likely to be overlapping face recognition abilities between participant groups who do, and do not, report face recognition difficulties but individuals falling within this overlapping range (exact scores will vary test by test) are currently rarely researched due to prevailing selection and classification methods.

Increasingly therefore, several research groups have begun to question whether the current approaches to classifying DP are appropriate and suggest that broader inclusion criteria might help inform our understanding of DP (Berger et al., 2022; Burns et al., 2022; Dalrymple & Palermo, 2016; DeGutis et al., 2023; Mishra et al., 2021; Stumps et al., 2020). Unless future research includes both the full range of individuals who report FR problems (potential DPs) *and* the full range of neurotypical individuals who score poorly on face processing tasks but do NOT report FR problems on the PI-20 or similar questionnaire, we are unlikely to be able to satisfactorily characterise DP.

4.1.5 The present study

In this study we therefore adopt a new approach that allows a more comprehensive understanding of the nature of the underlying deficits in participants who report severe and noticeable face recognition difficulties in everyday life (for simplicity we call these DPs). Specifically, we compare the classification of a group of 24 self-reported DPs on traditional accuracy measures alongside a new classification method using the Balanced Integration Score (BIS). BIS accounts for speed accuracy trade-offs possibly masking face and object processing impairments in some DPs. The aim of this study was to identify whether accuracy or BIS yields an objective measure that better accounts for self-reported face recognition ability.

Definitions

We use the term face processing generically to refer to the overall process involved in recognising a human face. More specifically, here *face recognition* and *face memory* are used to refer to the ability to say whether a newly learned target face has been seen before. Both the Old New faces task and the CFMT test face memory. We also use (object/bicycle) memory to mean the ability to say whether a specific bicycle has previously been seen. Face *memory* tasks do not require participants to be able to name the facial exemplar or to recall any biographical details about an individual. By contrast, when we refer to face *identification*, we mean the ability to identify a face, either by name or other biographical detail (e.g., Actor who plays Mr. Bean). The Famous Faces task tests face *identification*. It also tests face *familiarity* by which we mean the ability to correctly judge (yes /no) whether the face of a personally-known celebrity looks familiar.

4.2 Research transparency and openness

The data presented here were used to classify participants in a wider study which investigates face perception in developmental prosopagnosia and findings from the perceptual task battery will be reported separately in a future publication. Although exploratory in nature, we decided to pre-register the study after data collection had commenced, but prior to analysis, to avoid any suspicion of fishing for results or hypothesising after results were known. The preregistered hypotheses, analysis plan, study data and R analysis scripts are available via The Open Science Framework <https://osf.io/qne8d/>. Analysis of one of the initial screening tasks, the Matrix Reasoning Item Bank (Chierchia et al., 2019) showed that the self-reported DP group adopted a

different speed-accuracy strategy to controls, preferring to proceed more slowly and carefully on this visual processing task that did not involve faces (Lowes et al., 2023 under review). To account for these observed speed-accuracy trade-offs, we therefore report here one additional measure that was not preregistered - the balanced integration score or BIS (Liesefeld & Janczyk, 2019, 2022) and describe this further in section 4.3.4. We also conducted additional unregistered logistic regression analysis and regression analyses to formally assess whether BIS or our preregistered variables best classified potential DPs.

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

Legal copyright restrictions prevent public archiving of the Cambridge Face Memory Tests, the Cambridge Bicycle Memory Test, Old New Faces and the Famous Faces Test which can be obtained from the copyright holders in the cited references (details at <https://osf.io/f496b/>)

4.3 Methods

4.3.1 Participants

Participants were 24 individuals aged 8 – 71 years (7 men; 17 women) who self-reported severe everyday face recognition difficulties (DPs) and 110 age matched controls aged 6 – 74 years (50 men; 60 women). All participants were living in the UK, Ireland and USA and reported normal, or corrected-to-normal, vision as well as a sufficiently good level of English to understand the instructions and participant information. Participants needed access to a laptop or computer with stable broadband.

Preregistered exclusion criteria were as follows. Individuals with any neurodevelopmental or neurological condition; learning difficulty (other than mild dyslexia) or psychiatric illness; a history of major/moderate brain injury at any time or a mild head injury or concussion during the preceding 12 months or acquired prosopagnosia (i.e., any face recognition difficulty that developed suddenly).

We enquired whether participants had previously participated in other face recognition or face training studies or had attempted any of the test battery online. Control participants who had done so were excluded (number excluded $n = 0$). DPs who had

previously completed a face recognition test, but not face training, were included provided that at least three months had passed ($n = 7$, AF006, AF013, AF017, AF018, AF019 and CF008). Four of these reported prior participation in other studies but did not know which ones, two had attempted an (unspecified) online test and one child had been tested by a neuropsychologist using tasks not contained in our battery.

Participant data from any single test where responses were indicative of repeated random key press, technology failure, a clear lack of understanding of the task or failure to follow instructions were excluded. Following preliminary analysis, data from two child control participants were removed because their performance on two or more tasks suggested suboptimal effort or failure to follow instructions. Two adult controls were also excluded because of inconsistencies between their PI-20 scores which were borderline indicating possible difficulty, and their follow up interview which revealed no real evidence of lifelong subjective difficulty (e.g., PI-20 scores were influenced by a recent one-off failure to recognise a neighbour who was wearing a face mask). Because these two cases met neither the DP criteria (clear subjective lifelong impairment) *nor* the control criteria (no subjective impairment as measured by the PI-20), we excluded them.

Participants were recruited through media coverage, social media, personal networks, and prosopagnosia support groups. Data were collected online and participants offered a £10 gift voucher to recompense them for their time. Approval for the study was granted by the University of Stirling General University Ethics Panel (project code 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.

An overview of individual DP participant accuracy performance on the tests that follow is provided in Table 4.1.

Table 4.1 Individual case scores of self-reported DP participants

ID	Age	Gender	Self-report	Old New Faces		Cambridge Face Memory Tests		Famous Faces Test		CMT difference	
				<i>Effect size</i>	Estimated % of controls falling below case's score	<i>Effect size</i>	Estimated % of controls falling below case's score	<i>Effect size</i>	Estimated % of controls falling below case's score	<i>Effect size</i>	Probability that the standardised difference for a member of the control population would be greater than that of the case and in the same direction
			<i>Z</i>	<i>Z CC</i>		<i>Z CC</i>		<i>Z CC</i>		<i>Z-DCC</i>	
CF007	8	F	-5.15	0.30	61.0%	0.90	80.0%	-1.74	5.69%	2.10	2.80%
CF008	10	M	-7.18	0.12	54.5%	-0.08	23.1%	-3.33	0.19%	1.19	12.9%
CF059	16	F	-5.59	NA	NA	-2.34	1.66%	-0.57	29.0%	-0.43	35.4%
CF005	17	M	-0.66	-1.35	10.1%	-3.45	0.15%	-1.98	3.29%	-2.57	0.97%
AF016	23	F	-7.01	-1.35	10.1%	-0.83	21.3%	-2.63	0.87%	-0.33	37.2%
AF002	25	F	-7.28	-0.93	18.2%	-1.54	7.45%	0.42	65.6%	-1.14	13.7%
AF017	29	F	-6.60	-1.74	5.22%	-1.54	7.45%	-0.03	48.7%	-2.01	3.01%
AF004	31	M	-4.54	0.61	72.2%	-0.93	18.8%	-1.87	4.08%	-1.74	4.98%
AF007	31	F	-4.95	-0.57	29.4%	-0.53	30.6%	0.34	62.8%	-1.07	15.0%
AF009	42	M	-5.17	0.88	79.8%	-1.22	12.4%	-0.63	27.4%	-1.76	4.75%
AF013	44	F	-4.73	0.88	79.8%	0.92	80.9%	0.92	81.0%	0.19	42.5%
AF075	45	F	-3.64	0.29	60.8%	-0.80	22.2%	-1.41	9.18%	-0.63	26.8%
AF018	49	F	-5.72	0.88	79.8%	-0.91	19.3%	-4.31	0.02%	-0.82	21.4%
AF021	50	F	-3.09	0.88	79.8%	-1.55	7.39%	-2.04	2.96%	-2.23	1.93%
AF001	51	F	-4.62	-0.91	19.3%	-1.45	8.61%	-0.24	40.7%	-1.38	9.32%
AF010	53	F	-5.17	-9.83	0.00%	-3.04	0.40%	-2.63	0.88%	-2.48	1.37%
AF006	54	F	-6.16	-2.70	0.85%	-2.08	2.85%	-1.59	6.80%	-1.36	10.1%
AF003	55	F	-6.05	0.88	79.8%	-1.55	7.39%	-1.52	7.60%	-2.14	2.31%
AF060	61	F	-4.65	-0.91	19.5%	-0.42	34.3%	-1.91	4.02%	-0.85	20.6%
AF022	64	F	-4.65	-0.18	43.3%	0.04	51.7%	0.72	75.4%	-0.49	31.7%
AF098	67	M	-2.76	-3.13	0.39%	-1.74	5.56%	-1.45	8.88%	-1.16	13.6%
AF019	68	M	-7.75	-3.87	0.09%	-1.64	6.52%	-5.22	0.01%	-1.64	6.22%
AF099	70	M	-4.38	-3.13	0.39%	0.33	62.2%	NA	NA	-0.10	46.1%
AF008	71	F	-4.78	-0.18	43.3%	-0.52	31.1%	-1.37	10.1%	-0.91	19.0%

ID	Age	Gender	Self-report	Old New Faces RT		Cambridge Face Memory Tests RT		Cambridge Bicycle Memory Test RT	
				<i>Effect size</i>	Estimated % of controls slower than case's mean RT	<i>Effect size</i>	Estimated % of controls slower than case's mean RT	<i>Effect size</i>	Estimated % of controls slower than case's mean RT
			Z	Z CC		Z CC		Z CC	
CF007	8	F	-5.15	0.66	47.6%	-0.82	77.9%	1.97	3.90%
CF008	10	M	-7.18	-0.26	59.7%	-0.30	61.3%	0.98	17.6%
CF059	16	F	-5.59	NA	NA	2.08	2.8%	0.22	41.5%
CF005	17	M	-0.66	0.92	19.0%	0.39	35.5%	-0.82	78.4%
AF016	23	F	-7.01	0.76	23.2%	0.69	25.4%	-0.69	74.7%
AF002	25	F	-7.28	3.40	0.20%	1.16	13.5%	0.48	32.3%
AF017	29	F	-6.60	1.43	9.00%	-1.36	90.1%	-1.14	86.1%
AF004	31	M	-4.54	1.29	11.1%	1.27	11.5%	0.73	24.1%
AF007	31	F	-4.95	2.33	1.70%	4.14	0.04%	1.87	4.20%
AF009	42	M	-5.17	0.85	21.0%	4.34	0.02%	1.63	6.30%
AF013	44	F	-4.73	1.68	5.90%	0.13	45.1%	0.18	43.1%
AF075	45	F	-3.64	3.35	0.20%	3.03	0.40%	2.35	1.60%
AF018	49	F	-5.72	2.05	3.10%	1.97	3.50%	1.73	5.30%
AF021	50	F	-3.09	2.26	2.00%	0.48	32.2%	-0.11	54.1%
AF001	51	F	-4.62	1.54	7.50%	-0.88	80.0%	-1.54	92.6%
AF010	53	F	-5.17	2.28	2.00%	0.43	34.0%	-1.01	83.2%
AF006	54	F	-6.16	-0.27	60.3%	-0.92	81.0%	-0.59	71.6%
AF003	55	F	-6.05	1.48	8.30%	0.70	25.2%	0.83	21.4%
AF060	61	F	-4.65	2.02	3.40%	-0.35	63.1%	0.07	47.3%
AF022	64	F	-4.65	0.11	46.0%	0.55	29.9%	1.58	7.10%
AF098	67	M	-2.76	0.70	25.3%	-0.72	75.2%	0.05	48.1%
AF019	68	M	-7.75	3.03	0.50%	0.08	47.0%	0.82	21.8%
AF099	70	M	-4.38	4.39	0.03%	0.10	46.1%	-0.02	50.8%
AF008	71	F	-4.78	5.54	0.01%	0.72	24.8%	3.04	0.40%

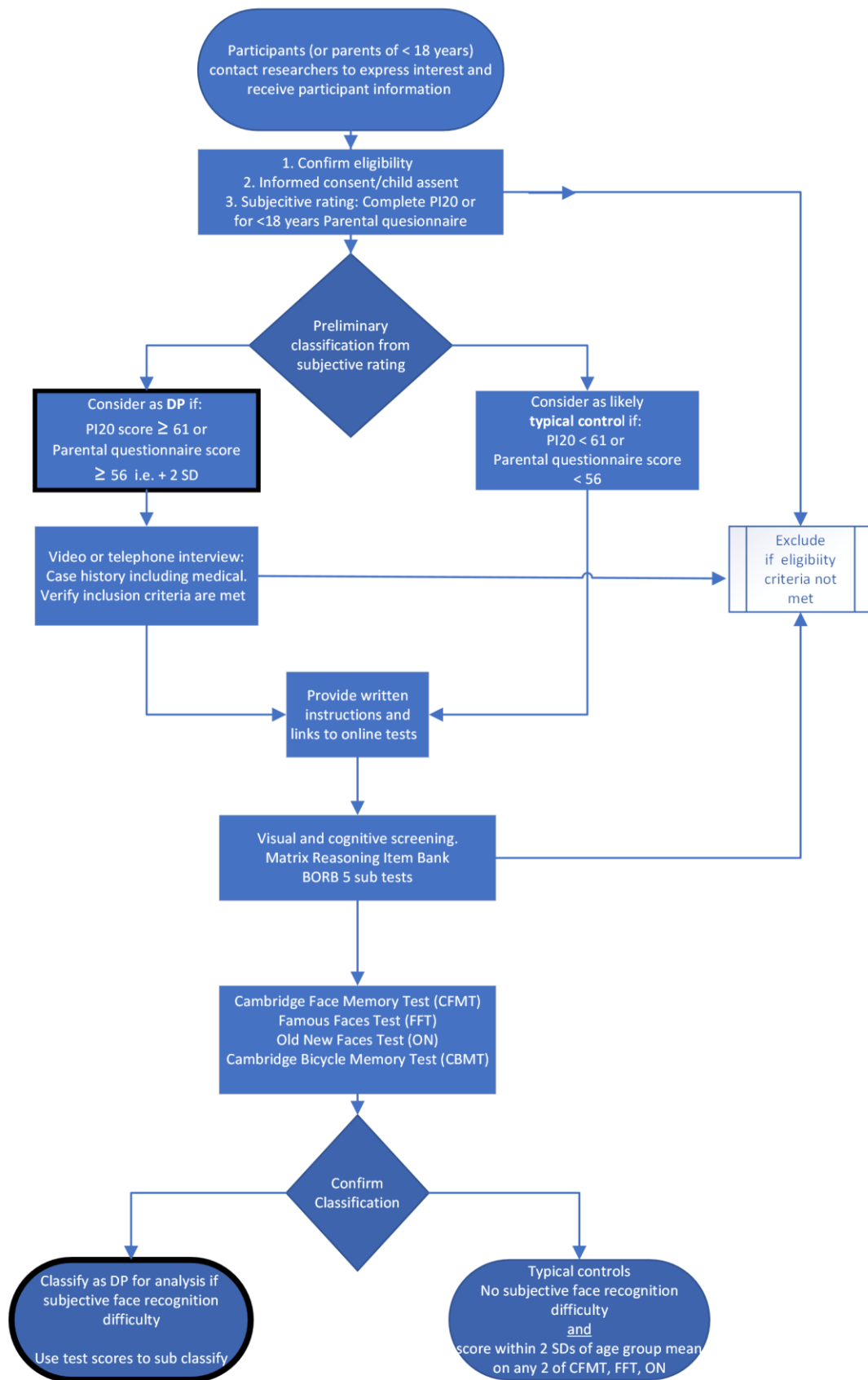
Note. Self report = standardised score on PI-20 or parental questionnaire. CMT difference = (Cambridge Face Memory accuracy z score – Cambridge Bicycle Memory accuracy z score). Mean RT is response time on correct trials. Effect sizes in bold indicate performance significantly worse than controls calculated using the SingleBayes_ES.exe computer programme for Bayesian tests of deficit (Z-CC) and the DiffBayes_ES.exe programme for Bayesian standardised difference test (Z-DCC) from Crawford et al., 2010 and Crawford & Garthwaite, 2007. Alpha = 0.05.

4.3.2 Materials

4.3.2.1 Background visual and cognitive screening.

To rule out deficit in low-level vision or more general cognition as explanations for impaired face processing test performance, we screened vision and non-verbal fluid reasoning. No participant scored significantly below chance on two or more of the following sub tests of the Birmingham Object Recognition Battery (BORB, Riddoch & Humphreys, 1993): Length of line, object decision, size of circles, line orientation, and position of gap. Following initial inspection, one child participant was excluded because their BORB RT data indicated suboptimal effort. Except for one participant (CF059, see below), all remaining DP and control participants had accuracy scores within 2 *SD* of the gender-matched age group mean on the Matrix Reasoning Item Bank (MaRs-IB) (Chierchia et al., 2019). One DP (CF059 female, 16 years) performed significantly below average on this task, indicating a wider cognitive difficulty. A small number of controls ($n = 20$) did not complete the MaRs-IB but their performance on other tasks was within typical norms so they were retained (n.b. reasoning serves as a check to ensure poor test performance in DP cannot be accounted for by impaired cognitive function or difficulty following test instructions). An overview of the recruitment and classification process is shown in Figure 4.1 below.

Figure 4.1 Overview of recruitment and classification process



4.3.2.2 Subjective report

To be initially classified as a potential DP, participants had to have subjective report of lifelong difficulty with familiar face recognition. For adults this was operationalised as a score of ≥ 61 on the PI-20 (Shah et al., 2015) as recommended by Tsantani et al., (2021). This cut-off represented roughly 2.5 standard deviations above the age group control means (and 3.6 *SD* above the mean in the 14 - 35 years age group). Inclusion and exclusion criteria were confirmed by follow up semi-structured interview which also captured details of daily face recognition experiences. For children and young people aged 6 - 17 years we required parental report of face recognition difficulties, classified as a score > 1.7 *SD* above the control group mean on a questionnaire comprising 32 items: 17 items from the PI-20 (Shah et al., 2015), rephrased where necessary to make it suitable for reporting a third party's ability, and 15 additional items identified as hallmarks of prosopagnosia for non-experts (Murray et al., 2018). Similar to the PI-20, a higher score indicates more difficulties with face recognition. Like adults, inclusion and exclusion criteria and case histories were confirmed through screening interviews with a parent. For ease of comparison with other tasks, negative self-report *z* scores indicate performance worse than the control average.

4.3.2.3 Face tests

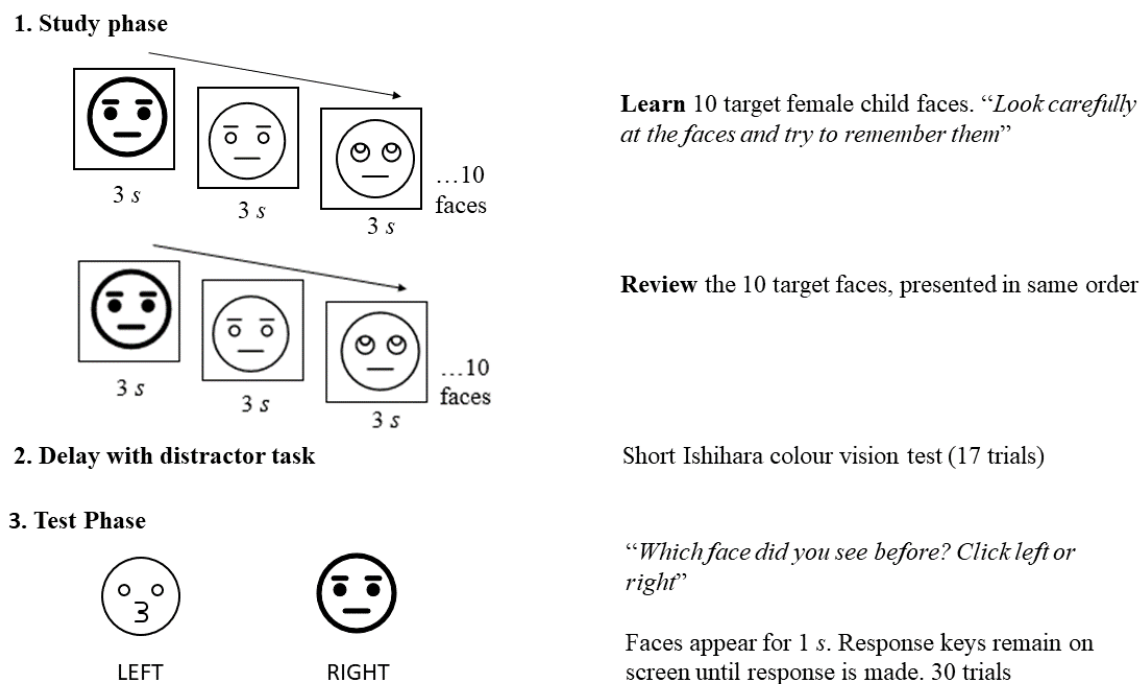
4.3.2.3.1 Old New Faces

The Old New Face task (see Figure 2) is a test of delayed face identity recognition matching. Full methods are described in the original paper (Dalrymple et al., 2014), both children and adults completed the version with child faces. Since *z* scores were computed using the means and *SD* of the appropriate control age group, any own age bias is therefore controlled for because each participant is compared to the typical score for their relevant age group (Figure 4.4). Briefly, participants memorise ten target child faces presented one at a time for 3 s each. Each target was then immediately shown again in the same order for 3 s. Both adults and children Participants were instructed to try to memorise the faces. At test, the task is to indicate which of two faces presented is the previously seen face. One target and a similar-looking distractor (age, expression, orientation) were presented simultaneously for 1 s. Response options appeared under each face ("LEFT" or "RIGHT") and remained

on screen until a response was made by mouse click. Targets appeared in randomised order three times each for a total of 30 trials alongside 30 unique distractors that were never repeated. Stimuli were presented in grayscale with hair, ears, and any obvious moles or blemishes removed.

DVs were accuracy and BIS (which uses mean RT on correct trials only). We administered an amended version of the test by introducing a distractor task between study and test. This took the form of a 17-item Ishihara colour vision test, lasting approximately two minutes. Data from one child control participant was not included in analysis of this test since their mean RT was 2.94 *SD* above the age group mean. Their other test performance was within typical norms, so they were retained as a participant.

Figure 4.2 Old New Faces test design



4.3.2.3.2 Cambridge Face Memory Tests (CFMT)

Three versions of the CFMT, suitable for different ages, were administered. The CFMT is considered the gold standard test for detection of prosopagnosia. It tests viewpoint-dependent and viewpoint-independent recognition memory for newly-learned faces. Impaired performance on this test is widely used to classify DP

since difficulty learning and individuating faces is the core behavioural manifestation of DP. There is a matched object recognition test and comparison of participants' standardised scores on both tests can indicate whether individuation deficits are face selective or also extend to object recognition - at least for the object class being tested. The DVs for all versions of the CFMT were accuracy and BIS. BIS is calculated using mean RT on correct trials only.

4.3.2.3.2.1 Cambridge Face Memory Test (adults and adolescents aged ≥ 14 years).

A full description of the methods can be found in the original paper (Duchaine & Nakayama, 2006). Briefly, participants study six target faces each from three different viewpoints. Cropped and greyscale faces are presented for 3 s. The test stage uses a three alternate forced choice (3 AFC) format, comprising one target and two distractor faces. In the first introduction test phase each target face is tested with three identical viewpoints. In the second test phase, after review, each target face is shown in a novel viewpoint from that learned at study and, finally, the noise test section introduces novel views of the target faces with added gaussian noise. There are 72 trials in total and chance = 33%.

4.3.2.3.2.2 Cambridge Face Memory Test - Kids (children 10 - 13 years).

The child version of the test (Dalrymple et al., 2014) follows a similar procedure to the CFMT but uses six child faces. There are 72 trials in total and chance = 33%.

4.3.2.3.2.3 Cambridge Face Memory Test - Young Kids (children 6 - 9 years).

Due to floor effects being observed in the CFMT-Kids in children < 10 years (Dalrymple & Duchaine, 2016; Dalrymple & Palermo, 2016), a shorter version of the test was designed (Dalrymple & Palermo, 2016). Here, participants study 12 unfamiliar target faces, one at a time, from three different viewpoints. Each identity is tested three times, once from each viewpoint, using a 3AFC paradigm. Like the introduction test phase in the CFMT and CFMT-Kids, testing occurs immediately after the study phase for each identity thus minimising memory demands. There are 36 trials in total and chance = 33%.

4.3.2.3.3. Cambridge Bicycle Memory Task (CBMT)

The CBMT (Dalrymple et al., 2014) is matched in format to the CFMT (adult and kid versions) but uses images of bicycles rather than faces to allow a wider object agnosia to be identified. The CBMT has been used with adults as well as children as young as seven years old and does not appear to have floor effects in this age group (Bate et al., 2020), is well matched in controls to the CFMT (Biotti & Cook, 2016), and has been argued to have better diagnostic properties than the car memory test (Barton et al., 2019). So, following Bate et al. 2020, the six-target version was used with all age groups. There are 72 trials in total and chance = 33%. As for the CFMT, the DVs were accuracy and BIS. The primary outcome measure of interest for classification purposes was the standardised CMT difference score which was computed as standardised face score minus standardised bicycle score.

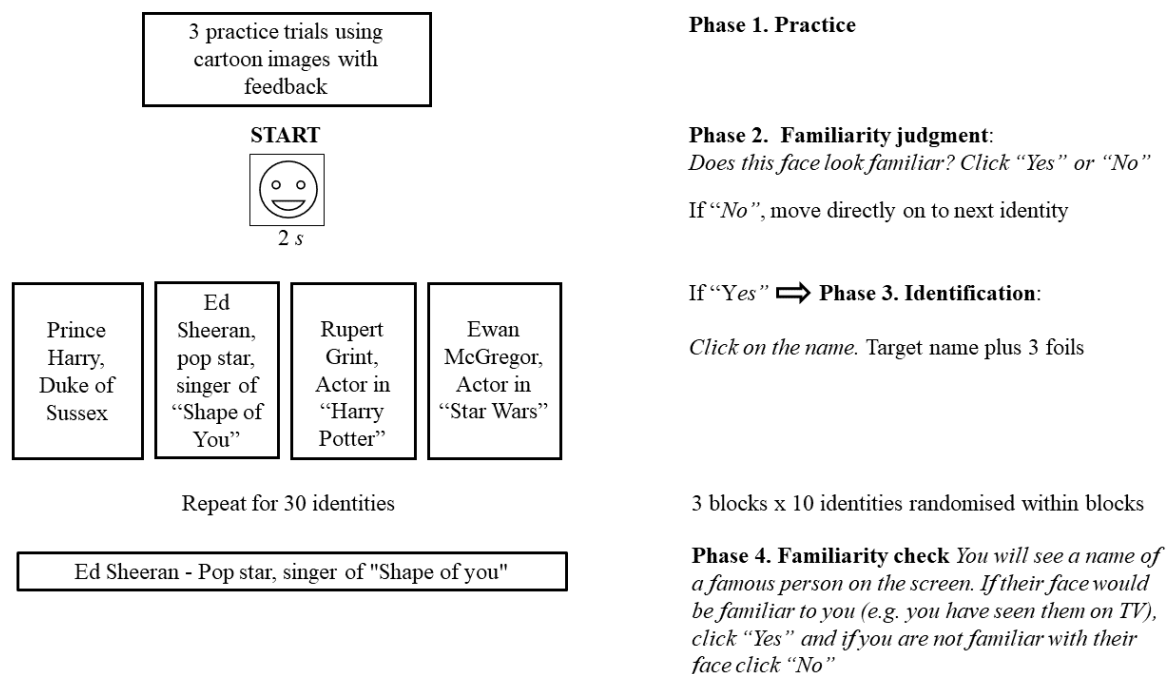
4.3.2.3.4 Famous Faces Test

Difficulty identifying familiar faces is the core deficit in DP. In contrast to the CFMT which tests memory for newly learned, unfamiliar faces, the FFT measures long-term familiar face recognition memory. We were not aware of any recent published famous face tests suitable for children and young people as well as adults, so a new test was devised and administered to all age groups. First, we used social media to informally poll parents of children aged 6-16 years. From the longlist of suggested famous identities, we selected 30 identities from multiple sectors covering sport, entertainment, music, politics, and royalty and gathered facial images from the internet. Pilot testing with eight typically developing participants in the UK aged 6-16 years showed that the median number of facial identities that participants reported knowing was 28.5/30 indicating that the chosen identities were likely to be familiar to children and young people across our target age range.

Participants saw 30 famous faces one at a time (Figure 4.3). Stimuli were presented in full colour on a black background with hair cropped but hairline and external contours retained. Identifying blemishes were removed and any jewellery blurred. Each face was presented for 2 s and participants indicated if the face was familiar by clicking “yes” or “no”. If “no”, they moved immediately onto the next trial and a new identity was shown. If “yes”, participants then had to click on the correct name/identity (e.g., Boris Johnson, UK Prime Minister- at the time of testing) from a choice of four (one target, three foils) before moving on to the next trial. Foils

were the descriptions of other famous identities matched for gender and approximate age and, as far as practical, profession. To discourage guessing in the initial familiarity judgement, instructions stressed that it didn't matter how many identities looked familiar. The test began with three practice trials using cartoon images to familiarise participants with the task and feedback was given. No feedback was provided during the test phase. Participants could take a break after every 10 trials. Stimuli and trial files are available from the authors on request. We do not have permission to share them publicly because images were sourced from the internet.

Figure 4.3 Famous Faces Test design



4.3.3 Procedure

All tests were administered using the online platform Testable (www.testable.org) with Google Chrome as the recommended browser, except the matrix reasoning screening task which was administered using Gorilla (www.gorilla.sc). Participants (or parents in the case of under 18s) were emailed a document containing written task instructions and links to each test and instructed to complete the tests in the prescribed order. Full onscreen instructions were also provided. Parents supervised children and could help with explaining tasks, but instructions stressed they must not help children with their responses. The tests

reported here were completed in the following order CBMT, FFT, CFMT and then Old New Faces. Testing took place over a minimum of three self-paced sessions. Each session included only one face recognition test alongside some perceptual tasks with no memory demands. Instructions recommended a minimum break between sessions of at least 12 hours for participants under 18 years and one hour for adults. In addition, participants were informed that they could stop at the end of any test to take additional breaks and parents were advised that children should take a break if they began to appear distracted or tired.

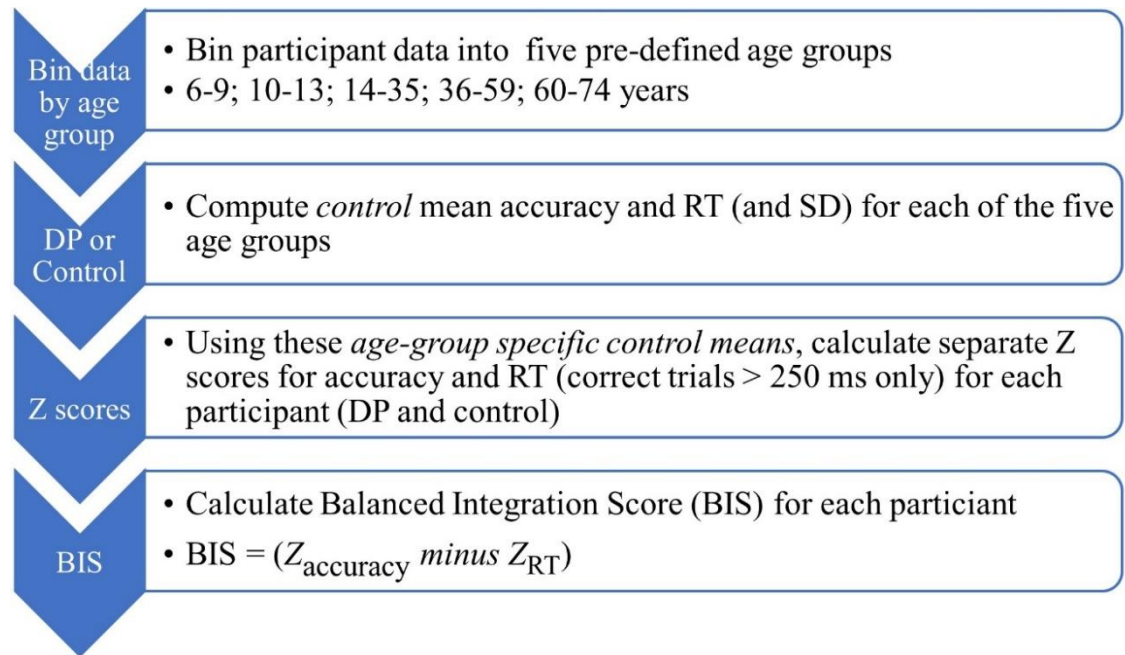
4.3.4 Analysis plan

First, we inspected the descriptive statistics and distributions for each test and compared these to published norms for controls where available. Next, test reliability was calculated using Cronbach's alpha for the sample as a whole and for the control and DP groups separately. We originally planned to include gender as a covariate, so we then checked for gender differences in all tests. Because there were no gender differences in all tests (all $ps > .05$) and there were no significant differences in gender distribution between the DP and control groups ($X^2 = 2.14$, $p = .144$, and all $ps > .144$ in individual age groups) data were collapsed across genders for analysis.

Finally, to allow comparison across tasks, we standardised all measures by calculating z scores (Figure 4.4). We centred z scores on control age group means because accuracy and RT varied on some tasks as a function of age. Age groups were predefined as 6-9, 10-13, 14-35, 36-59 and 60-74 years. The choice of age group broadly followed previous literature (e.g. Bowles et al., 2009) and was also partially driven by practical consideration as we aimed to have 20 controls in each age group. The child age groups were chosen primarily to be suitable for the CFMT-Kids (10-13 years) and CFMT-Young Kids (7-9 years). When deciding which tests to administer to adolescents 14-17 years, we followed Bate et al., 2015 and Bennetts et al. (2017) who administered the CFMT adult version to a prosopagnosic and a super recogniser (respectively) and typical controls aged 14 & 15 years. We inspected mean control performance for 14-17 year olds and 18-35 years old and found no difference with mean scores of 58.8/72 and 58.0/72 respectively. This age group standardisation allowed us to classify DP participants' performance across the full age range since here the z scores quantify participants' performance relative to their own age group on any given task. We checked within age groups for any significant age-accuracy

correlations and found none except among 6-9 year olds for Old New Faces accuracy ($r(10) = .664, p = .018$) and 10-13 year olds on the FFT ($r(20) = .44, p = .042$). We also report unstandardised mean accuracy, z score and BIS by age group for each test.

Figure 4.4 Overview of the method used to calculate z scores and BIS



The primary preregistered DV for the CFTM, CBMT and Old New Faces was accuracy (correct trials/total trials). This was because at the time of pre-registration we were not aware of BIS. As discussed in section 2 above, to account for the different speed accuracy trade off strategies observed between controls and DPs on the matrix reasoning screening task (Lowe et al., under review), we additionally calculated BIS ($Z_{\text{accuracy}} - Z_{\text{RT}}$) which can be thought of as accuracy adjusted for RT thereby controlling for differential speed-accuracy trade-offs.

For the FFT, three DVs were preregistered. In addition to the primary DV of accuracy (the proportion of identities known to the participant that were correctly identified), we also calculated the raw number of identities known to each participant (i.e., a familiarity check), and familiarity (the proportion of identities known to each participant that were reported as familiar). We did not compute RTs, and therefore

BIS, for the FFT since we considered reading speed to be an important potential confound.

The fifth and final preregistered measure of face memory was a global memory score computed for each participant from the mean of their standardised scores on the CFMT, Old New Faces, FFT and standardised CMT difference using pairwise deletion. We report two global memory scores, the first calculated using accuracy and the second using BIS.

4.3.5 Statistical analysis.

To calculate group differences, we conducted Bayesian independent samples *t* tests on the standardised scores using the default Cauchy prior with a scale of 0.707. We pre-registered Bayesian analysis because it enables the strength of evidence for both the alternative and null hypotheses to be compared and removes the need to correct for multiple comparisons (Gelman et al., 2012; Kruschke, 2010). For completeness, we also report 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. Data were analysed in R (R Core Team, 2021) using R Studio 2021.09.1 and the Tidyverse (Wickham et al., 2019) and jmv, version 2.3.4 (Selker et al., 2022) packages. Cronbach's alpha was calculated in SPSS 28.0.0.0 (IBM).

At the individual participant level, single case analyses were conducted using the SingleBayes_ES.exe computer programme for Bayesian tests of deficit (BTD) and the DiffBayes_ES.exe programme for Bayesian standardised difference test (BSDT, Crawford et al., 2010; Crawford & Garthwaite, 2007). The alpha level was set at 0.05. These analyses were preregistered. Our sample size of DPs was determined following previous literature and our control sample size (target of 20 participants per age group) was informed by McIntosh & Rittmo's (2021) study. This recommends a minimum neuropsychological control sample of at least eight participants and that twice that number is more desirable, but that increasing control sample size above 16 does little to meaningfully increase power.

Binomial logistic regression modelling and regression analysis were used to assess which outcome measures, or combination of measures, best predicted self-reported group membership as quantified by PI-20 or parental report scores. These

were not a pre-registered analyses but were added in order to formally assess whether the original variables or BIS best predicted group membership.

4.4 Results

The results section is structured as follows. We first report test reliability then group results for each of the four tests of face memory followed by the computed global (average) memory score. To illustrate the effect of taking RT as well as accuracy into consideration, we compare the group results calculated first using accuracy and second using BIS. Finally, we present logistic regression data showing which models best predicted group membership.

4.4.1 Reliability.

Table 4.2 shows reliability (Cronbach's alpha) for the overall sample and separately for the DP and control groups. Reliabilities are rarely reported in such detail but are needed in order to calculate maximum possible correlations between tests and their suitability for individual difference studies (see Bobak et al, 2023). Reliability for most tests was excellent or good; the Old New Faces test was acceptable for controls and good for DPs.

Table 4.2 Test reliabilities

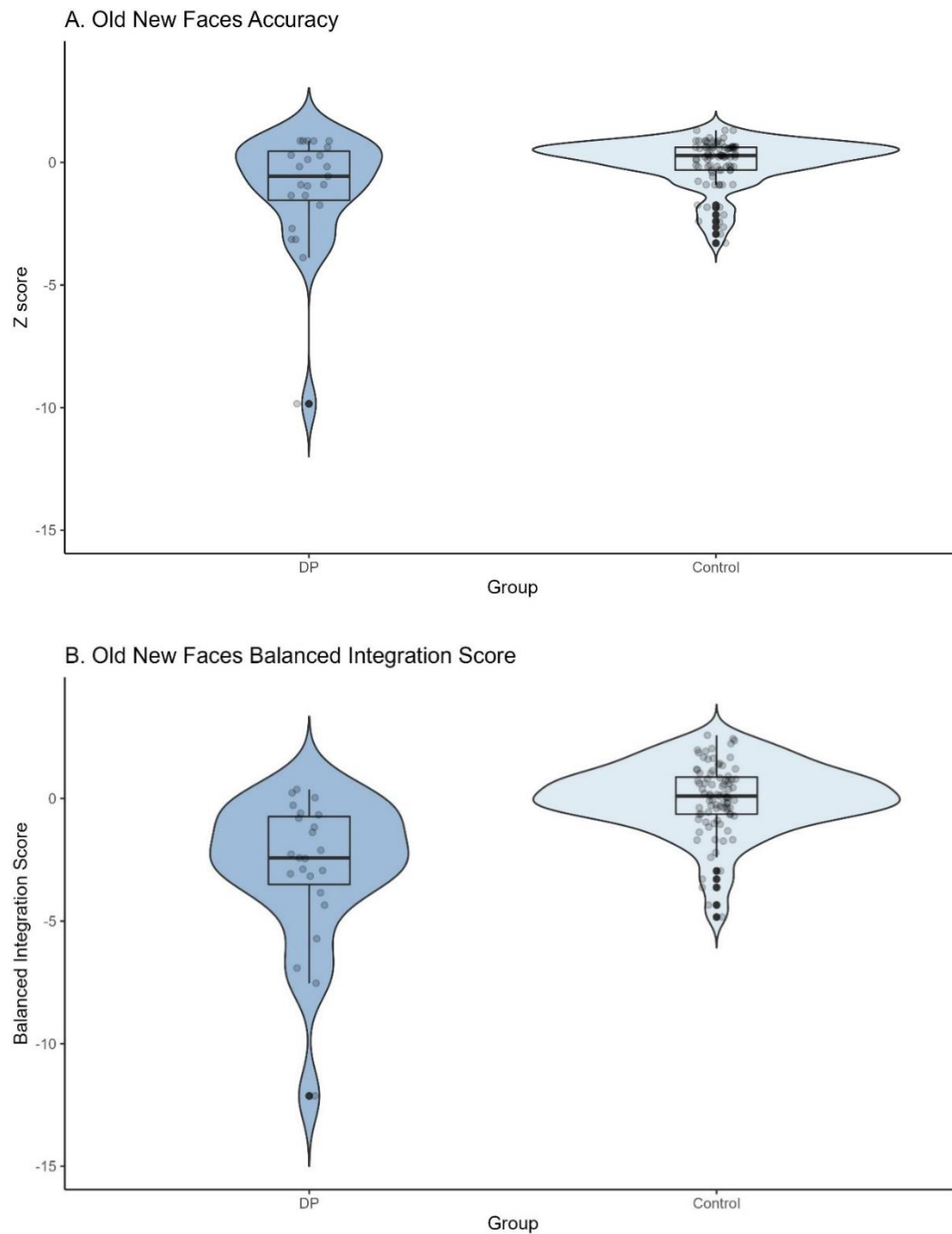
Test	Cronbach's alpha		
	Full sample	Controls	DPs
Old New Faces	.79	.77	.83
CFMT	.92	.92	.86
CFMT-Kids	.89	.90	-
CFMT-Young Kids	.92	.90	-
CBMT	.93	.93	.89
Famous Faces Test	.93	.93	.93

Note. CFMT = Cambridge Face Memory Test, CBMT = Cambridge Bicycle Memory Test. Only one DP completed the CFMT-Kids and CFMT-Young Kids meaning alpha could not be calculated separately for DPs on these tests.

4.4.2 Old New Faces

The potential DP group ($n = 23$, one self-reported DP participant did not complete this test) was, on average, less accurate than controls ($n = 87$, 23 participants did not complete this test) at judging whether a face had previously been seen or not. The results of Bayesian independent samples t test and Welch's t tests on standardised accuracy scores and BIS scores are reported in Table 4.8 and illustrated in Figure 4.5. There was strong evidence for a true difference between groups both for accuracy ($BF_{10} = 28.8$) and BIS ($BF_{10} = 5,090,000$). In other words, the alternative hypothesis (a true BIS difference exists between groups) is more than 5 million times as likely as the null hypothesis given the data. Descriptive statistics for each age group showing unstandardised accuracy scores and BIS are shown in Table 4.3.

Figure 4.5 Group difference (self-reported DPs vs controls) Old New Faces



Note: (A) shows accuracy and (B) BIS (Z_{accuracy} minus Z_{RT}). Each dot represents a single data point, the box shows the interquartile range (IQR) and the midline indicates the group median score. The end of each whisker line represents 1.5 x the IQR.

Table 4.3 Old New Faces accuracy and BIS by age group

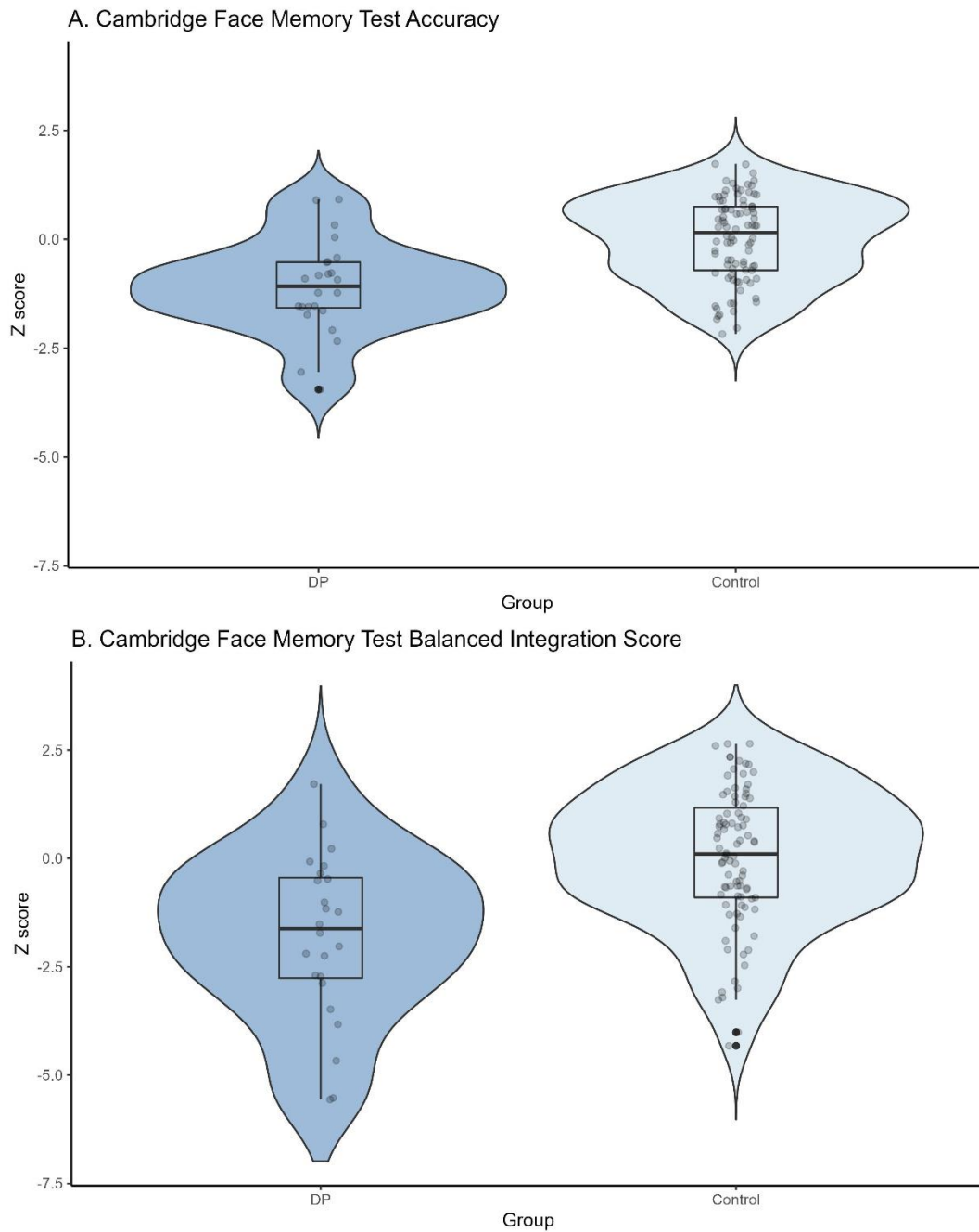
Proportion correct	DP				Control		
	<i>N</i>	<i>Z</i>	<i>M</i>	<i>SD</i>	<i>N</i>	<i>M</i>	<i>SD</i>
6 - 9 years	1	0.30	.90	-	12	.87	.09
10 - 13 years	1	0.11	.90	-	18	.88	.15
14 - 35 years	6	-0.89	.87	.07	21	.95	.09
36 - 59 years	9	-0.98	.90	.20	19	.95	.06
60 - 74 years	6	-1.91	.86	.08	17	.94	.05
BIS							
6 - 9 years	1	0.23		-	12	0.00	1.16
10 - 13 years	1	0.37		-	18	0.00	1.49
14 - 35 years	6	-2.58		1.23	21	0.00	1.57
36 - 59 years	9	-2.67		3.69	19	0.00	1.65
60 - 74 years	6	-4.54		2.73	17	0.00	1.37

Note. Chance = .5. There was only one potential DP in each of the two youngest age groups so *SD* could not be calculated separately for self-reported DPs in these age groups. Data from 23 controls is not available due to participant or technical error or participant drop out. Because BIS is a standardised measure, BIS scores for controls are 0 by definition.

4.4.3 Cambridge Face Memory Test (group and age group)

We analysed standardised accuracy scores to allow comparison across the different versions of the CFMT. Results are shown in Table 4.8 and illustrated in Figure 4.6. As expected, self-reported DPs were less accurate than controls and overall there was strong evidence of a true difference between groups for both accuracy ($BF_{10} = 3,290$) and BIS ($BF_{10} = 2,453$). Unstandardised accuracy scores and BIS are shown in Table 4.4.

Figure 4.6 CFMT group difference in performance



Note: Group difference (self-reported DPs vs controls) in performance using (A) accuracy (correct trials/total trials) and (B) BIS (Z_{accuracy} minus Z_{RT}). Each dot represents a single data point, the box shows the interquartile range (IQR) and the midline indicates the group median score. The end of each whisker line represents 1.5 x the IQR. Participants aged 6 - 9 years completed the CFMT Young Kids, ages 10 - 13 years completed the CFMT-Kids and participants age ≥ 14 years completed the CFMT.

Table 4.4 CFMT accuracy and BIS by age group

Proportion correct	DP				Control		
	<i>N</i>	<i>Z</i>	<i>M</i>	<i>SD</i>	<i>N</i>	<i>M</i>	<i>SD</i>
6 - 9 years	1	0.90	.97	-	15	.80	0.19
10 - 13 years	1	-0.47	.65	-	17	.76	0.14
14 - 35 years	7	-1.59	.59	0.14	21	.81	0.14
36 - 59 years	9	-1.27	.65	0.14	20	.81	0.13
60 - 74 years	6	-0.65	.62	0.13	17	.72	0.15
BIS							
6 - 9 years	1	1.71		-	15	0.00	1.74
10 - 13 years	1	-0.47		-	17	0.00	1.56
14 - 35 years	7	-2.79		1.82	21	0.00	1.47
36 - 59 years	9	-2.31		1.92	20	0.00	1.77
60 - 74 years	6	-0.72		0.74	17	0.00	1.63

Note. Chance accuracy = .33. There was only one self-reported DP in each of the two youngest age groups so *SD* could not be calculated separately for DPs in these age groups. 20 controls did not complete this test. Because BIS is a standardised measure, BIS scores for controls are 0 by definition.

4.4.4 Difference between face (CFMT) and object (CBMT) memory accuracy

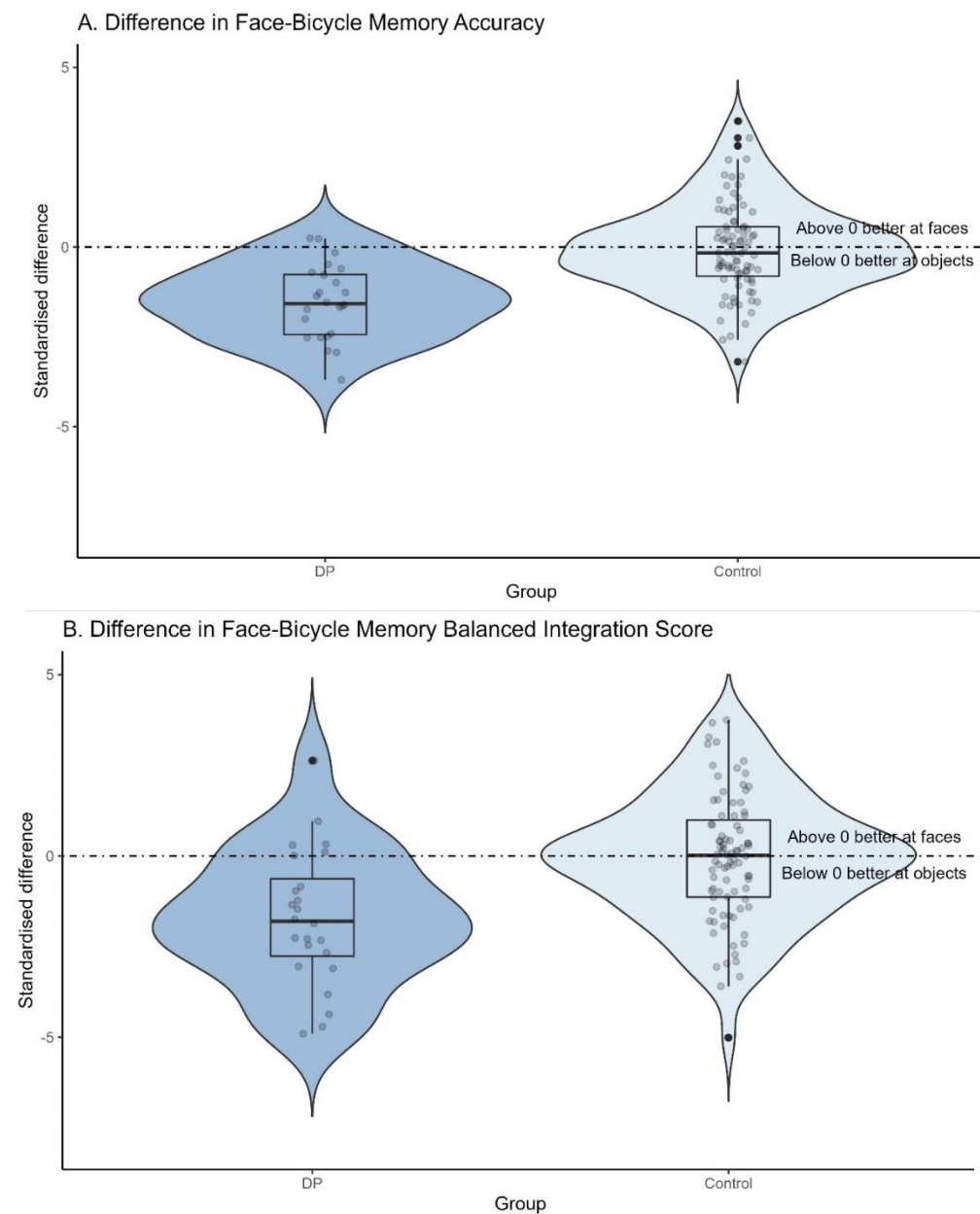
To investigate whether any memory difficulties were face-specific, we calculated a CMT difference score by subtracting each participant's CBMT *z* score from their CFMT *z* score (Figure 4.7). A difference score above zero indicates that a participant performed relatively better at faces than bicycles and a score below zero indicates that a participant performed relatively better at bicycles than faces. As a reminder, all *z* scores were centred on age group control means so this difference score is therefore a relative measure of face/bicycle memory accuracy compared to participants' own age control group. Unstandardised accuracy and BIS scores for the CBMT are shown in Table 4.5.

Table 4.5 Cambridge Bicycle Memory Test accuracy and BIS by age group

Proportion correct	DP				Control		
	<i>N</i>	<i>Z</i>	<i>M</i>	<i>SD</i>	<i>N</i>	<i>M</i>	<i>SD</i>
6 - 9 years	1	1.06	.72	-	15	.57	.15
10 - 13 years	1	0.50	.78	-	21	.69	.17
14 - 35 years	7	0.32	.85	.19	21	.81	.12
36 - 59 years	9	0.35	.86	.07	22	.81	.14
60 - 74 years	6	0.56	.86	.06	18	.78	.15
BIS							
6 - 9 years	1	-0.92		-	15	0.00	1.03
10 - 13 years	1	-0.48		-	21	0.00	1.12
14 - 35 years	7	0.28		1.93	21	0.00	1.57
36 - 59 years	9	-0.03		1.37	22	0.00	1.50
60 - 74 years	6	-0.36		1.03	18	0.00	1.23

Note. Chance accuracy = .33. There was only one DP in each of the two youngest age groups so *SD* could not be calculated separately for DPs in these age groups. Because BIS is a standardised measure, BIS scores for controls are 0 by definition.

Figure 4.7 CMT group difference



Note: Showing group difference (self-reported DPs vs controls) in participants' relative performance on the Cambridge Face Memory Test and Cambridge Bicycle Memory Test (Z_{CFMT} minus Z_{CBMT}) using (A) accuracy and (B) BIS ($Z_{accuracy}$ minus Z_{RT}). Each dot represents a single data point, the box shows the interquartile range (IQR) and the midline indicates the group median score. The end of each whisker line represents 1.5 x the IQR.

4.4.5 Famous Faces Test

Data from one potential DP were not analysed because they notified us of participant error during this task. Data from 23 controls is not available due to participant or technical error or participant drop out (i.e. some participants did not

complete all three sessions). No floor effects were observed in any age group, 2 *SD* below the control mean accuracy was always above chance (25%). As discussed above, the test design meant it was not appropriate to calculate BIS for the Famous Faces Test, instead we use only accuracy. As discussed below, FFT accuracy was a strong predictor of group and so remains a useful measure.

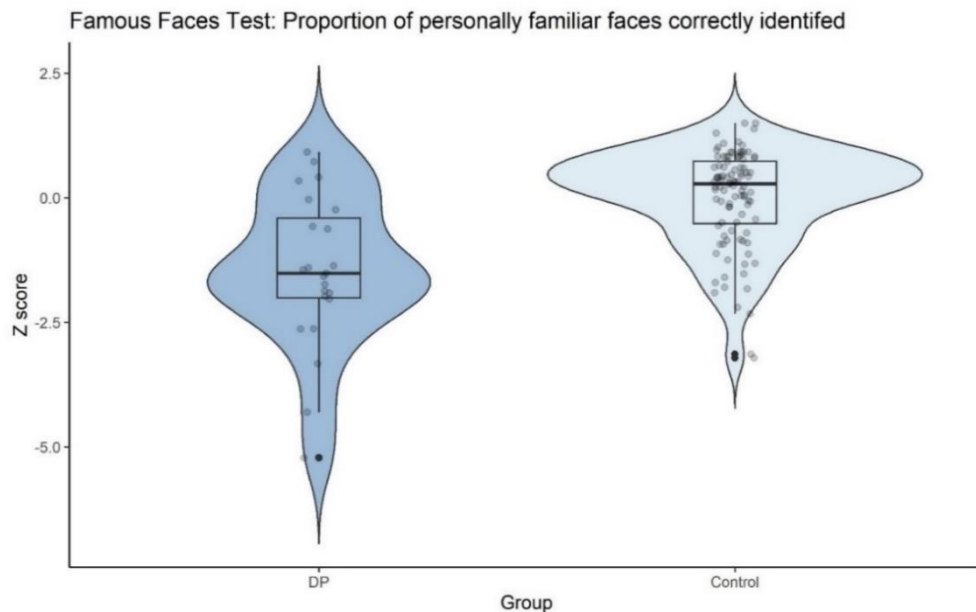
The raw number of famous faces known to participants was checked at the end of the test. At an age group level, the mean number of identities known ranged from 13.4 in young children (skewed by a higher proportion of non-UK controls, chosen to match the DP in this age group) to 24.2 in the 14 - 35 years age group. Considering UK control participants only, the mean number of identities known were: 16.4 (6 – 9 years); 21.2 (10 – 13 years); 25.7 (14 – 35 years); 23.1 (36 – 59 years); 20.4 (> 60 years).

Whilst number of known identities was not the main variable of interest and was measured in order to calculate personally familiar accuracy and familiarity scores for each participant, we noted that the DP group reported knowing significantly fewer famous faces than controls, $t(32.1) = 3.80$, $p < .001$ with extremely strong support ($BF_{10} = 133$) for a true group difference. This could be due to previously reported differences in media consumption (Dalrymple & Palermo, 2016).

4.4.5.1 Famous Face Test: Identification.

As shown in Figure 8, potential DPs ($n = 23$) on average, correctly identified (named) a lower proportion of faces than controls ($n = 100$). A Bayesian independent samples t test on standardised scores provided very strong evidence for a true difference between groups $BF_{10} = 204,436$. In other words, the alternative hypothesis (a true difference exists between groups) is more than two hundred thousand times as likely as the null hypothesis given the data (see Table 4.8). Unstandardised group-level accuracy descriptive statistics are shown in Table 4.6.

Figure 4.8 Group difference Famous Faces Test



Note: Showing group (self-reported DPs vs controls) in identification accuracy. Scores were calculated using only those facial identities that participants reported knowing. Each dot represents a single data point, the box shows the interquartile range (IQR) and the mid line indicates the group median score. The end of each whisker line represents 1.5 x the IQR.

Table 4.6 FFT identification accuracy by age group

Proportion correct	DP			Control		
	<i>N</i>	<i>M</i>	<i>SD</i>	<i>N</i>	<i>M</i>	<i>SD</i>
6 - 9 years	1	.33	-	15	.69	.21
10 - 13 years	1	.17	-	22	.75	.17
14 - 35 years	7	.85	.11	23	.93	.09
36 - 59 years	9	.70	.18	22	.89	.12
60 - 74 years	5	.67	.22	18	.86	.10

Note. Chance = .25 There was only one potential DP in each of the two youngest age groups so *SD* could not be calculated separately for DPs.

4.4.5.2 Famous Face Test: Familiarity

Similar to the primary outcome measure of famous face identification discussed above (“choose the name that matches the face”), the self-reported DP group was significantly less accurate at judging a known face as looking familiar vs controls (“does this face look familiar?”), $t(24.9) = 4.43$, $p < .001$, $d = 1.22$). A Bayesian independent samples t test also provided extremely strong evidence for a

true difference between groups ($BF_{10} = 6,130,000$). Notably, an almost perfect correlation was observed between participants' familiarity scores and identification scores in both the DP ($r(21) = .997, p < .001$) and control ($r(98) = .983, p < .001$) groups. This result suggests that, at least in this cohort, a sense of familiarity was not distinct from the ability to identify a face, however the 4AFC paradigm used in the identification phase would be expected to result in a higher familiarity/identification correlation than a paradigm which required participants to generate the name themselves.

We chose a 4AFC paradigm for this task because we reasoned it would be easier for children than a standard FFT paradigm. A typical FFT requires participants to provide a name, or some other unique identifying detail, from memory, unprompted, usually by writing or typing a response. We were therefore also concerned that parental interference might become an issue with such a design since children would be likely to need parental help to type and/or spell the celebrity's name and this could result in parents answering on behalf of children. Due to our design, it was possible therefore that the expected group differences would not be observed on this somewhat easier test (see Rivolta et al., 2013). We did not find this to be the case and observed strong group differences on the FFT (see Table 4.8) with, as expected, the DP group being less accurate at identifying famous faces known to them than the control group (Table 4.6). These results suggest that our test design did not unduly assist potential DPs to identify familiar faces vs controls.

4.4.6. Global face memory measures

Because it is possible that a participant may achieve a score on one test that may be higher or lower than their true ability due to chance or to random factors such as tiredness, we computed a global memory score by averaging participants' z scores across the four face memory measures of interest (Old New Faces, CFMT, CMT Difference and FFT). Descriptive statistics for the global memory scores are provided in Table 4.7. As shown in Figure 4.9, the self-reported DP group's global memory accuracy and global memory BIS means were both significantly lower than the respective control group means. This finding of lower DP performance vs controls, even when measured across multiple tests, suggests that the group differences observed for each individual face memory outcome measure are not due solely to noise, or chance, and is confirmed by much higher Bayes factors for global face

memory vs individual tests as shown in Table 4.8. Table 4.8 also shows that statistical analysis of the group differences in accuracy provided strong support for the alternative hypothesis and that the group difference was significant.

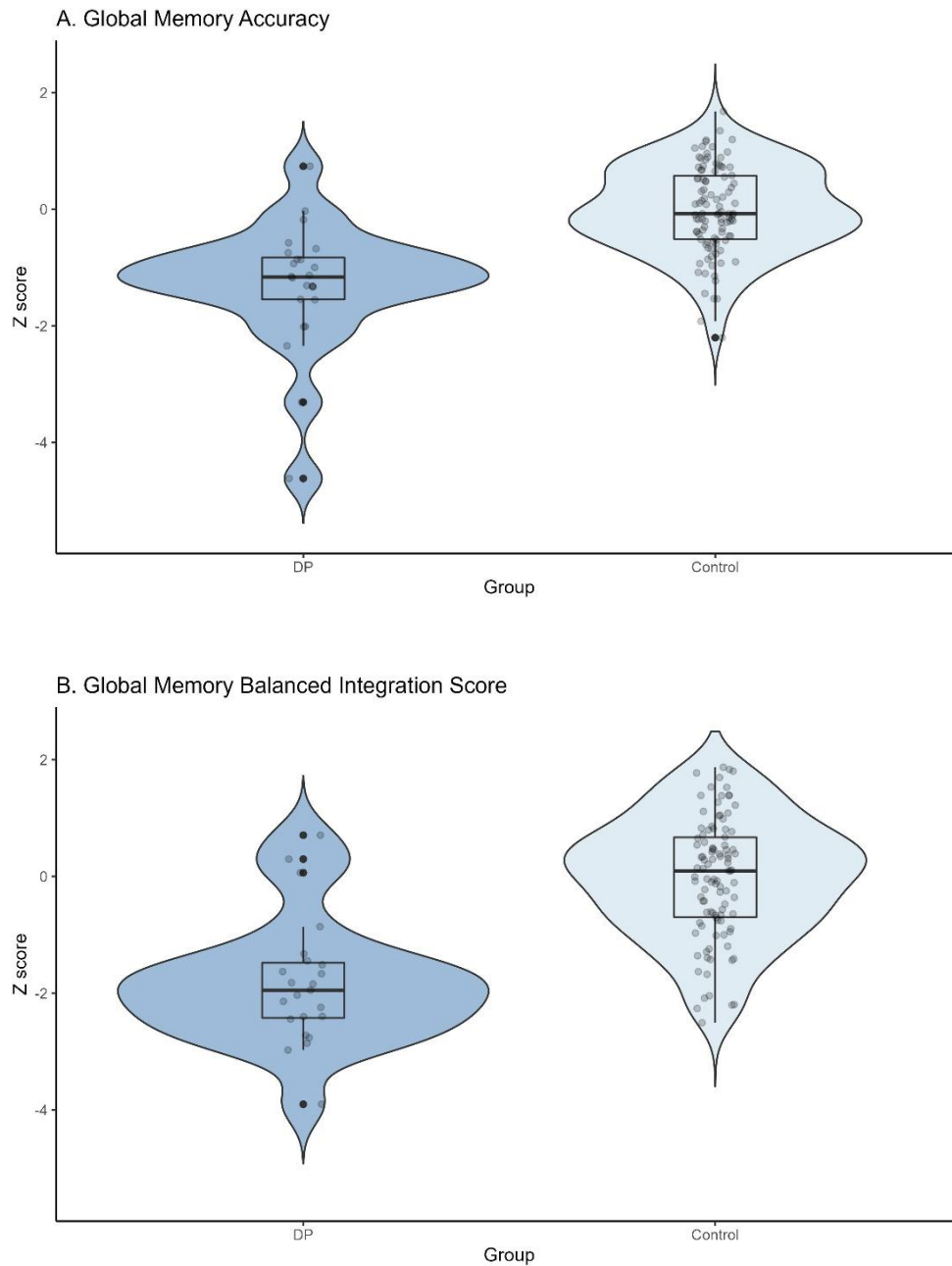
When considering group differences averaged across multiple standardised measures of face memory, the global memory BIS (i.e., accuracy adjusted for RT) showed a larger effect size than the global memory score calculated using accuracy alone (see Table 4.8). We also separately calculated group differences for adults (aged 18 years and over) and found an identical pattern of results (lower portion of Table 4.8), and very similar effect sizes. Although face processing continues to develop throughout childhood (e.g. Pascalis et al., 2011) and declines in later life (Bowles et al., 2009), our approach of using age-matched z scores to analyse group differences in this lifespan study ensures that any individual's face processing impairment is best classified with comparison to typical controls at a similar stage of development.

Table 4.7 Global face memory BIS by age group

Global face memory BIS	DP			Control		
	<i>N</i>	<i>M</i>	<i>SD</i>	<i>N</i>	<i>M</i>	<i>SD</i>
6 - 9 years	1	0.71	-	17	-0.10	1.06
10 - 13 years	1	-0.86	-	22	-0.05	0.98
14 - 35 years	7	-2.33	0.55	23	-0.05	0.97
36 - 59 years	9	-2.19	1.58	22	-0.02	1.06
60 - 74 years	6	-1.89	1.29	18	0.07	1.19

Note. There was only one potential DP in each of the two youngest age groups so *SD* could not be calculated separately for DPs. Mean scores were computed from CFMT BIS, CMT difference BIS, Old New Faces BIS and FFT identification accuracy.

Figure 4.9 Global face memory scores showing group differences (self-reported DPs vs controls) in mean face memory scores using (A) accuracy and (B)



Note: $BIS = (Z_{accuracy} \text{ minus } Z_{RT})$. Mean scores were computed from CFMT accuracy, CMT difference, Old New Faces accuracy and FFT identification accuracy. Each dot represents a single data point, the box shows the interquartile range (IQR) and the midline indicates the group median score. The end of each whisker line represents 1.5 x the IQR.

Table 4.8 Comparing effect sizes for the difference between potential DP and control group scores before (accuracy) and after (BIS) accounting for RT

	Accuracy			Balanced Integration Score		
	BF_{10}	p	d	BF_{10}	p	d
Old						
New						
Faces	28.8	.042*	-0.60	5,090,000	< .001***	-1.24
CFMT	3,290	< .001***	-1.08	2,452	< .001***	-1.03
CBMT	1.25	.044*	0.46	0.246	.766	-0.07
CMT difference	14,357	< .001***	-1.27	221	< .001***	-0.91
FFT	204,436	< .001***	-1.15	-	-	-
Global face memory	28,343,149	< .001***	-1.39	4,085,749,935	< .001***	-1.63

Adults aged 18 years and over only

	Accuracy			Balanced Integration Score		
	BF_{10}	p	d	BF_{10}	p	d
Old						
New						
Faces	6.08	.06	-0.60	80,900	< .001***	-1.34
CFMT	153	< .001***	-1.09	154	< .001***	-1.06
CBMT	1.49	.011*	0.60	0.265	.972	-0.01
CMT difference	550	< .001***	-1.27	86.0	< .001***	-1.04
FFT	2767	.001**	-1.12	-	-	-
Global face memory	57,332	< .001***	-1.36	9,166,061	< .001***	-1.78

Note. * $p < .05$, ** $p < .01$, *** $p < .001$. p values in bold remain significant after Bonferroni correction.

CMT difference = (CFMT – CBMT). RT (and consequently BIS) was not considered a relevant measure for the FFT since participants had to read names meaning RT would reflect reading speed and comprehension as well as face processing ability. Instead, we use only famous face identification accuracy (proportion of known famous faces that were correctly identified) in both versions of the global scores.

4.4.7 Which measures best predict group membership?

We used binomial logistic regression to formally assess which objective face memory measure, or combination of measures, best predicted self-reported group membership as classified by cut offs on the PI-20 or the parental report questionnaire. We developed five models (see Table 4.9). All models significantly predicted group membership, and all correctly classified over 94% of controls. Crucially however, the models' ability to correctly classify self-reported DPs varied greatly, ranging from only 16.7% (Model 5, with CFMT accuracy as the predictor) to 68.2% in the

best performing model (Model 2, comprising four separate predictor outcome variables (CFMT BIS, CMT difference BIS, Old New BIS, Famous Face identification accuracy)).

Crucially, Model 2 strongly predicted group membership and outperformed all other models. Model 2 correctly predicted 68.2 % of self-reported DPs and 94.9% of controls, explaining around 44% – 57% of the variance in subjective ratings of participants' face recognition ability.

Table 4.9 Logistic regression models predicting self-reported group membership (DP or control)

Model	Predictors	Overall model					% correctly classified	
		X^2	p	Adj R^2 McF ^a	Adj R^2 N ^b	AUC	DP	Control
1	Accuracy: CFMT, CMT Difference, Old New Faces, FFT	43.6	< .001	.41	.54	.90	59.1	97.4
2	BIS: Old New Faces, CFMT, CMT Difference; FFT accuracy	46.7	< .001	.44	.57	.90	68.2	94.9
3	Global memory accuracy	37.5	< .001	.31	.41	.87	29.2	95.1
4	Global memory BIS	45.6	< .001	.38	.50	.88	54.2	94.1
5	CFMT accuracy	20.3	< .001	.17	.25	.77	16.7	96.7

^aAdj R^2 McF = McFaddens R^2 ; ^bAdj R^2 N = Nagerkerke's

Model 2 Coefficients – group						95% Confidence Interval	
Predictor	Estimate	SE	Z	p	Odds ratio	Lower	Upper
Intercept	-3.03	.61	-4.96	< .001	.05	0.01	0.16
Old New Faces BIS	-0.63	.21	-3.00	.003**	.53	0.35	0.80
CFMT BIS	-0.08	.28	-0.28	.783	.93	0.53	1.61
FFT	-0.91	.31	-2.91	.004**	.40	0.22	0.74
CMT Difference BIS	-0.38	.28	-1.38	.168	.68	0.39	1.18

Note. Estimates represent the log odds of "group = DP" vs. "group = Control".

4.4.8 How does using BIS change classification of DP?

As seen in Table 4.8 above, the effect sizes for mean group difference were larger using BIS than using accuracy alone. We therefore next investigated how using BIS might change, or confirm, classification as a DP on our objective tests. Results are shown in Tables 4.10 and 4.11. As a reminder, to meet the initial classification as a potential DP, all participants firstly had to score atypically on the parental questionnaire or PI-20. We then sub-classified these self-reported “DPs” into three groups using the average of four objective measures: CFMT, FFT, Old New Faces and the difference between CFMT and CBMT scores (CMT difference). “Major DPs” scored more than 1.7 *SD below* their age group control mean; “Mild DPs” scored between 1 and 1.7 *SD below* the control means and “Subjective DPs” showed atypical self-report scores but scored within 1 *SD* of their age group control means, showing no objective impairment despite subjective report of face recognition difficulties. One participant (CF059) did not complete the Old New Faces test but was classified as Major DP because their CFMT z score (-3.45) indicated severe impairment and their FFT score was also below average.

As shown in Table 10, the key finding of our study was that 83.3% of the self-reported DP group showed objective face recognition deficits (20/24; 15 major, five mild) using the global memory BIS measure vs 58.3% who showed deficits when using global memory accuracy alone (14/24: five major, nine mild). This suggests that some self-reported DPs in our sample were able to achieve close to normal, or only mildly impaired, performance by trading speed for accuracy; however, once

RT was considered (using BIS), their impairment became apparent. This important finding demonstrates the value of BIS as a measure for classifying DP since these participants would otherwise have been missed if classification considered accuracy alone.

Three self-reported DP participants showed no objective impairment regardless of the measure used, we refer to these as having “subjective” DP since no objective impairment was observed. Additionally, one potential child DP showed a mild impairment ($Z = -1.32$) when considering global accuracy alone but was not considered objectively impaired once RT was accounted for (BIS) although their global memory BIS was still below average ($Z = -0.86$). In summary, many more of the participants with atypically high PI-20/parental questionnaire scores were classified as DP using BIS than using classical measures. Here we classify a major impairment as $z \leq -1.7$ but, for comparison, using a stricter $\leq -2 SD$ cut off, our data show that 12 participants would be classified as DP using the BIS measures vs only five when using accuracy alone.

Finally, we compared how classification using *individual* test scores, rather than the global face memory score shown in Table 4.10, would differ using accuracy vs BIS and these are reported in Table 4.11.

Table 4.10 Global memory accuracy and BIS classification for all self-reported DP participants

ID	Self report score	Global memory accuracy z score	Global memory BIS	DP classification (global memory accuracy)	DP classification (global memory BIS)	Change using BIS vs accuracy
AF019	-7.75	-3.31	-3.90	Major	Major	=
AF002	-7.28	-0.93	-2.24	Subjective	Major	↑
CF008	-7.18	-1.32	-0.86	Mild	Subjective	↓
AF016	-7.01	-1.32	-2.03	Mild	Major	↑
AF017	-6.60	-1.55	-1.51	Mild	Mild	=
AF006	-6.16	-2.01	-1.63	Major	Mild	↓
AF003	-6.05	-1.15	-1.66	Mild	Mild	=
AF018	-5.72	-1.33	-2.40	Mild	Major	↑
CF059	-5.59	-1.54	-2.85	Mild	Major	↑
AF009	-5.17	-0.75	-2.72	Subjective	Major	↑
AF010	-5.17	-4.62	-5.65	Major	Major	=
CF007	-5.15	-0.18	0.71	Subjective	Subjective	=
AF007	-4.95	-0.57	-2.76	Subjective	Major	↑
AF008	-4.78	-0.86	-1.84	Subjective	Major	↑
AF013	-4.73	0.74	0.30	Subjective	Subjective	=
AF060	-4.65	-1.13	-1.44	Mild	Mild	=
AF022	-4.65	-0.03	0.06	Subjective	Subjective	=
AF001	-4.62	-0.996	-1.33	Subjective	Mild	↑
AF004	-4.54	-1.17	-1.95	Mild	Major	↑
AF099	-4.38	-0.86	-2.40	Subjective	Major	↑
AF075	-3.64	-0.68	-2.44	Subjective	Major	↑
AF021	-3.09	-1.31	-2.14	Mild	Major	↑
AF098	-2.76	-2.01	-1.82	Major	Major	=
CF005	-0.67	-2.34	-2.97	Major	Major	=

Note. Major DP: $z \leq -1.7$ shown in red, Mild DP: $-1 \geq z \geq -1.69$ shown in yellow and Subjective DP: $z > -1$. Change column on far right indicates whether the DP classification severity rating changed when classifying using BIS vs accuracy; ↑ indicates rating increased, = indicates unchanged, ↓ indicates rating decreased.

Table 4.11 Individual cases compared using accuracy and BIS

ID	Age	Self report	Old New Faces	FFT	CFMT	CBMT	CMT Difference	Global face memory	Face memory measures	
									No. impaired (of 4)	Severe impairment on ≥ 2 measures?
Accuracy										
AF019	68	-7.75	-3.88	-5.22	-1.64	0.84	-2.48	-3.31	4	yes
AF002	25	-7.28	-0.96	0.41	-1.53	0.11	-1.64	-0.93	2	no
CF008	10	-7.18	0.11	-3.33	-0.78	0.50	-1.27	-1.32	2	no
AF016	23	-7.01	-1.35	-2.63	-0.83	-0.35	-0.48	-1.32	2	no
AF017	29	-6.60	-1.74	-0.04	-1.53	1.36	-2.89	-1.55	3	yes
AF006	54	-6.16	-2.70	-1.58	-2.09	-0.41	-1.67	-2.01	4	yes
AF003	55	-6.05	0.88	-1.52	-1.55	0.87	-2.42	-1.15	3	no
AF018	49	-5.72	0.88	-4.30	-0.91	0.08	-0.99	-1.33	1	no
CF059	16	-5.59	NA	-0.58	-3.45	-2.84	-0.60	-1.54	1 of 3	no*
AF009	42	-5.17	0.88	-0.63	-1.23	0.77	-2.00	-0.74	2	no
AF010	53	-5.17	-3.84	-2.63	-3.05	-0.12	-2.93	-4.62	4	yes
CF007	8	-5.15	0.30	-1.74	0.90	1.06	-0.16	-0.18	1	no
AF007	31	-4.95	-0.56	0.34	-0.53	1.02	-1.54	-0.57	1	no
AF008	71	-4.78	-0.17	-1.37	-0.52	0.84	-1.36	-0.85	2	no
AF013	44	-4.73	0.88	0.92	0.92	0.67	0.24	0.74	0	no
AF022	64	-4.65	-0.17	0.72	0.04	0.75	-0.70	-0.03	0	no
AF060	61	-4.65	-0.92	-1.92	-0.42	0.84	-1.27	-1.13	2	no
AF001	51	-4.62	-0.91	-0.24	-1.23	0.38	-1.60	-1.00	2	no
AF004	31	-4.54	0.62	-1.87	-0.93	1.58	-2.51	-1.17	2	yes
AF099	70	-4.38	-3.14	NA	0.33	0.09	0.23	-0.86	1 of 3	no*
AF075	45	-3.64	0.28	-1.40	-0.80	-0.02	-0.78	-0.67	1	no
AF021	50	-3.09	0.88	-2.03	-1.55	0.97	-2.52	-1.31	3	yes
AF098	67	-2.76	-3.14	-1.45	-1.74	0.00	-1.74	-2.01	4	yes
CF005	17	-0.66	-1.35	-1.98	-2.34	1.36	-3.70	-2.34	4	yes
Proportion impaired			0.35	0.65	0.50	0.04	0.67	0.58		0.33
Balanced Integration Scores										
AF019	68	-7.75	-6.92	-5.22	-1.72	0.02	-1.74	-3.90	4	yes
AF002	25	-7.28	-4.35	0.41	-2.70	-0.37	-2.33	-2.24	3	yes
CF008	10	-7.18	0.37	-3.33	-0.47	-0.48	0.01	-0.86	1	no
AF016	23	-7.01	-2.11	-2.63	-1.52	0.35	-1.87	-2.03	4	yes
AF017	29	-6.60	-3.17	-0.04	-0.17	2.50	-2.67	-1.51	2	yes
AF006	54	-6.16	-2.42	-1.58	-1.16	0.18	-1.34	-1.63	4	no
AF003	55	-6.05	-0.60	-1.52	-2.25	0.04	-2.29	-1.66	3	yes
AF018	49	-5.72	-1.17	-4.30	-2.88	-1.65	-1.23	-2.40	4	yes
CF059	16	-5.59	NA	-0.58	-5.53	-3.07	-2.46	-2.85	2 of 3	yes
AF009	42	-5.17	0.03	-0.63	-5.56	-0.85	-4.71	-2.72	2	yes
AF010	53	-5.17	-12.13	-2.63	-3.48	0.89	-4.37	-5.65	4	yes
CF007	8	-5.15	0.23	-1.74	1.71	-0.91	2.63	0.71	1	no
AF007	31	-4.95	-2.88	0.34	-4.67	-0.85	-3.82	-2.76	3	yes
AF008	71	-4.78	-5.72	-1.37	-1.23	-2.19	0.96	-1.84	3	yes
AF013	44	-4.73	-0.80	0.92	0.79	0.49	0.29	0.30	0	no
AF022	64	-4.65	-0.28	0.72	-0.51	-0.83	0.32	0.06	0	no
AF060	61	-4.65	-2.94	-1.92	-0.07	0.77	-0.85	-1.44	2	yes
AF001	51	-4.62	-2.45	-0.24	-0.35	1.92	-2.26	-1.33	2	yes
AF004	31	-4.54	-0.67	-1.87	-2.20	0.85	-3.05	-1.95	3	yes
AF099	70	-4.38	-7.53	NA	0.22	0.11	0.11	-2.40	1 of 3	no*
AF075	45	-3.64	-3.07	-1.40	-3.83	-2.37	-1.46	-2.44	4	yes
AF021	50	-3.09	-1.38	-2.03	-2.03	1.08	-3.11	-2.14	4	yes
AF098	67	-2.76	-3.84	-1.45	-1.02	-0.05	-0.97	-1.82	3	no
CF005	17	-0.66	-2.27	-1.98	-2.73	2.18	-4.90	-2.97	4	yes
Proportion impaired			0.70	0.65	0.67	0.17	0.67	0.83		0.71

Note. FFT = Famous Faces Test, CFMT = Cambridge Face Memory Tests, CBMT = Cambridge Bicycle memory Test, CMT difference = CFMT – CBMT, Global face memory = standardised mean average. BIS was not calculated for the FFT so the results are the same as for accuracy alone but are shown again in the second table for ease of

comparison. Major impairment $z \leq -1.7$ shown in red, mild impairment $-1 \geq z \geq -1.69$ shown in yellow. All scores are standardised and centred on age-matched control means. The two farthest right columns show firstly the number of independent face processing measures on which a participant was impaired (mild or major) and secondly whether the participant showed an impairment of at least 2 *SD* below control means on at least two of the four face memory measures, * indicates that participant completed three rather than four tasks.

4.5 Discussion

Twenty-four participants with self-reported face recognition problems (DPs) and 110 age matched controls completed four tests measuring different aspects of objective face recognition ability alongside a self-report questionnaire. To examine the best way of best identifying DP, we used both traditional accuracy measures (proportion correct) and a novel integrated measure, the balanced integration score (BIS), which adjusts accuracy to take account of RT while controlling for speed-accuracy trade-offs. At an individual participant level, 15 individuals who self-reported face recognition difficulties were classified as having a major face recognition impairment using BIS, but only four were classified as having a major impairment when using accuracy alone. Overall, observed between-group effect sizes for computed global (averaged) memory scores were also larger using BIS than using accuracy alone (see Table 4.8).

4.5.1 Is self-report a useful initial classification measure for DP?

Before comparing how the objective face memory measures were able to detect and classify DP, we first checked that self-report - which we used to make a preliminary classification of participants into potential DPs or likely controls - was in fact a valid basis on which to make this preliminary classification. The extent to which self-report is a valid indication of face recognition ability has been much debated. Furthermore, direct comparisons are difficult to make across studies using different self-report questionnaires. Nonetheless, previous studies have reported that self-reported and objective face recognition measures are, at best, modestly correlated in naïve typical perceivers (Bobak et al., 2019; Matsuyoshi & Watanabe, 2021; Palermo et al., 2017) and in individuals who were aware that they met the diagnostic criteria for DP before completing the self-report questionnaire (Murray & Bate, 2019). By contrast, studies of individuals with DP, mainly using the PI-20 (Shah et al., 2015), found that DPs *do* have insight into their own face processing

difficulties (Burns et al., 2022; Gray et al., 2017; Livingston & Shah, 2018; Tsantani et al., 2021; Ventura et al., 2018), or at least the fact that their face recognition ability was poor relative to others (Palermo et al., 2017). Our data also support the use of self-report measures for classifying DP. All the potential DP participants except one made contact with our lab because they (or a parent) believed they struggled with face recognition or had a family history of DP, but, unlike the DP participants in the Murray and Bate (2019) study, participants in the present study were *not* told prior to completion of the self-report questionnaire whether they met the ‘diagnostic’ threshold for DP. They were arguably therefore less likely to be influenced by a DP ‘diagnosis’ when subjectively rating their face recognition ability. However, four individuals reporting poor face recognition ability had previously participated in other face recognition studies which may have provided insight into their ability.

It has been reported that women rated their prosopagnosia symptoms as more severe than men (Murray & Bate, 2019) but we found no support for this. Among adult DPs ($n = 20$; 15 female, 5 male) who showed both objective and subjective impairment, we found no evidence of gender differences in PI-20 scores ($d = 0.32$, $p = .604$, and anecdotal evidence for the null hypothesis ($BF_{10} = 0.52$). It is possible that there is an interaction between gender and status (naive; informed) that should be explored further by researchers who disclose status to participants prior to administering a self-report questionnaire but this was not a relevant issue in our study.

Regression analysis showed that, overall, the standardised subjective rating score was a significant predictor of global face memory BIS, $F(1,124) = 57.7$, $p < .001$ explaining around 31% of the variance in objective scores (adjusted $R^2 = .31$). When considering adults and children separately, unsurprisingly adults’ subjective rating of their *own* face recognition ability was a better predictor of group than parental report of *their child’s* face recognition ability, explaining around 40% and 6% respectively of the variance in global face memory BIS. Binomial logistic regression analysis showed that parental report was nevertheless a significant predictor of group suggesting that parents did, on average, have insight into their children’s face memory ability in binary terms, i.e., whether it was much worse than average or not. However, parents’ ability to predict more precisely their child’s ability *relative* to their age group, as measured by global face memory BIS z score, was only just above chance ($p = .044$).

One factor that may have limited parental ability to accurately judge their child's face recognition ability was the parent's own face recognition ability. Face recognition is a highly heritable ability (Wilmer et al., 2010) and many reported cases of DP have a known family history (De Haan, 1999; Duchaine et al., 2007; Gruter et al., 2008; Lee et al., 2010; Schmalzl et al., 2006). In our study, the face recognition ability rating for three of the four child DP candidates was provided by a parent who themselves reported having difficulty recognising faces. Thus, it is likely that these parents lacked an accurate reference point for judging typical, and consequently atypical, face recognition ability.

4.5.2 DP performance on Old New Faces.

In line with previous work (Dalrymple et al., 2014), we found that Old New Faces was a useful test for classifying DP. Although the test could be criticised as being more of an image memory than a face recognition test since the images used at study and test are identical, Dalrymple, Garrido and Duchaine (2014) reported that all adult DPs ($n = 16$) showed impaired accuracy on the Old New Faces test. By contrast, 0/16 were impaired at a matched old/new houses test and only 1/16 was impaired in a matched old/new horses test. Among children, 4/6 DPs were significantly impaired vs controls on the Old New Faces test but – similar to adults – 0/6 were impaired on the matched object task, in this case an old/new flowers task. By contrast, the authors observed that 10/16 adult DPs were *unimpaired* on the Cambridge Face Perception Test accuracy (CFPT, Duchaine et al., 2007). Together, these results suggest firstly that the Old New Faces test is useful for identifying DP and that the impairment detected by the task is face specific. Secondly, the results suggest that the presence of a memory demand in the Old New Faces task - even when using the same image at study and test - produces different patterns of impairment compared with the CFPT which is also a test of face matching but without memory demands. Overall, Dalrymple et al. (2014) show that the Old New Faces task is a useful source of converging evidence of face recognition difficulties when the CFPT is not suitable. Our data support these findings, namely that the DP group, on average, produced significantly lower accuracy and BIS scores than the control group. Additionally, on the Old New Faces task we found stronger Bayesian evidence and larger effect sizes for a group difference for BIS vs accuracy alone.

4.5.3 DP performance on Famous Faces Test

Despite not being able to calculate BIS for FFT, since reading speed would have confounded RT, we nevertheless found this novel FFT to be a useful measure for classifying DP using identification accuracy. This measure produced the strongest accuracy effect size ($d = 1.15$, $p < .001$) of the four individual face memory tests we administered. Additionally, Bayesian analysis showed extremely high evidence for a group difference ($BF_{10} = 204,436$), again the highest of any single test (Table 4.8). To check that group differences were not driven by a small number of individuals with extreme scores, we conducted individual case analysis on all accuracy measures (Tables 1 and 11). On the FFT, 19/23 potential DPs produced identification accuracy scores below the age group mean, and of these, 9/23 scored significantly below mean control accuracy. A previous large-scale study (Bate et al., 2019) found evidence of a dissociation between memory for familiar faces (FFT) and memory for newly-learned faces in 63 of 165 individuals and suggested that long term memory for familiar faces (as indexed by a FFT) may be selectively impaired in DP. Our data support the use of an FFT as one of several measures to classify DP, even though BIS could not be calculated.

4.5.4 Do accuracy or BIS measures best predict group membership?

The logistic regression model that included three BIS face memory measures plus FFT accuracy (Model 2, see Table 4.9) was *the most sensitive* and classified the most self-reported DP participants (68.2%) as being objectively impaired. Importantly, Model 2's *highest sensitivity* did not come at the cost of reduced *specificity* as the ability of this model to classify controls was very similar (94.9% vs 96.7% for Models 2 and 5 respectively). In comparison, CFMT accuracy (Model 5), which is the measure traditionally used to diagnose DP, classified only 16.7% of our self-reported DP group. Although AUC was only slightly lower than this for Model 3, (AUC = .87, global face memory accuracy) and Model 4 (AUC = .88, global face memory BIS), these models were much less sensitive than Model 2, correctly classifying only 29.2% and 54.2% of self-reported DPs respectively.

One DP participant produced very low scores on three of the four tests. We therefore checked to ensure that this potential outlier was not unduly influencing results by repeating the logistic regression modelling with this participant removed.

Although exact values changed, the pattern of results was very similar, and Model 2 remained the best model for classifying DP.

We also reran these analyses on a sub-sample of only the self-reported DPs who showed mild or major impairment on the global face memory accuracy score, that is without the ten participants classed as “subjective DPs” and found the same pattern of results (see supplementary materials sections 2 and 3). Again, Model 2 with BIS as the outcome measure remained the best model ($AUC = .98$, $p < .001$). As would be expected from this approach (i.e. excluding the potential DPs who achieved typical accuracy due to atypical RT) the group effect size differences when comparing accuracy and BIS were smaller than when we analysed the full sample of potential DPs. This is because in this alternative approach, only the self-reported DPs with impaired accuracy are included in the analysis and so it logically follows that the utility of accuracy as a classification measure would improve. Nevertheless, the fact that BIS measures – even among this sub-sample - were more sensitive than accuracy measures for classifying DP further strengthens the value of BIS.

Our results confirm the importance of considering RT as well as accuracy when classifying DP and these findings are in line with recent studies (Fysh & Ramon, 2022; Stacchi et al., 2020) and a large literature review by Geskin and Behrmann (2018).

However, it is possible that RT (and therefore BIS) is more useful in some tasks and paradigms than others. Our data show that for the CFMT, a test specifically designed to detect DP, BIS added little or no additional information compared to accuracy alone. Effect sizes for both measures were similar, and indeed slightly larger for accuracy. Others have also questioned whether RT always adds value over accuracy. A recent study (DeGutis et al., 2022) investigating both accuracy and response times on a face matching task, the Benton Face Recognition Test (BFRT-c, Rossion & Michel, 2018) reported that RT alone did not reliably predict group membership. The BFRT-c is an updated version of the original face matching test (Benton et al., 1983) that was specifically designed to emphasise speed as well as accuracy. However, there are two important differences between the BFRT-c and the tasks used here. Firstly, despite its name, the BFRT is a perceptual task since it involves no memory demands. Secondly, the BFRT-c requires participants to click up to *three* faces from a choice of six meaning that motor control is likely to influence RT more than it would in our tasks where participants had to make a single response.

Finally, as the authors explain, the BFRT-c design means that it is not possible to analyse RT on correct trials only as is common practice in other face cognition tasks. Notably RT from correct trials *only* is the measure used to calculate BIS (Liesefeld & Janczyk, 2019). Despite this, some theoretical papers have incorrectly used RT on *all* trials when comparing BIS with other integrated measures of speed accuracy which could lead to confusion about how to calculate BIS (for a fuller explanation see Liesefeld & Janczyk, 2022). BIS may therefore not be informative on tasks where RT measures include both incorrect and correct trials.

Our findings support the use of BIS as an *integrated* measure that adjusts accuracy to account for RT. However, we make no claims about the use of RT as a sole measure which was the question DeGutis and colleagues (2022) investigated. As the authors correctly caution, if researchers wish to use RT *instead* of accuracy on a given task, RT should first be validated as a measure. In the present study we were instead interested in whether accuracy and RT *together* might be more informative than accuracy alone. Our data show that accuracy and RT together (BIS) explained more of the variance in self-reported face recognition ability than accuracy alone and, additionally, showed greater sensitivity for classification of DP than accuracy alone. A third benefit of BIS was that for the global face memory score and Old New Faces task, the observed effect sizes were larger for BIS than accuracy (on the CFMT there was little difference). This is an important consideration in neuropsychology research where effect sizes are typically modest and sample sizes often small, thereby limiting power (McIntosh & Rittmo, 2021). Using BIS thus provides a practical solution to ensuring that no impaired participants are omitted from analyses. This increased power to detect differences between different populations will allow for better understanding of deficits underpinning DP and provide pathways to effective training. Better classification will also ensure that neuropsychological research is more ethical. The British Psychological Society Code of Human Research Ethics calls for maximising the benefit to participants (point 2.4) from inception to dissemination (Oates et al., 2021). Thus by improving the methods of studying DP, our work contributes to that principle of ethical research.

In addition to offering a practical solution, BIS, as an integrated measure of both accuracy and RT, is a more ecologically valid approach to identifying DP than accuracy alone. This is because in typical social interactions, the amount of time an individual with DP has to make a correct identification is effectively the time it takes

the person they are interacting with to recognise the DP. It is time limited. If the DP has not recognised the face before the person greets or speaks to them, then this will appear to be a failure of recognition (even if - given much longer to study the face - the DP might have been able to identify the person). Response latency and accuracy are therefore *both* important elements with regard to ecological validity.

4.5.5 Use of global scores and choice of cut off

In their editorial of a special issue on DP, Bate and Tree (2017) argued that seeking converging evidence of impairment across *multiple* tests provides more compelling evidence of a true deficit as well as mitigating the risk that unimpaired controls may be misclassified as DP. The same argument is made for super recogniser research (Bate et al., 2018; Ramon, 2021). We therefore administered four separate tests and used these scores to compute a global face memory score (see table 4.7). Global scores showed large differences between DPs and controls and thus appear useful. However, logistic regression showed the global face memory scores produced slightly lower AUCs than Models 1 and 2 which used non-averaged scores from the individual tests. Our results therefore suggest that although global effect sizes are larger than those from individual tasks and therefore will increase the power to detect group level differences, the four *individual* BIS measures (CFMT BIS, CMT Difference BIS, Old New Faces BIS and FFT accuracy) best predicted group membership. This could be because averaging results can attenuate the insights provided by multiple individual test scores.

Researchers may need to decide whether they wish to prioritise sensitivity (classification), or ability to detect group difference (effect sizes). Individual BIS measures were slightly better for classification purposes and global measures showed larger effect sizes which may be an important consideration when sample sizes are small.

Finally, we also compared the patterns of impairment across the four independent face memory measures at the individual case level. Using this alternative, and more commonly used approach, results once again supported the overall conclusion that BIS is a more sensitive measure for classifying DP than accuracy alone. Data showed that more than twice as many (71%) of the self-reported DP participants showed severe objective impairment ($< -2 SD$) on at least two individual face memory BIS measures compared with only 33% who were severely

impaired on at least two accuracy measures. BIS therefore appears equally valuable whether using a global (averaged) score or multiple independent measures of face memory. Notably, although we used more liberal cut offs of $-1 SD$ to classify mild and $-1.7 SD$ for major impairment to classify potential DPs, *all* participants who were classified as impaired (mild or major) on global face memory BIS also showed *severe* impairment ($< -2 SD$) on two individual BIS measures suggesting that a more liberal cut off is justified provided multiple objective tests are administered.

4.5.6 Online vs lab-based data

We compared test results with previous literature and found that our control group mean accuracy data was broadly very similar, almost always within $1 SD$, to published lab-based studies testing similar age groups, thus giving us confidence in our results. However, one important difference we wish to highlight was that we observed a greater score variability, and thus larger standard deviations, compared with previous literature. For example, CFMT accuracy in both the 14-25 and 36-59-years control groups was 81%, almost identical to previously reported scores of 80.4% (Duchaine & Nakayama, 2006). By contrast, the standard deviations in these control groups were 14% and 13% respectively vs 11% in the original Duchaine and Nakayama (2006) study. Among control participants aged 60-74 years, mean accuracy data ($72\% \pm 15\%$) was again very similar to published age group norms for 60-69 year-olds of $70.14\% \pm 12\%$ (Bowles et al., 2009) but standard deviations were once again higher. We also observed similar patterns in the child data.

This finding is highly relevant for classification because is DP is ‘diagnosed’ or classified using the mean and standard deviations of the control group. For online research, we therefore caution against using “standard cut-offs” on popular tests such as the CFMT as these may not be valid for online data collection. Instead, online specific norms should be used. Further, our data support the need to use age group norms (Bowles et al., 2009) since our data also showed that accuracy and standard deviations both varied by age group resulting in different cut offs for mild and major impairment in each age group. This finding is also in line with results from the large sample of 165 DPs (Bate et al., 2019).

4.6 Conclusion

In conclusion, our key finding is that, whether using multiple individual scores or an averaged global score, and regardless of whether the cut off applied was $-1.7 SD$ or $-2 SD$, BIS was a more sensitive measure of difference between self-reported DPs and controls than accuracy alone. Furthermore, BIS better predicted group membership compared with accuracy alone. Our data show that using four measures (Old New Faces BIS, CFMT BIS, CMT difference BIS and FFT identification accuracy) alongside the PI-20 to classify DP *considerably improved sensitivity* (captured more DPs) with *no reduction in specificity* (did not decrease the proportion of controls correctly classified) compared with traditional accuracy measures. These results have important applied value for researchers who must identify and classify DP. Using measures that show strong effect sizes can increase power, an important consideration in research into rare conditions such as DP where large sample sizes are difficult to achieve. Improved classification will also allow a better understanding of the underpinnings of DP and avoid unnecessary exclusion of participants whose impairments may be masked by speed-accuracy trade-offs.

Acknowledgements

Thanks to Dr Gabriele Chierchia, Dr Kirsten Dalrymple, and Professor Brad Duchaine for kindly sharing their materials and online experimental set ups with us and to Professor Jane Riddoch for permission to create an online version of the Birmingham Object Recognition Battery. We would also like to thank undergraduate student Mhairi Webster for helping to develop the Famous Faces Test and all the participants and parents for taking part. We also thank the funders (ESRC and British Psychological Society Cognitive Section to Judith Lowes and the Leverhulme Trust ECF-2019-416 to Anna Bobak) for funding this research.

Chapter 5: An in-depth investigation of face perception in developmental prosopagnosia

Abstract

Developmental prosopagnosia (DP) results in difficulty recognising familiar faces but little is known about the exact stage of the face processing system where difficulties occur, or indeed whether this is identical in all cases of DP. DPs ($n = 20$) completed six tasks measuring identity and nonidentity aspects of face perception. Overall, the DP group showed impaired face perception, but their object (house) perception was similar to that of typical controls. Widespread impairments occurred at early, nonidentity, stages of face processing including face detection and face gender categorisation. Importantly, early-stage impairments were observed in most (13/20) of the DP group. A subgroup of DPs ($n = 6$) was impaired on ≤ 1 face perception task indicating broadly typical face perception and thus support for the idea of apperceptive (affecting both face perception and face memory) and mnemonic (affecting face memory only) sub types of DP. Three participants showed typical nonidentity face perception and were impaired only on the Cambridge Face Perception Test which tests facial identity perception. Among both DPs and controls, intraclass correlations showed clear intra-individual differences confirming that data from a single task (e.g. CFPT) cannot be taken as evidence of broader face perception ability.

5.1 Introduction

Developmental prosopagnosia is characterised by a severe, lifelong difficulty recognising familiar faces, often including those of friends and family, despite otherwise normal vision and IQ, and an absence of obvious brain damage.

5.1.1 What is the nature of the impairment in DP? A memory impairment or perceptual impairment, or both?

Typically, DP is studied within the framework of two leading face recognition models (Bruce & Young, 1986; Gobbini & Haxby, 2007; Haxby et al., 2000). These models posit that face processing is largely hierarchical, in that an early impairment of perception (or encoding), will result in downstream deficits at the retrieval stage (memory) and thus difficulty recognising familiar faces, but not the other way round.

Although such a dissociation between perception and memory was identified in visual agnosia (Lissauer, 1890) and in acquired prosopagnosia whereby memory for faces can be selectively impaired, leaving face perception unaffected (Davies-Thompson et al., 2014; De Renzi et al., 1991), it is still not clear if DP is primarily an impairment affecting face memory, face perception, or both. The most commonly-proposed subtypes of DP are mnemonic, sometimes called associative (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 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 a severe face memory difficulty).

The existence of potential subtypes has important theoretical and applied implications for the understanding and treatment of DP, an incurable condition that is estimated to affect between 1% and 8% of the population depending on the classification methods used (Bennetts 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 task most used to assess face perception in DP is the Cambridge Face Perception Test (CFPT, Duchaine et al., 2007) which tests face *identity* perception. Following guidelines (Dalrymple & Palermo, 2016) the CFPT is often used to diagnose or classify DP and at group level DP performance is usually impaired on this task (Bowles et al., 2009, for a review of effect sizes see DeGutis et al., 2023). However, looking solely at group level performance can mask informative individual differences *within* the DP sample; previous literature has found that between one third and one half of individual DPs show a significant impairment on the upright CFPT (Biotti, Gray, et al., 2017; Dalrymple, Garrido, et al., 2014), with the remainder performing within typical norms and therefore deemed ‘unimpaired’ despite often showing performance that is considerably below average (-1 to -2 *SD*). 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 and colleagues (2023) found that DPs with *milder* face memory impairments showed

greater impairment on the CFPT than DPs with more severe face memory deficits. 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.

A further limitation of the CFPT was raised by Bate and colleagues (2019) who suggested that, in the same way that there are perceptual sub processes of face perception, face memory itself may not be a unitary construct. PCA analysis of data from a large study ($n = 165$) of potential DPs (Bate, Bennetts, et al., 2019) found that the CFPT (Duchaine et al., 2007), testing face identity perception, 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, Garrido, et al., 2014) to dissociate face perception and face memory 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.

Similarly, there is uncertainty around whether holistic processing, whereby faces are processed as an indivisible whole, is atypical in DP. Aberrant holistic processing has been widely suggested as a key underlying deficit in DP and a recent scoping review (Nørkær et al., 2024) identified 28 separate studies of holistic processing in DP. Some studies (e.g., (Avidan et al., 2011) report atypical holistic face processing, or atypical holistic processing of the eyes (DeGutis, Cohan, et al., 2012) while others found evidence of normal holistic processing (Garrido et al., 2008; Susilo et al., 2010). However, the term holistic processing is not well defined (Rezlescu et al., 2017) and three different paradigms are commonly used to measure face holistic processing; composite faces (Young et al., 1987), face inversion effect (FIE, Yin, 1969)), and part-whole tasks (Tanaka & Farah, 1993). Rezlescu et al. (2017) found, of these, only the FIE moderately predicted face identity perception in typical perceivers and argued that the inversion, composite and part-whole effects tap distinct mechanisms, and thus that holistic processing is not a unitary construct. Consistent with this, Boutet et al. (2021) used factor analysis to assess whether three paradigms commonly believed to index holistic processing do in fact tap the same underlying construct. Results indicated a two-factor solution with the part-whole

effect and a configural featural detection task loading onto one factor and the composite face effect loading onto the second factor. They argue this provides evidence against a common holistic mechanism for face processing. Further, the authors suggest that the observed individual differences in holistic processing might be accounted for by a differential ability to process the *featural* aspects of faces. I do not address holistic processing in detail here since my interest is mainly in upright face perception. However, I use the FIE as an index of holistic processing since based on Rezlescu et al.'s findings (2017) discussed above, see also Chapter 2.

5.1.2 Limitations of the literature to date

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 low reliability (e.g., CFPT), from which it is difficult to make wider generalisations about the condition. We know that even among typical perceivers, there are low correlations across different face matching tasks (Bobak et al., 2023) suggesting that stimulus properties and experimental design must be considered when interpreting results. A similar lack of consistency across different face tasks has been found among super recognisers 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. At the very least, the use of a more comprehensive test battery can eliminate three alternative possibilities for the current mixed findings on face perception deficits in DP (Corrow et al., 2016). The mixed findings could result from i) measurement error on single task, ii) the fact that the single stage of face processing tested differed across studies making comparison impossible, or iii) paradigm-specific effects.

A handful of published DP studies have administered a battery of perceptual tests and these provide useful insight into the perceptual sub processes underpinning face identity recognition (e.g., Esins et al., 2016). Some were limited by small sample sizes (Ulrich et al., 2017), or by the use of face stimuli whose effects have been shown by Yovel and Duchaine (2006) to be due to stimulus properties (e.g. Behrmann et al.,

2005; Schmalzl, Palermo, & Coltheart, 2008; Ulrich et al., 2017) because tasks only tapped one subprocess i.e. face identity perception cannot therefore provide insight in the nonidentity aspects of face processing.

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 2SD 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., 2022; DeGutis et al., 2023, 2024) and should therefore be included in DP research. My recent work also suggests that slower reaction times can lead to ‘typical’ accuracy scores and mask impairments in DP (Lowes et al., 2024). 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, I included participants with mild objective impairment but who report severe day to day difficulties.

In summary, recent reviews of DP (Bate & Tree, 2017; Robotham & Starrfelt, 2018) highlighted the need to systematically test the different sub processes and levels of processing involved in face recognition to better understand the precise nature of the deficits that occur in DP. In addition, recent 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. I addressed many of these concerns in the present study.

5.1.3 The present study

In this study, I took a novel approach to characterising perceptual impairment in DP. I assembled a comprehensive test battery that hierarchically taps each of the multiple stages of face processing mapped in a recent literature review of tests of upright face processing in used in DP (Robotham & Starrfelt, 2018) and is informed by Bruce & Young model of face recognition (Bruce & Young, 1986; Young & Bruce, 2011). Importantly, I used *multiple* measures of face perception alongside a matched object perception task, thus allowing a much richer understanding of the precise nature of the face perception deficits in DP. Such insights can inform the

design of bespoke remedial interventions, which to date have typically focused on improving DPs' ability to better discriminate small differences in facial configurations.

The aim of this study was to precisely map the perceptual underpinnings of DP and, importantly, to do this at both a *group* and *individual* level. As recommended by Bate and Tree (2017), I tested participants across the adult lifespan (age 14 to 71 years). My study design mitigated the risk of overestimating the prevalence of perceptual impairments in DP by using subjective report and only face *memory* tests for classification (Lowes et al., 2024a). I 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 *nonidentity face perception* (face detection, face categorisation, gender and age categorisation, feature and configuration change discrimination) followed by *face identity perception*. I then compared participants' performance on these perceptual measures to their performance on four face memory measures.

The pre-registered research questions (available at <https://osf.io/qne8d/>) to be addressed in this study were:

- RQ1: Are all, or some, DPs impaired at face detection?
- RQ2: Are all or some DPs impaired at facial gender categorisation?
- RQ3: Are some or all DPs impaired at judging facial age?
- RQ5⁹: Do children and/or adults with DP show deficits in configural face processing? And are any deficits observed domain specific or process specific?
- RQ6: Is face perception ability related to face identity recognition?

The paper is structured as follows. First (section 5.3), I provide a general methods overview of participants, procedure and statistical approaches that apply to all my research questions. Next (section 5.4), I present each research question

⁹ Pre-registered research question four could not be completed due to lack of children (see 2. Data transparency) but for ease of comparison with the original preregistered research questions and hypotheses we use the original question numbers here and omit question 4 and hypothesis 4.

covering motivation, hypotheses, test methods, results and discussion and report test reliability. In Section 5.6, I consider the *patterns* of individual level results across all tasks (I provide formal single case analysis in Appendix 4 Tables S4a-e). The final section (section 5.6) is a general discussion of this chapter.

5.2 Data transparency statement

Although this study is exploratory in nature, it tests a large number of sometimes competing hypotheses. For transparency, I therefore decided to preregister the study after data collection had started but prior to analysis. The study hypotheses, analysis plan, data and R scripts are available via The Open Science Framework https://osf.io/s57zq/?view_only=ec7f0738fccc4cd8a7f075dc9b5bc932). Here I report findings from the main perceptual test battery that I assembled to test the preregistered hypotheses. Both of our previous papers reporting classification (Lowes et al., 2024a) and screening (Lowes et al., 2024b) of this cohort show that the DP group adopted a different speed-accuracy strategy to the control group, preferring to proceed more slowly and carefully compared to controls. To account for these differential speed-accuracy trade-offs which masked the true extent of face identify recognition ability deficits among many DPs, I therefore report an additional measure that was not preregistered – the balanced integration score or BIS (Liesefeld & Janczyk, 2019, 2022).

Results from the four tests used to initially classify participants as DP or typical controls have been reported in full elsewhere (Lowes et al., 2024a) and are provided in Table 5.3 below. Here I use the classification method that was found to be most sensitive for classifying the DP group rather than the pre-registered method that performed less well when the two were compared.

Some hypotheses were not tested due to lack of children in the sample¹⁰. Preregistered hypothesis 5d was included by error; it relates to a test that was ultimately not included in my battery for reasons of overall test duration. The preregistered analysis plan specified one-way tests for multiple measures that tested

¹⁰ 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.

a directional hypothesis, however I decided to revert to the more conservative two tailed tests for all group and single case analyses.

5.3 General methods

5.3.1 Participants

DP participants (DPs) were 20 individuals aged 15 – 71 years (6 men; 14 women). Participant demographics are shown in Table 5.1 below. No gender differences were observed between groups ($X^2 = 1.11$. $p = .292$) and so data were collapsed across genders for analysis.

Table 5.1 Participant demographics

Age group	Gender	Group		Total
		Control	DP	
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

Results from four individual face memory measures that were used to classify this sample have been reported separately (Lowes et al., 2024a) and are shown in Table 5.2. Participants were classified as DP if they reported lifelong subjective face recognition difficulties (a score of at least 61 on the PI-20; Shah, Gaule, Sowden, et al., 2015) or an adapted version for parents of adolescents) and objective face memory impairments (scoring at least 1 *SD* below the age matched control mean on a global face memory score). The global face memory score was calculated by averaging z scores of 1) balanced integration scores (BIS) from the Old New faces task (Dalrymple, Garrido, et al., 2014), 2) the Cambridge Face Memory Task

(CFMT) (Duchaine & Nakayama, 2006), 3) a CMT difference score calculated by subtracting CFMT scores from Cambridge Bicycle Memory scores (Dalrymple, Garrido, et al., 2014), 4) the proportion of famous faces correctly identified (Lowes et al., 2024a).

On this global objective face memory measure, 15/20 DPs showed a *major* impairment (≤ -1.7 SD) versus age matched controls and 5/20 showed a *minor* impairment (< -1 SD). Notably, all the DPs with a major impairment and 4/5 DPs with minor impairment showed a major impairment (≤ -1.7 SD) on two *separate* objective face memory measures (Table 5.2) and would therefore be widely considered to have DP. The z scores for the fifth ‘mild’ DP (AF006) were -2.42 on Old New Faces and between -1 and -1.7 on each of the CFMT, FFT and the CMT difference score and they were therefore considered to have consistently below average face memory performance and classified as DP.

Four additional potential DPs (CF007, CF008, AF013 and AF022) completed the test battery but are *not* included here either because they did not consistently show objective impairment on the classification measures, despite reporting severe difficulty recognising faces in everyday life. Thus, all the DP participants reported here meet the widely accepted classification criteria of subjective difficulty *plus* objective impairment on a minimum of two tests.

Some controls did not complete all tasks due to participant or computer error, or drop out part way through the battery, so group comparison analyses are calculated using pairwise deletion¹¹. To allow a meaningful assessment of face perception ability *across multiple tasks* I excluded from 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; 34 women).

All participants were living 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). Preregistered exclusion criteria were: Any neurodevelopmental or neurological condition; a learning difficulty (other than mild dyslexia) or psychiatric illness; acquired prosopagnosia; a history of major/moderate brain injury at any time or a mild head injury or concussion during the preceding 12 months. Participants were

¹¹ Also known as available-case analysis. To aid statistical power, group analyses were run using all available data, even if a participants had missing data on another task.

recruited through media coverage, social media, personal networks, and prosopagnosia support groups and offered a £10 gift voucher to recompense them for their time.

Table 5.2: DP Z scores on tests used for classification

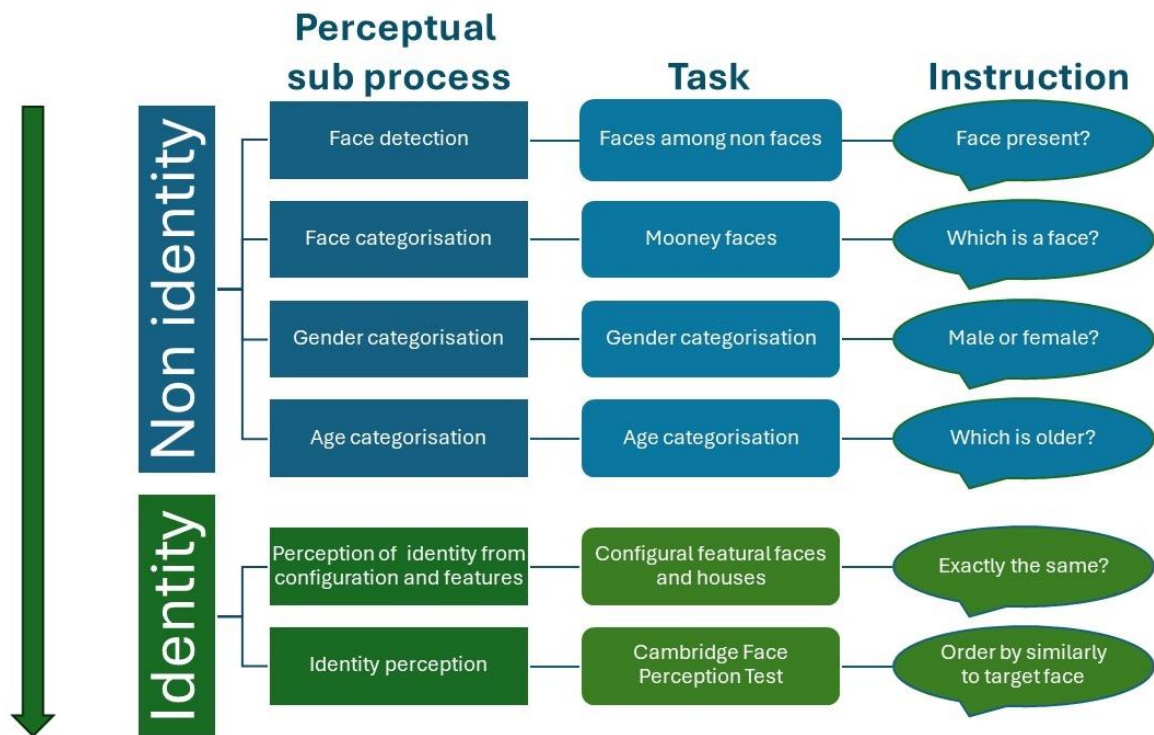
ID	Age	PI-20	Global face memory (BIS)	Old New Faces (BIS)	FFT	CFMT (BIS)	CBMT (BIS)	CMT difference (BIS)	No. of tests showing impairment (out of 4)
AF019	68	-7.75	-3.90	-6.92	-5.22	-1.72	0.02	-1.74	4
AF002	25	-7.28	-2.24	-4.35	0.41	-2.70	-0.37	-2.33	3
AF016	23	-7.01	-2.03	-2.11	-2.63	-1.52	0.35	-1.87	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. Data from four additional participants with suspected DP are not included here because they were classified as subjective DPs in Chapter 4 i.e. their global face memory BIS score was within 1 SD of control means.

5.3.2 Overview of tasks

For ease of reading, the tasks used to test each research question hypothesis are described in section four prior to the results of that question. Figure 5.1 gives an overview of the tasks, the perceptual sub process that each investigates, and the decision participants must make.

Figure 5.1 Overview of tasks



5.3.3 Procedure

Data were collected online using Testable (www.testable.org, Rezlescu et al., 2020). Participants (or parents, in the case of under 18s) received written task instructions containing links to each test and were instructed to complete the tests in a fixed order¹². Full onscreen instructions were also provided. Parents supervised under 18s and could assist by explaining tasks, but instructions stressed they must not help children with their responses. 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 also stop at the end of any test to take additional breaks.

¹² Faces among non faces, Cambridge Face Perception Test, age categorisation, gender categorisation, configural and featural faces and houses, and finally Mooney faces.

5.3.4 Statistical analyses

5.3.4.1 Group analyses

To calculate group differences, I performed Bayesian independent samples *t* tests on the standardised scores using the default Cauchy prior with a scale of 0.707. I pre-registered Bayesian analysis in addition to frequentist analyses because Bayesian analysis enables the strength of evidence for both the alternative and null hypotheses to be compared and removes the need to correct for multiple comparisons (Gelman et al., 2012; Kruschke, 2010). I report 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 non-parametric alternatives were required, I performed Wilcoxon Rank Sum tests and Bayesian Mann Whitney U tests with 5 chains of 1000 iterations. Test reliabilities are reported in Table 5.3 below.

5.3.4.2 Individual cases.

At the individual participant level, I performed single case analyses using the SingleBayes_ES.exe computer program for Bayesian tests of deficit (BTD), or the DiffBayes_ES.exe program for Bayesian standardised difference test (BSDT) (Crawford et al., 2010; Crawford & Garthwaite, 2007). The alpha level was set at 0.05. The DP sample size was determined following previous literature and my control sample size (target of 20 participants per age group) was informed by McIntosh & Rittmo's (2021) study. This recommends a minimum neuropsychological control sample of at least eight participants and that twice that number is more desirable, but that increasing control sample size above 16 does little to meaningfully increase power.

I previously reported a novel and important early finding from the screening and classification analysis, namely that a clear speed-accuracy strategy difference was observed between DPs and controls on both non-face (Lowes et al., 2024b) and face tasks (Lowes et al., 2024a). To account for observed differences in speed-accuracy strategies between DPs and typical controls, I therefore additionally calculated and analysed balanced integration scores (BIS) (Liesefeld & Janczyk, 2019, 2022) in the following tasks: Face among non faces; gender categorisation; age categorisation; configural and featural house and face tasks. The BIS was not a preregistered measure but helpfully adjusts accuracy (a preregistered measure) for

response time (RT), another preregistered measure, into a single integrated score thereby controlling for differential speed-accuracy trade-offs both within and between participants. For RT all trials < 250 ms were removed from analysis. BIS is calculated as $Z_{\text{accuracy}} - Z_{\text{RT}}$ and the formal calculation for BIS is shown in supplementary materials.

For completeness and transparency I report both the original pre-registered primary outcome measures (either accuracy or RT alone) and BIS for most tasks. I did not consider BIS an appropriate measure for two face perception tasks; first the Mooney face task because the task was to click very precisely on the eyes of one of three images that contained a face and I reasoned that motor control could disproportionately influence RTs. Second the Cambridge Face Perception Test (CFPT) since task instructions emphasise accuracy over speed, and again motor control and familiarity with a mouse/touchpad strongly influence RTs because participants must respond by clicking and dragging six face images to place them in the correct order of similarity to the target image. Results on these two tasks are therefore reported using only the original preregistered outcome measures, namely accuracy and inversion effect (CFPT) and accuracy (Mooney faces).

I was interested in whether the DPs who showed the largest impairment on various face perception measures would be the same individuals who showed the largest impairment on the four face memory measures I used to classify the DP group, in other words was the lowest performing DP on one task also the lowest scoring on others? To answer this, I estimated Intraclass Correlation Coefficients (ICC) and their 95% confidence intervals using the irr package (Gamer et al., 2019), based on mean rating using consistency and a two-way mixed effects model (model type C1). I calculated ICCs separately for the DP and control groups using single measures (i.e. each individual test z score) and then compared the correlations by checking for overlapping confidence intervals. Overlapping confidence intervals suggest that there is no meaningful difference between groups. I use Koo & Li's guidelines (2016) 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.

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 I 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) I applied the Benjamini Hochberg (BH) correction (Benjamini & Hochberg, 1995) to my 13 primary outcome measures (Faces among non faces RT, Faces among non faces RT normalized inversion index, Mooney face accuracy, age categorisation BIS, gender categorisation BIS, CFPT accuracy, CFPT normalised inversion index, configural face BIS, configural featural house BIS, configural featural face advantage score, configural featural spacing advantage scores, 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; Selker et al., 2022; Wickham et al., 2019). Cronbach's alpha was calculated in SPSS 28.0.0.0 (IBM) and Bayesian analyses in JASP (JASP Team, 2023).

5.4 Research questions and results

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

Table 5.3 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 categorisation	.40	.20	.58	.85	.85	.85
Age categorisation	.63	.12	.86	.90	.90	.86
Configural featural	.85	.86	.83			
CFPT	.72	.69	.68			

5.4.1 Research question 1 Are all, or some, DPs impaired at face detection?

Hypotheses	Stage of perceptual processing	Test	Dependent variable (s)
1a: If DPs are impaired at detecting faces, then their RTs on face-present upright trials will be longer than those of controls in a Faces among non-faces task.	Face detection	Faces among non faces	RT BIS
1b: If DPs are impaired at holistic processing, then they will show a reduced inversion effect on 'face present' trials compared with controls in a Faces among non faces task.	Face detection	Faces among non faces	Inversion effect (RT on correct upright face-present trials minus RT correct inverted face-present trials).
1c: DPs will be less accurate than controls in a Mooney face task	Face categorisation	Mooney	% correct

5.4.1.1 Motivation for the research question

This task tests face detection, the earliest stage of face processing. Face detection is thought to be relatively automatic in typical perceivers and studies with both adults and children have found that detection of a face in a scene is rapid, but that large inversion effects are observed. The FIE has been hypothesized to be due to holistic processing being disrupted by inversion or because the internal face template, to which all incoming faces are thought to be compared, is developed though

particular experience with upright faces. For efficient face recognition to take place, individuals must first be able to rapidly *detect* the presence of a face within a scene or among other objects, and also *categorise* 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. Here I test detection using the Faces among non faces task (Garrido et al., 2008) since the authors found group level RT impairments on this task in adult DPs. Further, a child study using this same task (Dalrymple & Duchaine, 2016) reported that 3/7 children with DP showed impaired face detection. Together, these findings suggest that some, but perhaps not all, DPs are impaired at face detection. 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 recognising faces.

Categorisation and detection are slightly different sub processes of face perception (Bindemann & Lewis, 2013). Neuroimaging studies have identified neural correlates of face categorisation as early as 100 *ms* after stimulus onset (Liu et al., 2002; Zhen et al., 2013). Categorising two tone Mooney faces is considered to be a top-down process that depends on being able to access stored representations of previously seen faces. Mooney faces lack obvious featural cues that DPs may rely on for face identification thus theoretically forcing participants to rely on configural or holistic processing. Mooney faces are therefore an interesting task when testing DPs and a deficit at this early stage of face processing would strongly suggest a perceptual deficit in addition to a memory deficit. Although previous work reported that Mooney faces performance was impaired in a patient with acquired prosopagnosia (Steeves et al., 2006), other studies (Towler et al., 2016; Ulrich et al., 2017) found that DPs did *not* show deficits on tasks using Mooney faces, despite Towler and colleagues (2016) reporting that Mooney and naturalistic faces recruited shared neural populations. Whether face categorisation is atypical in DP therefore remains an open question.

5.4.1.2 Task: Faces among non faces (Garrido et al., 2008)

For a full description of methods please see Garrido et al., 2008. Briefly, participants were shown a fixation cross followed by a 5 x 5 grid of images, presented for 8 s. In two thirds of the trials, a single face is present. The remaining target-absent trials contain no face. The task is to spot when a face is present and to indicate this by key press as quickly as possible. If no face is present, participants were instructed

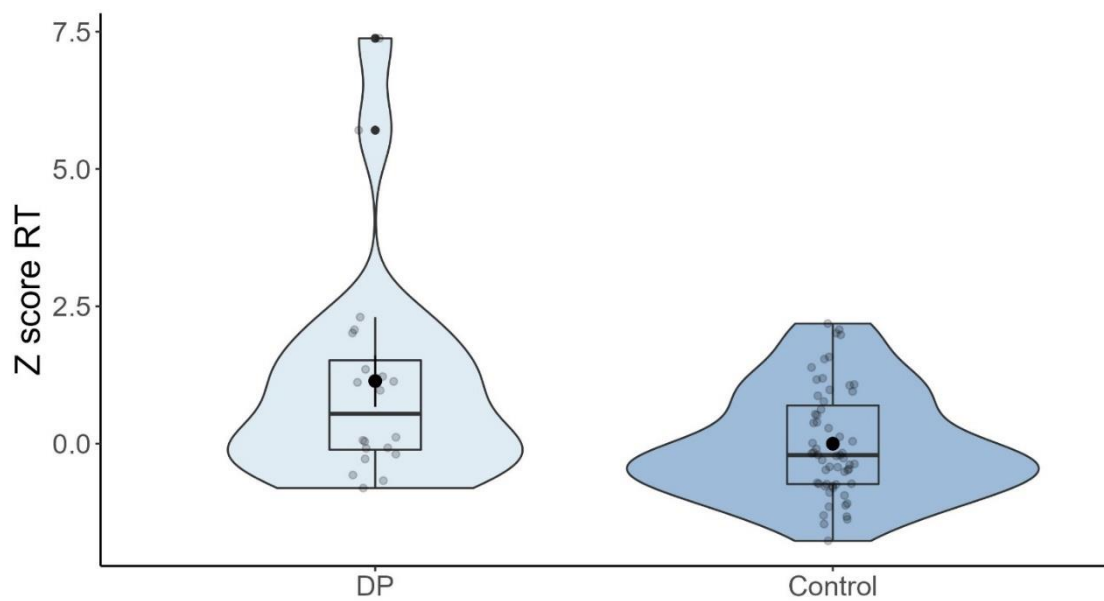
not to press any key. The test starts with four practice trials with feedback. No feedback was provided in the experiment itself. The test employs a blocked design, 37 upright face trials are followed by a second block of 37 inverted face trials. If present, the face is the only inverted image in the grid. Trial order was randomised within blocks. At the end of the first block, participants were instructed that in the following trials the face, if present, would be inverted and shown an example of an inverted face-present trial. Participants had the opportunity to take a short break between blocks. In this task, upright accuracy is typically very close to ceiling. The main outcome variable of interest is therefore RT on the correct upright face-present trials, with longer RT among DPs thought to indicate impaired face detection ability. The second outcome variable was inversion effect (calculated as RT correct upright face-present trials *minus* RT correct inverted face-present trials).

5.4.1.3 Hypothesis 1a results

DPs were predicted to take longer to detect the presence of a face versus controls. One control participant's data were highly atypical with higher accuracy and faster RT for inverted versus upright trials and because this is the opposite of the pattern robustly observed on face perception tasks among typical perceivers I removed this outlier's data from the analysis. In both the control and DP groups, the relationship between upright accuracy and RT was negative but NS ($p > .05$ in both groups).

As shown in Figure 5.2, DPs took longer on average to correctly detect the presence of an upright face among distractors than their age matched control group. Descriptive statistics are shown in Table 5.4. Because neither RT nor BIS scores were normally distributed, I conducted Wilcoxon rank sum tests, a non-parametric alternative to t tests, to assess whether the between-group difference was significant. Results indicated the DP group were significantly slower to respond correctly than the control group, $W = 826$, $p = .008$, which remained significant after Benjamini Hochberg adjustments. Bayesian Mann Whitney tests provided anecdotal evidence for a true between-group difference, $BF_{10} = 1.88$. These results provide initial support for hypothesis 1a since longer RTs likely indicate an impairment.

Figure 5.2 Group differences Faces among non faces



Note: Mean RT is calculated on upright correct trials. 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 x the IQR.

Next, to account for possible differential speed accuracy trade-offs (SATs) that might otherwise mask accuracy impairment, I also calculated BIS. A Wilcoxon rank sum test with continuity correction showed that once accuracy was adjusted for RT, the DP group scored, on average, significantly lower than the control group, $W = 361$, $p = .010$. Bayesian analysis showed anecdotal evidence for a group difference $BF_{10} = 1.94$. Taken together, the results support hypothesis 1a with the DP group showing group level impairment in face detection.

Table 5.4 Faces among non faces descriptive statistics

RT (ms)	DP				Control		
	<i>N</i>	<i>Z</i>	<i>M</i>	<i>SD</i>	<i>N</i>	<i>M</i>	<i>SD</i>
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	<.001	1.65
36 - 59 years	8		-1.12	2.59	21	<.001	1.73
60 - 74 years	5		-2.75	2.84	17	<.001	1.42
Total	20		-1.79	3.84	59	<.001	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 standardised 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, the smaller the number the greater the effect of inversion on RT.

To investigate further I inspected results at individual case level (Appendix 4 supplementary Table S4a). Data showed that 5/20 (25%) of DPs were significantly impaired on the primary measure of RT and a further 4/20 showed mild impairment of at least 1 *SD* above age matched controls. The observed group differences cannot therefore be accounted for by a small number of DPs with unusually long RT since almost half of the DP group showed impairment. When inspecting BIS, a similar proportion (6/20) of individual DPs showed major impairment ($Z \leq -1.7$). Therefore, because on this task RT and BIS produced very similar results at group and individual level due to both groups achieving accuracy close to ceiling, I proceeded to use RT, as preregistered, to assess inversion effects.

5.4.1.4 Hypothesis 1b results

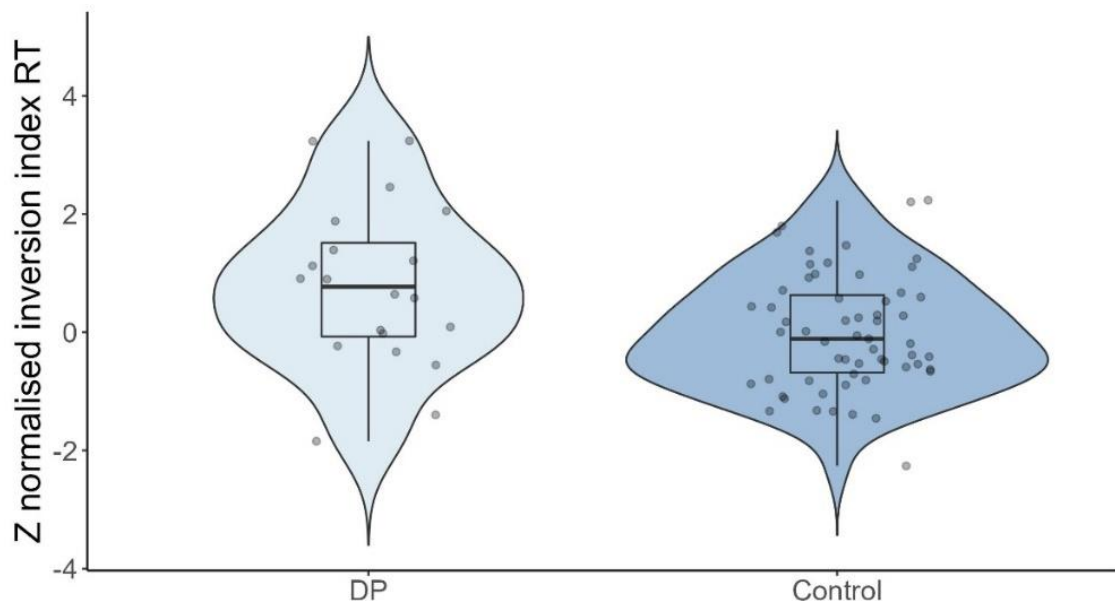
As predicted by hypothesis 1b, when the face present within the array was inverted, the DP group showed an attenuated face inversion effect ($M_{\text{diff}} = -707$ ms, SD 494) versus the control group ($M_{\text{diff}} = -888$ ms, SD 367). In other words, both groups took longer to correctly detect the presence of an inverted compared with upright face (this showing a FIE), but importantly the RT increased *less* among DPs than among typical controls. I first assessed the difference in inversion effect using the preregistered subtraction method (upright RT *minus* inverted RT) and next standardised the inversion effects centred on each control age group mean RT to account for any possible differences in the magnitude of typical face inversion effects between age groups. Data were not normally distributed, I therefore conducted a Wilcoxon rank sum test with continuity correction which indicated that the group difference in inversion effect was NS, $W = 753$, $p = .067$. Bayesian Mann-Whitney tests resampled with 5 chains of 1000 iterations, provided converging weak evidence for the null hypothesis, $BF_{10} = 0.59$. 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.

Due to the observed RT differences between groups, I also computed a *normalised* inversion RT index following previous literature (Avidan et al., 2011) calculated as follows: Inversion Index = [(upright RT – inverted RT) / (upright RT + inverted RT)]. I once again standardised this index based on control age group means. This normalised 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 one control participant with a mean upright RT of 1000 ms (representing a 50% increase in RT) versus a second DP participant with a mean upright RT of 2000 ms (representing a relatively smaller increase in RT of 25%). Inversion index data showed that the DP group's mean RT time was *relatively* less affected by inversion compared with age matched controls.

Results indicated mixed support for hypothesis 1b (Figure 5.3) . The DP group showed reduced, although not absent, holistic processing of faces compared with controls as indexed by face inversion effects. Both groups showed a similar pattern of results - taking longer to correctly detect inverted versus upright faces and thus *both groups* showed evidence of holistic face processing. The DP group's face

inversion effect was significantly *reduced* versus controls when using the normalised method but was not significantly different when using the subtraction method 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 5.3 Face inversion effect: Faces among non faces



Note: Inversion Index = (upright RT – inverted RT) / (upright RT + inverted RT). Scores > 0 indicate a relatively smaller inversion effect versus age matched controls. 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 x the IQR.

The observed group level inversion index differences do not appear to be driven by a small number of DP participants with extreme scores. At an individual level (Table 5.14), 8/20 DPs (40%) showed at least a mildly reduced inversion index ($Z > 1$); of these 4/20 (20%) showed a major reduction in face inversion index versus controls ($Z \geq 1.7$). By contrast, only 2/60 (3.3%) of control participants showed a major face inversion index impairment.

5.4.1.5 Task: Mooney Faces (Verhallen & Mollon, 2016)

I administered a shorter version of the original task described in detail in Verhallen and Mollon (2016). Pilot data collected online for another study in our lab ($n = 65$) using 144 Mooney faces was examined and I selected the ten trials with the highest item-total correlations to create a much-shortened test. I reasoned that these

trials would be the most informative for identification of a possible impairment and would allow a relatively short, but still meaningful, test to be designed. Test length considerations are important when compiling a lengthy test battery as I had here. Mean accuracy for the ten chosen items ranged from 40% to 64.5%. Item-total correlations ranged from 62.6% to 73.8%. Briefly, the test procedure was as follows. Participants were instructed to calibrate their screen, and that calibration was essential for response precision on this task. In each trial, participants saw three scrambled black and white images presented simultaneously, only one of which was a face. The task was to choose the image that showed a face and to indicate this 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 of the image containing a face and chance is 33%. The task began with three example trials, no feedback was given in the test itself.

5.4.1.6 Hypothesis 1c results

Most participants scored close to ceiling (Table 5.5) meaning accuracy data were not normally distributed (Shapiro-Wilk $p < .001$) so I performed a Wilcoxon rank sum test with continuity correction to assess group differences. Results indicated no significant group accuracy difference ($W = 753$, $p = .062$, rank biserial correlation = 0.21) and I was unable to reject the null hypothesis. In line with this finding of no significant group difference, a Bayesian Mann Whitney test provided converging anecdotal evidence for the null hypothesis ($BF_{10} = 0.62$). At individual level, a small number (3/20) of DP participants (AF010, AF018 and AF098), were significantly impaired on the Mooney task (see Appendix 4 supplementary table S4a. In each of these cases, the Bayesian point estimate of the percentage of the age matched control population with a lower accuracy score than the case was less than 1.4%. Interestingly, two of these individuals also showed impairment on the Faces among non faces detection task (longer RT versus controls, see 4.1.1 above) and together these results provide converging evidence that these participants are significantly impaired at the earliest stage of face processing, that is face detection.

Table 5.5 Descriptive statistics Mooney faces

Proportion correct	DP				Control		
	<i>N</i>	<i>Z</i>	<i>M</i>	<i>SD</i>	<i>N</i>	<i>M</i>	<i>SD</i>
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.

5.4.1.7 Discussion

The finding (supporting hypothesis 1a) that the DP group was significantly slower than controls to *detect* a face among objects replicates the original behavioural study using the same task (Garrido et al., 2008). Further, my findings that some but not all individual DPs showed impaired face detection is in line with previous work that employed the Faces among non faces task in both children and adults with DP (Dalrymple & Duchaine, 2016; Garrido et al., 2008). Consistent with these behavioural findings, a previous neuroimaging study of three family members with DP (Németh et al., 2014) concluded that DP is due to difficulties filtering faces from non-faces, since DPs showed reduced face sensitivity to the N170 that was driven not by a smaller face-elicited N170 in DP, but instead by a larger noise-elicited N170.

I included a face detection, or arguably a face categorisation, task (hypothesis 1c) using Mooney faces in the hope this would provide converging evidence for the hypothesis 1a that DPs are impaired at face detection. However, data did not support hypothesis 1c; both frequentist and Bayesian analyses supported the null hypothesis. My finding of no significant group-level difference between DPs and controls is line with data from other studies that tested DPs using Mooney faces (Towler et al., 2016; Ulrich et al., 2017), albeit these studies used different test versions and different outcome measures to ours, thus preventing direct comparison.

However, at individual case level a slightly different pattern emerged; 3/20 individual DPs in our sample showed significant impairment on the Mooney task, contrary to the findings of Ulrich et al. (2017) who reported that 0/11 DPs were impaired. One possible explanation for this discrepancy is that their DP sample was smaller ($n = 11$) than ours ($n = 20$) so may not have been large enough to capture what appears to be a relatively infrequent impairment in DP. Another difference is

that Ulrich and colleagues administered a different version of the Mooney test which comprised more trials but used a simpler 2 AFC format that did not require participants to select the eye area of the image. Their version appears to have been easier than the shorter version I created, as demonstrated by the fact that mean accuracy in both their DP and control groups was 98% (versus 82% and 80% respectively in my sample) and may have lacked the required sensitivity to detect group differences. The tasks are not therefore directly comparable. Although only a small number (3/20, 15%) of the DPs in my sample showed impairment on the Mooney face task, notably two of these three were *also impaired* on the alternative face detection task (Faces among non faces), thus providing converging evidence that these individuals have low level face perception difficulties.

5.4.2 Research question 2: Are all, or some, DPs impaired at categorising facial gender?

Hypothesis	Stage of perceptual processing	Test	Dependent variable (s)
2: DPs will be less accurate than controls at judging gender.	Nonidentity encoding of facial gender	Gender categorisation	% correct BIS

5.4.2.1 Motivation for the research question

Accurate face identify recognition presumably requires the target's gender and approximate age (see 4.3 age categorisation) be judged quite rapidly. Successful categorisation 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 et al., 2012; Duchaine et al., 2006; Le Grand et al., 2006) and this therefore remain an open question.

5.4.2.2 Task: Gender categorisation

Participants were presented with 30 face images, one at a time, and the task was to judge whether the face was male or female (Figure 5.4). Participants indicated their response by keypress (*F* for female and *M* for male). All stimuli were

naturalistic, White adult faces aged 18-54 years with no obvious facial hair or make up and were taken from the University of Michigan Face database Cal Pal ageing set (Minear & Park, 2004). Stimuli were presented in full colour on a black background. Hair was cropped but hairline was retained. The test began with three practice trials showing cartoon characters and feedback was given. Each stimulus remained on screen until a response was made. No feedback was given during the 30 test trials which were presented in randomised order, with an equal number of male and female faces.

Figure 5.4 Gender categorisation task



Note: Example stimuli from the gender categorisation task. From left to right female, female, male, male. Stimuli were presented one at a time. The outcome variables are proportion correct and balanced integration score.

5.4.2.3 Hypothesis 2 results

Likely due to mean accuracy being close to ceiling in both groups (Table 5.6), Wilcoxon rank sum tests indicated no significant group difference in standardised accuracy ($p = .696$, $BF_{10} = 0.34$) but showed that the DP group was on average significantly slower to make a correct response (mean RT 2186 ms, SD 863 ms) versus typical controls (mean RT 1669 ms, SD 527 ms), $W = 863$, $p = .004$).

I therefore repeated the group 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 (Figure 5.5). Bayesian Mann Whitney tests provided converging anecdotal support for a true group difference, $BF_{10} = 1.49$.

At individual level, single case analyses (Appendix 3, Table S4a) revealed that three DPs showed significantly impaired gender categorisation accuracy versus an age-matched control population. Two additional DP participants showed a mild impairment ($z < 1$).

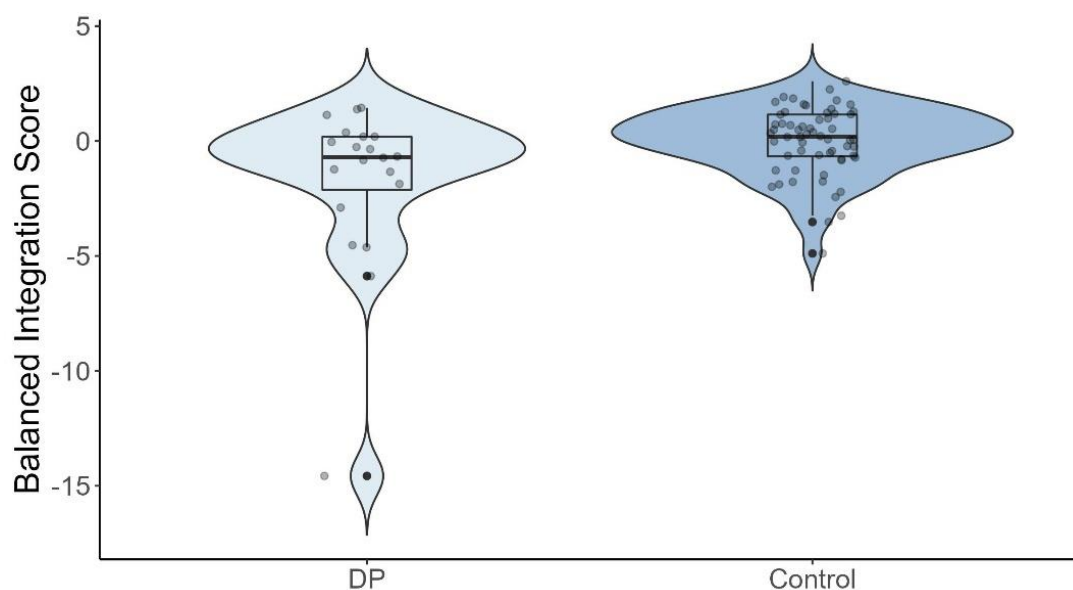
Table 5.6 Descriptive statistics gender categorisation task

Proportion correct	DP				Control		
	<i>N</i>	<i>Z</i>	<i>M</i>	<i>SD</i>	<i>N</i>	<i>M</i>	<i>SD</i>
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				<i>Z</i>		
14 - 35 years	7	-2.80	5.38	21	<.001	1.47
36 - 59 years	8	-0.53	1.88	21	<.001	1.77
60 - 74 years	5	-2.26	2.75	19	<.001	1.63
Total	20	-1.76	3.63	60	<.001	1.48

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

Figure 5.5 Comparing group performance on the gender categorisation task



Note: Each dot represents a single data point, the box shows the interquartile range (IQR) and the midline indicates the group median score. The end of each whisker line represents 1.5 x the IQR. In both groups (DP and control) mean accuracy was close to ceiling so BIS was calculated ($Z_{\text{accuracy}} - Z_{\text{RT}}$) to account for DPs' slower RTs and to control for differential speed accuracy trade-offs.

5.4.2.4 Discussion

Results provided support for hypothesis 2. The DP group was significantly slower than age matched controls. An integrated measure (BIS) showed DPs were significantly more impaired than controls once response time was accounted for. This finding is in line with Esins et al. (2016) who reported group level ($n = 16$) gender categorization impairment in DP and another recently study (Marsh et al., 2019).

Previous work has reported no difference in ability between DP and control groups (e.g. Chatterjee & Nakayama, 2012; DeGutis et al., 2012). However, major differences in stimuli type mean my findings are difficult to compare directly to these studies which used artificially generated and/or morphed facial stimuli comprising part male and part female identities and therefore lacked a ground truth. Previous work has used a sorting paradigm (Chatterjee & Nakayama, 2012) whereas I used a binary male/female judgement. My task used ecologically valid, naturalistic images of human faces, from which cues such as make up, facial hair, jewellery and hair (though not hairline) were absent, which arguably should have been easier for DPs to categorise than morphed androgynous faces, thus making differences between DPs and controls more difficult to detect. It is possible that DPs may categorise artificial, or morphed, faces differently to faces they encounter in everyday life, despite the advantages the former offer in tight control of stimulus properties. However, considering the two studies that found group differences, one used natural faces (Esins et al., 2016) and the other morphed (Marsh et al., 2019) so the stimulus difference is evidently not decisive.

At individual case level, my finding that 6/20 (30%) DPs showed a major gender categorisation impairment is strikingly similar to the proportion reported elsewhere in the literature. For example, despite finding an absence of group-level difference, Chatterjee and Nakayama (2012) nevertheless report that 5/19 (26.3%) of individuals with DP showed impairment of at least $z = -1.5$ on a gender categorisation task. Similarly, Marsh et al. (2019) supplementary data shows 5/16

(31%) of individual DPs were impaired and Esins et al., (2016) found over half (9/16 or 56.3%) were impaired ($z < -1.8$). Taken together, these individual level results suggest that one quarter to one half of individual DPs have considerable difficulty categorising gender. This finding appears robust because a similar proportion of DPs were impaired across several studies emanating from different research groups and using different paradigms. I suggest that it is possible that the conflicting group-level findings in the literature to date may be explained not just by methodological differences, but also by the relatively modest DP sample sizes which may have differed in their balances of DP participants who were impaired, and relatively unimpaired, at gender categorisation.

5.4.3 Research question 3: Are some or all DPs impaired at judging facial age?

Hypothesis	Stage of perceptual processing	Test	Dependent variable (s)
3: DPs will be less accurate than controls at judging facial age.	Nonidentity encoding of facial age	Age categorisation	% correct BIS

5.4.3.1 Motivation

As with the gender categorisation task, I included this facial age categorisation 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. I reasoned that including a second perceptual categorisation 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 (e.g. Chatterjee & Nakayama, 2012; Dobel et al., 2007) and thus whether DPs can make accurate age categorisation decisions when presented with naturalistic faces is another open question.

5.4.3.2 Task: Age categorisation

Participants had to judge which of two simultaneously presented faces was older and indicate their choice by clicking on the older face (see Figure 5.6). Each pair of faces was White, the same gender, presented in full colour with neutral

expression and hair cropped but hairline remaining. Stimuli were taken from the University of Michigan Face database Cal Pal ageing set (Minear & Park, 2004). Three age conditions were created: A small age gap of 10-15 years; a medium age gap of 16-30 years; and a large age gap of > 30 years. The test began with two simple practice trials showing a baby and an elderly person and feedback was given. Stimuli remained on screen until a response was made. No feedback was given during the 30 test trials which were presented in randomised order with an equal number of male and female pairs. The position of the older face (left or right) was counterbalanced.

Figure 5.6 Age categorisation task



Note: Example stimuli from the age categorisation test. Left: Male, medium age difference trial (older on left). Right: Female, large age difference trial (older on right). The outcome variables are accuracy and BIS.

5.4.3.3 Hypothesis 3 results

As with the gender categorisation task, most participants achieved accuracy scores at, or close to ceiling, meaning data were not normally distributed (Table 5.37). A Wilcoxon rank sum test with continuity correction indicated no significant group difference in standardised 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 anecdotal support for the null hypothesis for both accuracy ($BF_{10} = 0.33$) and BIS ($BF_{10} = 0.58$).

At individual level, single case analysis showed that one DP was significantly less accurate versus age matched controls (Supplementary table S4a). However, when RT and differential speed-accuracy trade-offs were controlled for using BIS, results were that 3/20 DPs showed a major impairment ($z \leq -1.7$) and 4/20 showed a mild impairment versus age-matched controls.

Table 5.7 Descriptive statistics age categorisation task

Proportion correct	DP				Control		
	<i>N</i>	<i>Z</i>	<i>M</i>	<i>SD</i>	<i>N</i>	<i>M</i>	<i>SD</i>
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	< .001	2.05
36 - 59 years	8	-0.14	-0.49	1.32	21	< .001	1.80
60 - 74 years	5	0.07	0.34	0.62	18	< .001	1.45
Total	20	-1.61	-1.61	5.37	60	< .001	1.29

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

5.4.3.4 Discussion

Contrary to the prediction of hypothesis 3, I 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 that found ceiling effects (Dobel et al., 2007; Nunn et al., 2001), the absence of significant group difference is likely due to ceiling effects as the test was designed to be suitable for children. Nevertheless, reduced BIS performance was observed in the DP group ($z = -1.61$). To my knowledge, the literature shows no evidence for DP group level impairments for facial age judgement and this task was included primarily to see if it might provide converging evidence for early-stage face perceptions deficits at the individual case level. Although fewer individual DPs showed deficits on the age categorisation task (1/20 for accuracy and 3/20 for BIS) than for gender categorisation, an important finding is that two of the three participants who were severely impaired when categorising age *also* showed a significant impairment when judging facial gender (Table 5.14). This suggests these individuals have more generalised difficulty encoding early-stage nonidentity facial information.

5.4.4 Research question 5: Do participants with DP show deficits in configural or featural face processing? Are any deficits observed domain specific or process specific?

Hypotheses	Stage of perceptual processing	Test	Dependent variable (s)
5a: If DP is <i>domain specific</i> then DPs will be <i>more</i> impaired at detecting changes to upright <i>faces</i> than to upright houses compared with controls.	Nonidentity encoding of configuration and features	Configural featural	% correct BIS Face advantage score [(upright face correct – upright house correct)/(upright face correct + upright house correct)]
5b: If DP is <i>process specific</i> and driven by a deficit in configural processing, then DPs will be <i>more</i> impaired at detecting changes to <i>spacing</i> than to features compared with controls across both faces and houses.	Nonidentity encoding of configuration and features	Configural featural	% correct BIS Spacing advantage score [(upright spacing accuracy – upright feature accuracy)/(upright spacing accuracy + upright feature accuracy)].
5c: If DP is driven by a deficit in <i>domain-specific holistic processing</i> , then DPs will show a <i>reduced</i> inversion effect for faces, but not houses, compared with controls.	Nonidentity encoding of configuration and features	Configural featural	% correct Face inversion House inversion

5.4.4.1 Motivation for the research question

I used this task to test three competing accounts of the face perception impairments observed in DP (hypotheses 5a to 5c above, for a full discussion 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 and 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 for spacing changes to houses are *smaller* than inversion effects for face spacing changes (or are absent), this would suggest that face processing mechanisms are domain-specific (i.e. specialised for faces) as opposed to process specific (i.e. specialised for configural processing more generally) thus supporting the face-specific holistic hypothesis (Yovel & Duchaine, 2006).

A review (McKone & Yovel, 2009) of conflicting findings from different configural featural face tasks found that performance varied as a function of stimulus properties such as colour and contrast, and as a function of task properties such as set size. Therefore, to allow comparison with previous DP research, I chose to use a task and stimulus set developed and tested by Yovel and Duchaine (2006). Yovel & Duchaine (2006) argue their configural featural task, using the Ann faces set, can provide evidence for the *face-specific holistic hypothesis* as it uses a matched houses test. Correct detection of spacing changes tests participants' ability to process second order spatial relations (configural processing). If inversion effects for spacing changes to houses are smaller than inversion effects for face spacing changes (or are absent), this would suggest that the processing mechanisms involved are specific for faces as opposed to process specific for configural processing, thus supporting the face-specific holistic hypothesis.

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., 2014; DeGutis et al., 2007; Schmalzl et al., 2008) understanding whether either, or both, abilities are impaired in DP in general – or in individual DPs - can inform more personalised approaches to rehabilitation.

5.4.4.2 Task: Configural featural (Yovel & Duchaine, 2006):

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 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 (e.g., between the eyes or windows) or to *features* themselves (e.g., mouth

or door). Stimulus pairs were presented consecutively, each for 500 ms, with an inter-stimulus interval of 1000 ms. The task used a blocked design (upright faces; feature change x 20 trials, spacing change x 20 trials; inverted faces; feature change x 20 trials, spacing change x 20 trials). Next, this blocked order was repeated showing house stimuli making a total of 160 trials (80 upright, 80 inverted). Trial presentation was randomised within the eight blocks and participants could take an optional break between blocks.

Feedback during piloting indicated that some child control participants (whose results are not reported here due to no children being classified as DP) anticipated the differences between the faces to be more obvious than they were, so I amended the instructions (see supplementary information) slightly from the original to make them simpler to understand and to show participants two examples of the types of changes they might see. Example trials showed a different facial identity to the one used at test. The test began with two practice trials using cartoon faces to familiarise participants with the task. Prior to analysis, I inspected RT and accuracy data to check for suboptimal effort, which I defined as performance at, or below, chance *and* $RT < 250$ ms on more than 10% of trials. No participant was removed on this basis and only a tiny minority of trials (18/12800) showed $RT < 250$ ms.

5.4.4.3 Hypothesis 5a results (testing the domain-specific hypothesis)

The face advantage score measures the extent to which participants are *relatively* more accurate at detecting changes to houses versus faces (within - participant measure) compared to typical controls of similar age (between participant measure). Negative face advantage scores indicate an advantage for houses. Face advantage scores were normally distributed (Shapiro wilk $p = .964$) and descriptive statistics for the component and derived scores are provided in Table 5.8.

Table 5.8 Descriptive accuracy statistics configural featural task

	DP				Control		
	<i>N</i>	<i>Z</i>	<i>M</i>	<i>SD</i>	<i>N</i>	<i>M</i>	<i>SD</i>
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 normalised difference score calculated as $[(\text{faces} - \text{houses})/(\text{faces} + \text{houses})]$. A face advantage score > 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 z scores < 0 indicate that the magnitude of the relative impairment for faces is greater than that observed in typical controls.

As Table 5.9 below shows, face advantage results provide support for hypothesis 5a and therefore for a *domain-specific* account of DP. As predicted by this account, at group level, DPs were significantly impaired in the faces condition but not in the house condition. Further, at individual level as shown by the face advantage score, the DP group showed significantly reduced ability to discriminate faces compared to houses, relative to age matched controls (including after B-H adjustments were applied for multiple comparison). Bayesian analysis provided converging anecdotal support for true group differences in face discrimination and additionally 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 Appendix 3 Table S4b.

Table 5.9 Configural featural task: Comparing DP and control group performance

	DP	Control			
	<i>M (SD)</i>	<i>M (SD)</i>	BF ₁₀	<i>p</i>	Effect size
Face advantage	-0.54 (0.60)	< .001 (0.98)	2.49	.005**	<i>d</i> = 0.67
Face accuracy	-0.56 (0.87)	< .001 (0.98)	2.21	.021*	<i>d</i> = 0.60
House accuracy	0.11 (1.41)	< .001 (0.98)	0.29	.331	<i>rbb</i> = 0.11
Exploratory analyses					
Face BIS	-1.10 (1.39)	< .001 (1.52)	7.62	.005**	<i>d</i> = 0.76
House BIS	-0.21 (1.95)	< .001 (1.57)	0.25	.890	<i>rbb</i> = 0.02

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

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 ($Z_{\text{accuracy}} - Z_{\text{RT}}$) using upright trials only.

Group differences in standardised 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.

5.4.4.4 Hypothesis 5b results (testing the process specific account)

The *process-specific* account of DP posits that DP can be accounted for by a wider deficit in the ability to discriminate *configural changes*, that is changes to the spacing between parts or features. The predictions that follow from this hypothesis are first that DPs should first show relatively *more* impairment when detecting spacing versus featural changes, and second, that this greater *relative* difference would be observed for both face *and* houses

I first calculated accuracy in the *spacing* (face spacing and house spacing conditions collapsed) and *feature* conditions (again for both faces and houses together). I next computed a normalised *spacing advantage* score [(upright spacing accuracy - upright feature accuracy)/(upright spacing accuracy + upright feature accuracy)] (

Table 5.10). Finally, I converted these scores to z scores centred, as always, on age group control means. Because data were not normally distributed (Shapiro Wilk $p < .05$), I conducted Wilcoxon rank sum tests with continuity correction and their Bayesian equivalent to assess group differences.

Results did not show support for hypothesis 5b (Figure 6.7). No significant difference in spacing advantage was observed between DPs and controls, $W = 513.5$ $p = .337$, rank biserial correlation (r_{bb}) = 0.11. A Bayesian Mann Whitney test showed moderate converging support for the null hypothesis ($BF_{10} = 0.29$). At individual case level, I again found no support for hypothesis 5b. Findings were that none of the 20 DPs showed only a configural impairment, the pattern predicted by the process-specific hypothesis. A minority of DPs (3/20) showed significant featural impairment, of whom one (CF009) also showed a significant configural impairment. Dissociation analyses revealed only one participant with any significant dissociation (AF019; $z = 1.03$ spacing, $z = -3.16$ features $Z\text{-DCC} = 4.16$, Appendix 3 Table S4b).

Table 5.10 Descriptive statistics: Configural and featural task spacing and features accuracy

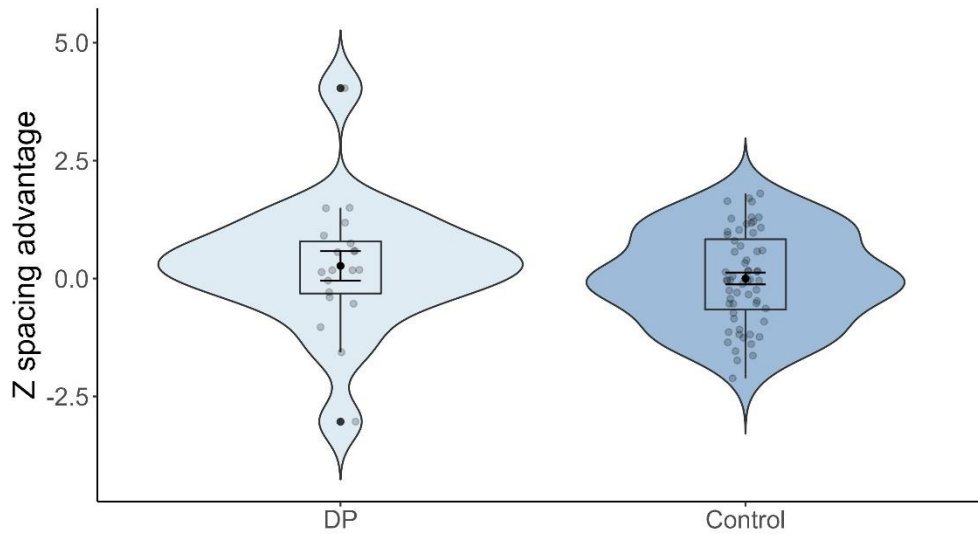
Spacing	DP			Control		
	<i>N</i>	<i>M</i>	<i>SD</i>	<i>N</i>	<i>M</i>	<i>SD</i>
14 - 35 years	7	.73	.12	21	.78	.10
36 - 59 years	8	.73	.08	21	.74	.10
60 - 74 years	5	.70	.12	18	.66	.13
Total	20	.72	.13	60	.73	.12

Features						
14 - 35 years	7	.72	.09	21	.78	.09
36 - 59 years	8	.77	.10	21	.78	.07
60 - 74 years	5	.61	.13	18	.70	.10
Total	20	.71	.12	60	.75	.09

Spacing advantage						
14 - 35 years	7	.00	.03	21	.00	.07
36 - 59 years	8	-.02	.05	21	-.03	.07
60 - 74 years	5	.07	.16	18	-.03	.09
Total	20	.01	.11	60	-.02	.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 normalised 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.

Figure 5.7 Comparing group performance spacing versus features



Note: The spacing advantage Z score is normalised ($[(\text{upright spacing accuracy} - \text{upright feature accuracy}) / (\text{upright spacing accuracy} + \text{upright feature accuracy})]$) and thus measures whether an individual participant performs best in the features or spacing conditions. Here $Z = 0$ indicates that relative standardised difference in a participant's performance between conditions is the same as the mean difference score among age-matched controls; z scores > 0 indicate relatively better performance in the spacing versus the feature condition in comparison with a typical control and scores < 0 indicates a relatively better performance in the feature condition. The prediction of hypothesis 5b is that the DP group score would be shifted below zero, which it was not. Each dot represents a single data point, the box shows the interquartile range (IQR) and the midline indicates the group median score. The end of each whisker line represents $1.5 \times$ the IQR. Error bars show 1 SEM.

5.4.4.5 Hypothesis 5c results (testing the domain-specific holistic processing account).

Data did *not* show support for hypothesis 5c, the domain-specific holistic processing account, since groups showed similar standardised inversion effects for both faces *and* houses. The DP and control groups were both less accurate on inverted compared with upright face trials and an independent samples t test showed that the between-group difference was NS, $t(33.3) = 1.18$, $p = .245$ Cohen's $d = 0.30$. Bayesian independent t tests provided converging anecdotal evidence for the null hypothesis, $BF_{10} = 0.47$.

As predicted by this hypothesis, 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 NS, $t(31.4) = 0.19$, $p = .847$ Cohen's $d = .05$. Bayesian independent samples t tests showed converging moderate

support for the null hypothesis, $BF_{10} = 0.27$. Taken together, results indicate that at a group level DPs showed broadly typical inversion effects for faces and houses. Face and house inversion data were normally distributed (Shapiro Wilk $ps > .93$). Dissociation analyses (Crawford & Garthwaite, 2005; Crawford et al., 2010) indicated that face and house inversion effects did not dissociate in the DP group.

5.4.4.6 Discussion

This task was used to test three competing hypotheses that follow from different accounts of DP. I found only support for the *domain-specific* account of DP. The overall group accuracy scores in my sample were very similar to those reported by Yovel and Duchaine (2006) for both faces and houses and consistent with their findings, I 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). My results support the view that DP does not result from overall poor configural processing but instead is face selective.

Inversion effects among DPs using this stimulus set were not reported by Yovel & Duchaine (2006) but my findings that DPs do not show significantly reduced face inversion effects (FIE) compared with typical controls are consistent with previous work (Dalrymple, Garrido, et al., 2014; Klargaard et al., 2018) yet inconsistent with other studies (Avidan et al., 2011; Russell et al., 2009; Shah, Gaule, Gaigg, et al., 2015) which found reduced or absent FIE in DPs. However, although the face inversion effect (Yin, 1969) is robust and has been replicated many times, it has recently been argued that face inversion effects in typical perceivers may be task specific (Gerlach & Mogensen, 2024) and this may account for the inconsistent results across the literature.

5.4.5 Research question 6a: Is face perception ability related to face identity recognition?

Hypothesis	Stage of perceptual processing	Test	Dependent variable (s)
6a: Face perception ability is related to face memory ability	All	CFMT	BIS
		CBMT	CFMT-CBMT BIS
		Old New Faces	BIS
		FFT	% correct (Z)
		Mooney	% correct (Z)
		Faces among non faces	RT (Z)
		Configural featural	% correct (Z)
		CFPT	Upright accuracy

5.4.5.1 Motivation for the research question

Overall, I was 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? I asked whether performance on the five previously-described measures of upright face perception [Mooney faces, Faces among non-faces, age categorisation, gender categorisation and configural featural face discrimination], plus one additional, widely-used measure of face identity perception, the Cambridge Face Perception Test (CFPT, Duchaine et al., 2007) see 4.5.2) might be related both to each other and to face identity recognition ability as measured by intra-class coefficients (ICC).

I included upright accuracy on the CFPT as one of the ICC input measures because it a very commonly used measure of face perception in DP, despite its relatively low internal consistency which it makes less than ideal as a sole measure of face perception. Several studies of DP (e.g. Bate et al., 2019; Dalrymple et al., 2014; Nørkær et al., 2023) used the CFPT as the only 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 generalisations 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 generalise performance on a single face perception task to infer ability across other face perception tasks.

5.4.5.2 Additional task: The Cambridge Face Perception Test (CFPT)

Full methods are described in the original paper (Duchaine et al., 2007). Briefly, participants have 60 seconds to arrange six grayscale, morphed faces in correct order of similarity to a target face at the top of the screen with the most similar face at the left and the most dissimilar at the right. Participants are instructed to *“Please try to get the faces in the PERFECT order, do not stop when you think they are only good enough! Use all the time if necessary. If you finish before the time runs out, click the “I’m done” button. A timer will show you how much time you have left. If the timer reaches zero, the current order of the faces freezes and you will need to move on to the next trial.”* The test begins with two practice trials to familiarise participants with the task, but no feedback is given. The test comprises a total of 16 trials, 8 upright and 8 inverted. 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, I converted the error score into an upright accuracy score following previous literature (Dalrymple, Garrido, et al., 2014; Rezlescu et al., 2012) i.e. $[(\text{maximum errors} - \text{participant error}) / \text{maximum errors}]$. The secondary outcome measure of interest is the inversion score $[(\text{upright errors} - \text{inverted errors}) / (\text{upright} + \text{inverted errors})]$ 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. 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 comparison.

Table 5.11 Descriptive statistics Cambridge Face Perception Test

	DP				Control			
Upright								
accuracy	<i>N</i>	<i>Z</i>	<i>M</i>	<i>SD</i>	<i>N</i>	<i>Z</i>	<i>M</i>	<i>SD</i>
14 - 35 years	7	-1.38	0.63	0.07	20	< .001	0.74	0.08
36 - 59 years	8	-1.65	0.61	0.11	21	< .001	0.75	0.09
60 - 74 years	5	-0.25	0.64	0.08	17	< .001	0.68	0.11
Total	20	-1.20	0.63	0.08	58	< .001	0.73	0.10
Inversion index								
14 - 35 years	7	0.81	-0.17	0.07	20	< .001	-0.30	0.17
36 - 59 years	8	1.51	-0.08	0.12	21	< .001	-0.30	0.15
60 - 74 years	5	-0.45	-0.23	0.10	17	< .001	-0.17	0.15
Total	20	0.77	-0.15	0.11	58	< .001	-0.26	0.16

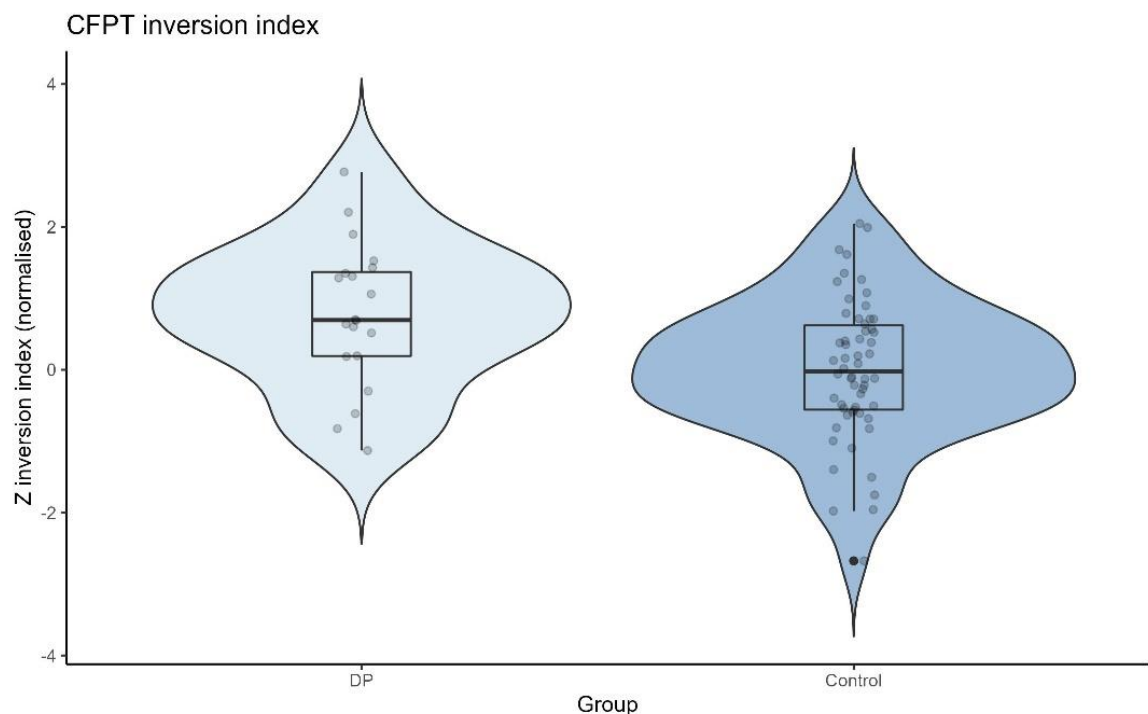
Note: Upright accuracy = [(maximum errors - participant error)/maximum errors]; inversion index = [(upright errors - inverted errors) / (upright + inverted errors)]. *Z* scores are centered on control age group means to allow comparison with typical controls of similar age. A positive *Z* score for the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

Due to weak Bayesian evidence for the null hypothesis (no group difference in the magnitude of the FIE in the faces among non faces task), I decided, post-hoc to explore an additional pre-registered inversion measure, the Cambridge Face Perception Test (CFPT) inversion index. I was interested to see whether this would provide converging evidence of absence of reduced face inversion effects among DPs. Following Avidan et al. (2011) and others, I calculated an inversion index¹³ for CFPT computed as [(upright errors – inverted errors)/(upright + inverted errors)]. An inversion index < 0 would show that a participant made proportionately more errors in the inverted condition relative to their performance in the upright condition i.e.

¹³ The preregistered inversion score did not specify *normalised* scores but, as outlined in section 4.1.2, normalised scores are a more appropriate measure of inversion effects when comparing groups with mean differences in upright (i.e. baseline) scores as was the case for the CFPT scores in our sample.

showed a typical face inversion effect with smaller numbers indicating a greater relative effect of inversion. I next converted the inversion index to a Z score centered on age group control means. Using this measure results were that DP group performance was relatively less impacted by inversion ($M = -0.15$, $SD 0.11$) versus controls ($M = -0.26$, $SD = 0.16$), descriptive statistics are shown in Table 5.11. Wilcoxon rank sum tests conducted on the standardised inversion indices showed this group difference was significant ($W = 817$, $p = .007$), including when Benjamini Hochberg adjustments were applied to control for multiple comparison. Bayesian analysis indicated converging anecdotal evidence for a true group difference $BF_{10} = 1.71$ (Table 5.13).

Figure 5.8 Comparing CFPT inversion effects DP versus control groups



Note: Inversion index = [(upright errors – inverted errors)/(upright + inverted errors)]. Z scores > 0 indicate that a participant showed a reduced face inversion effect versus age-matched controls. Each dot represents a single data point, the box shows the interquartile range (IQR) and the midline indicates the group median score. The end of each whisker line represents 1.5 x the IQR.

The fact that I observed strong, significantly reduced, face inversion effects for the DP group on the CFPT but not on the configural featural task is rather difficult to interpret. This perhaps suggests that intact holistic processing is more critical for

face individuation (the ability to identify a face as a specific exemplar, as tested by the CFPT) than for face discrimination (the ability to tell faces apart from one another, as tested by the configural featural task). Using different tasks, Stantić et al (2022) found face *perception* (how similar are these faces?) and face *matching* (is this the same identity, yes or no?) to be independently impaired in DP. It is not possible to repeat their analysis using this test battery but one explanation for this apparent discrepancy could be that, using their terminology, the configural featural task is more of a matching task (is this the same identity, yes or no?) whereas the CFPT taps perception (the instruction is to arrange faces by order of similarity to a target face) and so, if as Stantić and colleagues propose, these abilities are distinct in DP, it follows that a performance on these tasks are not necessarily related.

5.4.5.3 Hypothesis 6a results

To test consistency of performance across multiple tasks, I calculated the intraclass correlation coefficient (ICC, see Statistical analysis section 5.3.4). I input the individual z scores from each upright face perception task (Mooney face accuracy, Faces among non faces BIS, gender categorisation BIS, age categorisation BIS, Configural featural faces BIS, CFPT 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, see Lowes et al., 2024a). The ICC results using the original accuracy (or RT) measures are shown in Appendix 3 Table S3.

ICC estimates (Table 5.12) did not support hypothesis 6a. Using Koo and Li's cut-offs (2016) 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. To assess whether consistency was similar between the DP and control groups I checked whether the 95% confidence intervals overlapped between groups since for ICC the 95% confidence intervals are considered a better measure than the point estimates. Table 5.12 shows that 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. In other words, similar inconsistencies in rank performance were observed for both the DP and control groups and the inconsistencies were not specific to the group with poorer face processing ability

(DP). Instead, heterogeneity of performance was observed among DPs and also typical controls.

Because the ICC estimates across the 10 variables input into the preregistered, ICC analyses were unreliable, I conducted additional exploratory (unregistered) 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. I also calculated the ICC of the preregistered *global* (averaged) face memory and face perception scores because taking the average performance across multiple tasks can provide a more robust, less noisy measure of performance. In all cases I computed ICCs separately for the DP and control groups since I was interested to see whether the consistency of ranking would be similar between groups. The exploratory results showed that the correlations, once 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.

5.4.5.4 Discussion of hypothesis 6a

The lack of consistency among DPs, in other words the fact one DP might be ranked bottom on one test, but not on the others, supports the view that DP is a highly heterogeneous condition. This 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) but, to my knowledge, this 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 perception) is not related to similar ranking on global memory. The inconsistency of DP performance across face

perception and face memory task can also be seen visually to some extent in the table showing individual DP z scores (Table 5.14).

Like DPs, the typical control group 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 correlation estimates among typical controls, showed good to moderate reliability. These findings suggests that in typical perceivers, there is a relationship between overall face perception and face memory ability that is not observed in DPs and supports the idea of apperceptive and mnemonic sub types of DP (Biotti et al., 2019; Stollhoff et al., 2011; Ulrich et al., 2017).

Table 5.12 Intraclass correlation estimates

		95% interval	confidence	F test with true value 0			
	Intraclass Correlation (model type)	Lower Bound	Upper Bound	Value	df1	df2	p
<u>Pre-registered</u>							
DP group <i>all</i> individual measures	.105 (C,1)	.016	.285	2.17	17	153	.007
Control group <i>all</i> individual measures	.209 (C,1)	.130	.321	3.65	50	450	<.001
<u>Exploratory</u>							
DP group individual face <i>perception</i> measures only	.343 (C,1)	.167	.577	4.13	19	95	<.001
Control group individual face <i>perception</i> measures only	.209 (C,1)	.114	.334	2.59	56	280	<.001
DP group <i>average</i> face perception & face memory	.373 (C,2)	-.584	.742	1.59	19	19	.159
Control group <i>average</i> face perception & face memory	.749 (C,2)	.580	.850	3.98	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 I 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 categorisation BIS, gender categorisation BIS, configural/featural face BIS, CFPT upright accuracy).

5.4.6 Summary of group level results

Table 5.13 below summarises the group results.

Table 5.13 Summary of hypothesis testing results at group level

	Data support hypothesis?	Cohen's <i>d</i>	<i>rbb</i>	<i>p</i>	BF ₁₀
Compared with age matched typical controls:					
1a Face detection: DP will be slower to detect faces	Yes		.40	.008**	1.88
1b Face detection: DPs will show reduced normalised face inversion effects	Yes		.35	.020*	1.08
1c Mooney faces: DPs will be less accurate	No		.28	.062	0.60
2 Gender categorisation: DPs will be less accurate	Yes		.37	.015*	1.49
3 Age categorisation: DPs will be less accurate	No			.129	0.71
5a Configural featural: DPs will be <i>more</i> impaired at faces than houses	Yes	0.67		.005**	2.49
5b Configural featural: DPs will be <i>more</i> impaired at spacing than features	No	-0.22		.430	0.38
5c Configural featural: DPs will show reduced face inversion effects	No	0.3		.245	0.47
6a 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 (BF₁₀). The Benjamini Hochberg (BH) method was used to correct for multiple comparison across 13 measures with the false discovery rate set at 0.05, *p* values in bold remained significant after BH adjustment (see Supplementary table S2). The gender and age categorisation outcome measures are Balanced Integration Score.

5.4.7 Research question 6b: Do all DPs show face perception impairments?

Hypothesis 6b: Some, but not all, DPs will be relatively unimpaired at face perception compared to face memory.

5.4.7.1 Motivation for the research question

The DP literature contains many mixed results at group level and individual case level and the condition is widely reported as heterogeneous (for a review see Corrow et al., 2016). Consistent with this, it is clear from DP studies which report individual case level data (e.g., Bate et al., 2019; Dalrymple & Duchaine, 2016; Ulrich et al., 2017) that perceptual deficits are common, though not universal, in DP. Neuropsychology sample sizes are typically small and therefore the likelihood of detecting significant group differences may depend on the make-up of a given sample (Bennetts et al., 2022) or be driven by one or two participants with extreme impairments. Inspection of individual level data can therefore provide useful insights into milder, though common impairments, and elucidate the perceptual heterogeneity in DP. Limitations of previous studies which helpfully reported individual level data include small sample sizes ($n \leq 11$, e.g., Duchaine & Nakayama, 2004; Huis in 't Veld et al., 2012; Schmalzl, Palermo, & Coltheart, 2008; Ulrich et al., 2017) or a focus on a single perceptual sub process (Bate et al., 2019; Biotti et al., 2019; Dalrymple et al., 2014; Duchaine & Nakayama, 2004; Marsh et al., 2019) which preclude a more complete understanding of which sub processes may be affected.

5.4.7.2 Hypothesis 6b: Individual DP results

Table 5.14 illustrates patterns of upright face perception in the DP group. Individual DP z scores from each task are listed and for comparison I show the original measures of accuracy or RT, and also BIS where appropriate.

Table 5.14 DP z scores

ID	Mooney	Face detection		Age categorisation		Gender categorisation		Configural featural faces		CFPT	Global face perception		Global face memory	Tests impaired
	Accuracy	RT	BIS	Accuracy	BIS	Accuracy	BIS	Accuracy	BIS	Accuracy	Accuracy	BIS	BIS	
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
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	4
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	4
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	4
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	4
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
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	3
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
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	3
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	3
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	2
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	2
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
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
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
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	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	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	1
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	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
% DPs														
impaired	15%	50%	30%	5%	35%	25%	40%	30%	35%	50%	15%	40%	100%	65%

Note: Major impairment $z \leq -1.7$ shown in red, mild impairment $-1.69 \leq z \leq -1$ shown in yellow. All scores standardised and centred on age-matched control means. BIS = Balanced Integration Score. A positive z score on RT indicates impairment. 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 control.

5.5.3 Discussion of individual level results

Following the DSM-5 approach to classifying neuropsychological disorders (Sachdev et al., 2014) findings were that 11/20, or 65%, of the DP group had a *perceptual* impairment for faces, scoring at least 1 *SD* below the age matched control mean on more than one test of face perception.

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 0) can be clearly seen. Across the six face perception tasks, the median number of tasks on which individual DPs showed impairment was 3 [range 0-5] indicating a consistent difficulty encoding a visual percept of a face. However, the *patterns* of impairment were highly variable between participants. For example, 50% of the DP group showed difficulty with face detection 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 small 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 generalise 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). My findings using both a larger sample size ($n = 20$) and a greater number of face perception tasks ($n = 6$) are a major advancement on previous work.

Second, although DP is usually thought of as a condition that affects face *identity* recognition, the majority of DPs (13/20) showed impairment on at the earliest stages of face perception impairment- *face detection* and *gender categorization*, as indicated by a z score < -1 on at least two nonidentity sub processes of face perception tasks [face detection (Faces among non faces); face categorisation (Mooney faces); age categorisation; gender categorisation] indicating that the perceptual difficulties in DP commonly, though not always, arise before facial *identity* is encoded. This is consistent with hierarchical models of face processing (Bruce & Young, 1986; Haxby et al., 2000, 2002). Deficits in face identity perception 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 I extend these findings by showing that DP also commonly affects nonidentity aspects of face perception.

Third, a small minority (15%) of DPs (AF01, AF016 and AF020) showed typical early-stage face perception and were impaired *only* on the CFPT, suggesting relatively spared nonidentity face perception but a specific difficulty with *identity based* face perception tasks (or, more 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, although below average in 5/6 cases, were all 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 my 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. Biotti and colleagues (2019) suggest that methodological considerations including the CFPT's poor internal reliability (leading to noisy results), fewer trials and extended presentation time (which may allow DPs to employ compensatory strategies) may reduce the likelihood of DPs falling statistically below typical performance (i.e. ~ 2 *SD*) and thus being considered to have 'unimpaired' face perception. However, this explanation cannot account for my findings since I administered six face perception tasks which, with the exception of the CFPT, showed good internal reliability in both DPs and controls. Furthermore, I followed the DSM-5 approach (Sachdev et al., 2014) which sets a more liberal threshold for impairment of -1 *SD* on more than one task, and nevertheless found that 7/20 (35%) of the DPs in my sample would be considered to have *no* face perception impairment.

5.6 General discussion

20 DP and 60 age matched controls completed a battery of perceptual tasks. At group level DPs showed impairments on early stage, nonidentity, subprocesses aspects of face processing (face detection and gender categorisation) as well as significantly reduced ability versus controls to discriminate changes to the configuration and parts of faces. I found that 13/20 (65%) of the DP group showed an objective face perception impairment of ≤ -1 *SD* versus the age matched control group mean on at least two separate face perception tasks (9 with major impairment, 4 with mild). Across the full test battery, a high level of inter and intra individual variability was observed. My 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 important findings with important theoretical implications.

5.6.1 DP showed impairments in nonidentity face perception

Research questions 1 – 3 investigated the early, nonidentity 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 task) and on a facial gender categorisation task using an integrated measure (BIS). By contrast, there was no support for a significant *group* difference for Mooney faces or age categorisation. However, at the single case level, a notable finding is that the *majority of DPs* (13/20) showed impairment at these very early stages of face perception, as indicated by a z-score < -1 on at least two nonidentity face perception tasks (face detection, Mooney faces, age or gender categorisation). 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 perceptual and one memory task (Biotti et al., 2019), however these considerations cannot account for my findings since my study design mitigated both these potential biases.

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 my findings that DPs are impaired when encoding semantic, *nonidentity* aspects of faces (e.g. gender) is consistent with this notion and furthermore suggest 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 sub processes with impaired face identity recognition.

5.6.2 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.

5.6.3 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, I 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, the group difference was not driven by one or two outliers, perceptual impairments were widespread in the DP group; 13/20 DPs (65%) showed an impairment of at least 1 *SD* alongside their or face memory deficit. By contrast, only 5/60 (8.3%) of controls showed impairment on this global face perception (BIS) measure, a similar proportion to the 9% of controls with mild perceptual impairments recently reported by DeGutis et al. (2024).

5.6.4 Sub types of DP

5.6.4.1 Apperceptive subtype (impaired face perception and face memory)

The largest subgroup of DPs ($n = 13$) showed widespread face perception deficits, affecting early-stage nonidentity 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.

My 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).

5.6.4.2 Mnemonic subtype (unimpaired perception but impaired face memory)

I also found robust evidence for a mnemonic subtype of DP characterised 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*. Of these, 6/20 showed impairment only on a single task, and 1/20 showed no impairment on any task indicating that - despite clear face memory difficulties - their face *perception* ability was similar to typical perceivers'. The finding that around one third of my 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). Further, analysis of published data from a large study of 165 potential DPs (Bate, Bennetts, et al., 2019) showed that 40.4% were *unimpaired* on the CFPT (i.e., 21 of the 52 participants who showed significant impairment on the CFMT and a famous faces task). My 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 the small sample size.

One criticism levelled at some other studies estimating the prevalence of face perception difficulties in DPs is that of “double dipping”. Importantly, I did not use perceptual scores for classifying DP therefore I can be confident that DPs with a perceptual impairment were not over-represented in my sample because of the chosen inclusion criteria. To further ensure that I did not introduce an artificial distinction between the perceptual performance of controls and DPs, I adopted a conservative approach and retained control participants who showed perceptual difficulties. This approach means that my control sample is more representative of the full perceptual ability spectrum. Thus, the present control group means are lower, and group effects smaller, than they would be had I excluded control participants who produced low perceptual scores yet reported no subjective face recognition difficulty.

5.6.4.3 Potential 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 nonidentity 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 can show major impairments on several other aspects of face perception and yet not the CFPT. Taken together with the ICC analyses, the results show that it is not possible to generalise performance from one face perception task to another.

5.6.5 Theoretical implications

I 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 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. I therefore recommend the use of multiple tasks assessing different stages of face perception.

5.6.6 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., 2022; DeGutis et al., 2023). I agree since, consistent with their results, my 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 sub type of DP and showed *no* face perception impairment had 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). I would therefore also argue for the inclusion of DPs with ‘mild’ impairment in research to better characterise the condition since the degree of face perception deficits did not predict the degree of face memory deficits. Indeed, some DPs with unimpaired face perception (on my measures) in fact showed the most substantial face memory deficits.

5.6.7 Limitations

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 showed good internal reliability and I therefore used BIS as an integrated measure.

I included the CFPT as a measure of face identity perception since it 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 my battery so the finding of a possible third sub type 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.

5.6.8 Directions for future research

Due to the need to limit the overall size of the test battery, I inevitably had to exclude other potentially informative tests. I did not, for example, include an identity matching task but recent work (Stantić et al., 2022) suggests that face matching, face perception, and face memory may independently contribute to face identity recognition and that threshold of evidence needed to judge two images of an individual as being the same person may differ across groups with DPs requiring more evidence than typical perceivers. In line with this, a small study by White and colleagues (2017) reported that DPs are primarily impaired at telling faces together (have fewer correct hits on match trials) rather than telling faces apart (but see Berger et al., 2022). If true, this further supports the inclusion of a regular face matching task in future work.

The clear evidence for sub types of DP alongside more detailed understanding of the perceptual sub processes that may be impaired, or relatively preserved, at individual DP level presented in this thesis 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 individualised approach to rehabilitation of DP.

5.7 Conclusion

Two subtypes of DP were identified. The largest DP subgroup showed widespread face perception *and* face memory deficits (the *apperceptive* subtype). A second subgroup showed fairly typical face perception yet a distinct face memory deficit (the *mnemonic* subtype). For the first time my 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 nonidentity face

perception tasks was within the typical range. If using CFPT as the sole perceptual task, these participants would be incorrectly classified as having a mnemonic impairment.

I also 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 categorising facial gender suggesting early-stage perceptual difficulties. However, I 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. I therefore recommend the use of multiple tasks assessing different stages of face perception.

Chapter 6: This condition impacts every aspect of my life: A survey to understand the experience of living with developmental prosopagnosia

Recap and purpose of this chapter

The previous chapters investigated the cognitive underpinnings of DP and considered which measures best classified the condition. Results confirmed a heterogeneous presentation. At individual level different subprocess of face recognition were affected in the DP group. During the recruitment stage of the main studies reported in the previous chapters I conducted screening interviews with all the prospective participants who had indicated that they believed they had face recognition difficulties to confirm their eligibility (i.e. a lifelong difficulty, no visual impairment or history of head injury). During these interviews I asked these participants why they believed their face recognition was particularly poor. The examples provided and explanations given were rich, detailed, fascinating and extremely diverse but I had not sought ethical consent to record the interviews. I therefore decide to investigate the subjective experiences of living with DP in a more systematic way using a mixed methods survey that built on the insights that the initial screening interviews and email correspondence with potential DPs had provided. This survey forms the basis of the present chapter.

6.1 Introduction

Developmental Prosopagnosia (DP) is a lifelong condition characterised by a severe difficulty recognising familiar faces, despite otherwise normal vision, IQ and memory and in the absence of brain damage. It can result in severe psychosocial difficulties, including embarrassment, social anxiety, and difficulties in interpersonal relationships (Dalrymple, Fletcher, et al., 2014; Yardley et al., 2008). It is also a safety concern, particularly for children who have the condition and may not recognise their guardians or mistake a stranger for one (Dalrymple, Fletcher, et al., 2014).

Although DP has a hereditary element (De Haan, 1999; Lee et al., 2010) there is currently no genetic test or biomarker for the condition. Instead, diagnosis or classification are usually made based on impaired performance on one or more behavioural face processing tests plus subjective report following broad guidelines introduced in 2016 (Dalrymple & Palermo, 2016). In the UK, and much of the rest of the

world, it is almost impossible to obtain a diagnosis of DP from a medical professional (Burns, 2023b; Faceblind UK, n.d.) meaning that individuals seeking an explanation for, or confirmation of, their face recognition difficulties often contact university researchers. However, the choice of tests and cut offs varies widely between research groups (see DeGutis et al., 2023 for a review), in part because the question of which tests, or combinations of tests, are most sensitive for detecting DP remains an open one but also due to the different research questions being addressed.

Several recent studies (Bate, Bennetts, et al., 2019; Burns et al., 2022; Lowes et al., 2024) reported that around half of the individuals who present to research groups with subjective face recognition difficulties show no objective impairments when using the most widely adapted classification approach. I refer to these as “subjective DPs” and others have called them “excluded DPs”, (Burns et al., 2022, p. 4293). There are several possible explanations why these individuals may not show objective difficulties in laboratory tests. First, individuals could be exaggerating their difficulties when providing subjective report and are not truly impaired. Second, the tests currently used may be insufficiently sensitive and/or the cut offs commonly used (< -2 or $-1.7 SD$) are too conservative since milder difficulties also result in severe problems in everyday life (Burns et al., 2022; Lowes et al., 2024a). Third, individuals with severe deficits may lack insight into their difficulties (Murray et al., 2018). A recent mixed methods study (Portch et al., 2023) suggested that individuals with DP may find it almost impossible to disentangle the effect of the compensatory strategies which many have developed from their limited face recognition ability itself. If so, then self-report may not always correlate strongly with performance on objective tasks which use tightly controlled face stimuli and experimental conditions that are designed to reduce the effectiveness of such compensatory strategies. It is important therefore to better understand the types of difficulties that individuals with DP, including “subjective DPs”, experience as a more parsimonious explanation for the high proportion of subjective DPs for whom current laboratory tests cannot detect their specific difficulties.

It is also widely reported in the literature that DP is a heterogeneous condition both in terms of presentation and severity (Corrow et al., 2016). Although it has been proposed that the observed heterogeneity in DP may be due, at least in part, to methodological factors (Corrow et al., 2016), in-depth research with members of the same family has revealed different patterns of face processing impairment between family members (Johnen et al., 2014) suggesting that certain sub-processes of face processing may be

differentially impaired in DP. Consistent with the latter study, screening interviews I undertook as part of a large study of objective impairment in DP (Lowes et al., 2024a) provided anecdotal evidence that individuals with DP can experience the condition quite differently in everyday life and provided motivation to explore this question more systematically.

We know much about how such individuals with poor face recognition ability perform on a wide range of lab-based tests (for reviews see e.g. Corrow et al., 2016; Geskin & Behrmann, 2018), but so far very little about what relationship, if any, exists between lab-based test scores and real-world experience (for one exception see Portch et al., 2023). A recent scoping review (Nørkær et al., 2024) identified that between 1976 and 2022 most research published into DP in adults used lab-based experimental or neuroimaging methods. Of the 224 studies identified, only four (Adams et al., 2020b; Grüter & Grüter, 2007; Murray et al., 2018; Yardley et al., 2008) employed qualitative approaches, of which two (Adams et al., 2020b; Murray et al., 2018) were based on the same sample. The review concluded that “it is likely that collection of subjective experiences by participants might open our eyes to as of yet unanswered, and unasked, questions.” (Nørkær et al., 2024, p. 172).

Although DP is a standalone condition, it commonly co-occurs with other neurodevelopmental conditions (Svart & Starrfelt, 2022), including object agnosia (Geskin & Behrmann, 2018), topographical agnosia (Bate, Adams, et al., 2019; Klargaard et al., 2016) and autism (Griffin et al., 2021). A recent meta-analysis of face recognition abilities in autism (Griffin et al., 2021) found that on face identity recognition tasks the average autistic individual will produce scores lower than around 81% of neurotypical individuals. Additionally, it has been estimated that 36% of the autistic population would meet the clinical cut off for DP (Minio-Paluello et al., 2020). While DP research typically excludes participants with other neurodevelopmental conditions in order to study ‘pure’ DP, evidence that DP commonly occurs with other conditions means that the true proportion of individuals living with severely impaired face recognition is likely to be considerably higher than the commonly proposed estimates of 1 - 4% of the population with DP alone (Bennetts et al., 2017; Kennerknecht et al., 2008, but see also DeGutis et al., 2023). Building a better understanding of the experience of living with lifelong face identity recognition difficulties is therefore an important avenue of research in its own right (Nørkær et al., 2024) as well as because of the prevalence of the condition and its potential for debilitating social impact.

6.1.1 The present study

The overall purpose of the present study is to generate a better understanding of the *range* of ways in which the experience of DP manifests in everyday life, and compare the diversity of this experience, from both psychosocial (how does it feel to have difficulties recognising people you know, what is most challenging about living with poor face recognition) and quantitative perspectives. Qualitative insights may identify potential areas of difficulty not tapped by existing face processing tests and could help inform the development of new sensitive measures for detecting and understanding the underpinnings of DP. The present study also investigates how experiences of face recognition difficulties in everyday life may relate, if at all, to performance on widely used objective tests of face processing. To the best of my knowledge, it is the first study to assess relationships between the estimates of real-life performance and individuals' objective face test scores and the first to attempt to obtain quantitative estimates of DPs' face identity recognition ability outside the lab. For the first time I investigate and quantify DPs' experiences of requesting reasonable adjustments or accommodations in the workplace or elsewhere. Such adjustments are increasingly commonly requested, and provided, for other forms of neurodivergence but nothing is known about the experience of requesting these for DP which is currently not classified as a disability and therefore not covered by the UK Equalities Act (Equality Act 2010, 2010), despite negatively affecting quality of life (Dalrymple, Fletcher, et al., 2014; Yardley et al., 2008). Finally, this study is the first to ask what DPs themselves believe should be the priorities for future research into the condition.

6.1.2 Research questions

Overall, I was interested in how poor face recognition abilities affected individuals in their everyday lives and the diversity of this experience within the DP sample. The following broad research questions were preregistered (<https://osf.io/b24ug>):

- How many individual faces can people with DP typically recognise and can DPs recognise close friends and family?
- Is there a relationship between real life reported and objective measures such as PI-20 (Shah, Gaule, Sowden, et al., 2015) and CFMT (Duchaine & Nakayama, 2006) scores?
- What aspects of DP do people find most difficult – and conversely are there any positives?

- What is DPs' experience of seeking a diagnosis?
- What is DP's experience of disclosing their difficulties to others? Have institutions been willing to introduce adaptations and, if so, have these been helpful?
- What do DPs believe the priorities for future research into DP should be?

6.2 Methods

The protocol (<https://osf.io/b24ug>) and survey questions (https://osf.io/dxrb5/?view_only=2dc6276380f34da596c781f198868cb0) were preregistered

6.2.1 Context and sampling strategy

Purposive sampling was used to recruit 29 adults (19 women, 10 men), who had previously contacted the research team because of face recognition difficulties. Participants were recruited from two prosopagnosia research groups: 13 from University of Stirling (experimental results from 12 of the 13 are reported in Lowes et al., 2024a) and 16 from Prof. Brad Duchaine's research group at Dartmouth College, USA. All participants lived in the UK as we wished to examine experiences in a UK context. Objective and subjective face recognition difficulties were confirmed in all participants using the same PI-20 questionnaire (Shah, Gaule, Sowden, et al., 2015) and four objective standardised face memory measures that were previously used to classify the Stirling cohort; these were the Cambridge Face Memory Test (Duchaine & Nakayama, 2006), Old New Faces (Dalrymple, Garrido, et al., 2014), a score measuring the difference between performance on the CFMT and a matched bicycle memory test (Dalrymple, Garrido, et al., 2014), and a Famous Faces Test (Lowes et al., 2024a). One additional potential participant (DA027) showed subjective, but no objective, difficulties and was therefore not invited to complete the survey after the initial screening. All eligible participants reported in Chapters 3-5 who had given permission to be contacted again were invited to participate in this final study.

Although autism and DP co-occur frequently (Griffin et al., 2021), I included autism (diagnosed or self-diagnosed) in the exclusion criteria since it is likely that autistic traits and experiences could mediate or moderate other experiences I was interested in e.g. of social interactions or experiences of seeking healthcare support. My aim in the first instance was to understand the experience of 'pure' prosopagnosia.

Thus, I investigate the experience of a cohort of individuals (DPs) who meet the widely accepted definition of DP, showing both subjective and objective face recognition difficulties. My study does not aim to provide a representative sample of DPs. Sample size was guided primarily by resource limitations: The maximum number of potential participants is constrained by the number of qualifying individuals with prosopagnosia already known to the researchers and time constraints. I initially anticipated 30 participants; 31 eligible participants started the survey but 2 did not complete it, leaving total of 29 complete responses which were analysed.

6.2.2 Ethical issues pertaining to human subjects

Approval for the study was granted by the University of Stirling General University Ethics Panel (reference 14876). All participants provided informed consent, including for the use of de-identified quotes, and research was conducted in accordance with the principles of the Declaration of Helsinki. Data were collected using unique identifiers. The study design and research questions were pre-registered (https://osf.io/dxrb5/?view_only=2dc6276380f34da596c781f198868cb0) prior to data collection. Several possible qualitative approaches were considered at the pre-registration stage, of these I chose framework analysis which was considered the most suitable once the data were explored.

6.2.3 Data collection methods and instruments

Data were collected using a mixed methods approach. Participants completed an online survey comprising a combination of open response, multiple choice and closed response questions. The qualitative questions were designed to encourage detailed responses and many qualitative but also quantitative questions were informed by information provided by potential DPs when they first contacted us. Data collection ran from August 2023 to February 2024 using the Online Surveys platform (www.onlinesurveys.ac.uk). Participants received a £5 gift voucher upon completion. A draft survey was piloted with two prosopagnosic participants using a “think aloud” process. I met with the participants who verbally answered the draft survey questions, explained how they interpreted each question and why they answered as they did. Based on the resulting feedback and suggestions, the wording of several questions was refined to improve clarity. All participants reported here completed the same final survey and questions are provided in Appendix 4. The survey typically took between 20 and 59 minutes to complete.

6.2.4 Participant characteristics

Participant demographics are shown in Table 6.1 below. Participants ($N = 29$; 19 female, 10 male) lived in the UK and were aged between 26 and 73 years ($M_{\text{age}} 53.5 \pm 13.8$). All participants reported normal, or corrected-to-normal, vision, no neurodevelopmental or psychiatric conditions (other than mild dyslexia), or history of brain injury. Three respondents provided somewhat ambiguous responses to specific questions about the onset of their face identity recognition (FIR) difficulties but were included in the analysis since they had all previously confirmed that their FIR issues were lifelong and none of their qualitative responses suggested sudden onset.

6.2.5 Data processing

Where a participant provided a range for their numerical estimates (e.g. 10-20), I input the median value (in this example of a range between 10 and 20, the median would be 15). Face memory test scores were standardised and centred on control age group means (see Chapter 4 for a detailed description of this approach).

Table 6.1 Participant characteristics

ID	Age Group	Gender	Employment status	Family history of face recognition difficulties?	PI-20 score
DA012	16-35	Female		Yes	91
DA005	16-35	Female	Not working due to illness, disability or family reasons	Yes	78
AF017	16-35	Female	Working	Yes	83
AF004	16-35	Male	Working	Yes	68
DA009	36-59	Female	Working	Don't know	81
DA013	36-59	Male	Working	Don't know	77
DA002	36-59	Female	Working	No	70
AF009	36-59	Male	Working	Yes	86
DA001	36-59	Female	Working	Yes	87
AF075	36-59	Female	Unemployed but available to work	No	72
AF125	36-59	Female	Not working due to illness, disability or family reasons	No	92
AF001	36-59	Female	Working	No	81
DA022	36-59	Female	Working	Yes	95
DA015	36-59	Male	Unemployed but available to work	No	82
DA028	36-59	Male	Working	No	85
AF010	36-59	Female	Working	Don't know	86
AF006	36-59	Female	Working	Yes	95
DA020	36-59	Female	Working	No	84
AF003	36-59	Female	Working	Yes	94
AF060	60-74	Female	Retired	No	76
DA018	60-74	Female	Retired	No	86
DA033	60-74	Female	Working	Yes	77
DA014	60-74	Male	Retired	No	87
DA017	60-74	Female	Retired	No	67
DA004	60-74	Male	Working	Don't know	75
DA006	60-74	Male	Retired	No	71
AF019	60-74	Male	Retired	Don't know	99
AF099	60-74	Male	Retired	Yes	74
AF008	60-74	Female	Retired	Yes	77

Note: The PI-20 (Shah, Gaule, Sowden, et al., 2015) is a validated questionnaire measuring self-reported prosopagnosic traits; higher scores indicate more difficulties and the cut off used was 61. Participant IDs beginning AF were recruited via Stirling and DA via Dartmouth.

Additionally, for completeness, the objective and PI20 score used to classify these participants is reported below.

ID	Age	Gender	PI-20	Old Face accuracy	New Faces BIS	FFT	CFMT accuracy	CFMT BIS	CBMT accuracy	CBMT BIS	CMT Difference accuracy	CMT Difference BIS	Data included in chapter
DA012	26	Female	91	-0.19	-2.38	-5.40	-1.92	-3.98	-1.12	-2.20	-0.79	-1.78	6
AF017	29	Female	83	-1.74	-3.17	-0.04	-1.53	-0.17	1.36	2.50	-2.89	-2.67	3,4,5,6
DA005	29	Female	78	0.19	-2.27	0.31	0.76	-1.61	0.17	-0.89	0.60	-0.72	6
AF004	31	Male	68	0.62	-0.67	-1.87	-0.93	-2.20	1.58	0.85	-2.51	-3.05	3,4,5,6
DA009	37	Female	81	-0.28	-3.54	-0.01	-1.74	-3.75	0.82	-2.08	-2.56	-1.67	6
DA013	39	Male	77	-1.94	-3.06	-0.57	-0.03	0.14	0.39	1.44	-0.43	-1.30	6
DA002	40	Female	70	-4.72	-6.12	-0.27	-2.38	-1.46	-1.21	-0.62	-1.18	-0.84	6
AF009	42	Male	86	0.88	0.03	-0.63	-1.23	-5.56	0.77	-0.85	-2.00	-4.71	3,4,5,6
DA001	43	Female	87	-7.50	-8.19	-4.42	-2.71	-2.30	-0.03	0.79	-2.67	-3.08	6
DA027*	44	Female	76	0.28	0.33	0.92	0.71	0.01	1.14	1.09	-0.43	-1.08	
AF075	45	Female	72	0.28	-3.07	-1.40	-0.80	-3.83	-0.02	-2.37	-0.78	-1.46	3,4,5,6
AF001	51	Female	81	-0.91	-2.45	-0.24	-1.23	-0.35	0.38	1.92	-1.60	-2.26	3,4,5,6
AF125	52	Female	92	-0.83	-3.12	0.43	-1.00	-2.74	1.03	0.05	-2.03	-2.79	6
AF010	53	Female	86	-9.84	-12.1	-2.63	-3.05	-3.48	-0.12	0.89	-2.93	-4.37	4,5,6
DA022	53	Female	95	-0.28	-1.18	-2.65	-2.28	-2.41	0.18	0.61	-2.46	-3.03	6
AF006	54	Female	95	-2.70	-2.42	-1.58	-2.09	-1.16	-0.41	0.18	-1.67	-1.34	3,4,5,6
DA015	54	Male	82	-4.72	-5.95	-2.75	-1.64	-1.44	0.29	-0.30	-1.92	-1.14	6
DA20	54	Female	84	0.28	-0.10	0.40	-1.21	-1.37	1.35	2.06	-2.56	-3.44	6
DA28	54	Male	85	-1.39	-3.65	0.25	-1.10	-1.94	-0.46	-1.74	-0.64	-0.20	6
AF003	55	Female	94	0.88	-0.60	-1.52	-1.55	-2.25	0.87	0.04	-2.42	-2.29	3,4,5,6
AF060	61	Female	76	-0.92	-2.94	-1.92	-0.42	-0.07	0.84	0.77	-1.27	-0.85	3,4,5,6
DA18	66	Female	86	-3.47	-5.84	-5.27	-1.10	-5.17	-0.63	-3.30	-0.46	-1.86	6
DA33	66	Female	77	-2.80	-3.23	-1.68	-0.26	-0.23	0.01	-0.52	-0.28	0.29	6
DA14	67	Male	87	-8.80	-9.60	-4.67	-2.95	-2.89	-0.73	-1.05	-2.22	-1.84	6
DA17	67	Female	92	0.53	0.03	1.40	-1.47	-1.74	-0.17	0.08	-1.30	-1.82	6
AF019	68	Male	99	-3.88	-6.92	-5.22	-1.64	-1.72	0.84	0.02	-2.48	-1.74	4,5,6
DA4	68	Male	75	-3.47	-3.59	-3.22	-0.45	-0.55	1.22	1.01	-1.67	-1.56	6
DA6	69	Male	71	-0.80	-1.62	0.07	-0.45	-0.37	0.85	0.58	-1.30	-0.95	6
AF099	70	Male	74	-3.14	-7.53	NA	0.33	0.22	0.09	0.11	0.23	0.11	3,4,5,6
AF008	71	Female	77	-0.17	-5.72	-1.37	-0.52	-1.23	0.84	-2.19	-1.36	0.96	3,4,5,6

Note: * Subjective DP: Not invited to complete the survey and data excluded from analysis. FFT = Famous Faces Test, CFMT = Cambridge Face Memory Tests, CBMT = Cambridge Bicycle memory Test, CMT difference = CFMT – CBMT, Global face memory = standardised mean average. All scores are standardised and centred on age-matched control means.

6.2.6 Data analysis qualitative approach

The qualitative survey data were analysed using framework analysis (FA), a five-stage approach to thematic analysis described in Spencer & Ritchie (Spencer & Ritchie, 1994) that is suitable for both contextual and diagnostic research questions. FA allows codes to be generated both deductively (e.g. from theoretical models of face processing) and inductively (from questions). FA was chosen for several reasons. First, it enables data to be analysed systematically yet flexibly and crucially it also allows *between* case and *within*-case analysis thus allowing individual patterns of responses to be assessed and

compared (Spencer & Ritchie, 1994). Second, FA permits us to analyse the diversity of DPs' experience without prioritising, giving - or appearing to give - additional weight to more common responses (cf. content analysis). My aim is not to describe a "typical" experience of DP but rather to capture the richness and diversity of lived experience. In line with Pyett (2003), I consider each person's experience to be valid and potentially informative, regardless of whether it is typical of the wider cohort, or not. In neuropsychology, there is a long history of single case studies that have resulted in important theoretical insights and therefore an individual's experience, or pattern of responses, does not need to be widely found to have potentially important theoretical and/or explanatory value. Third, I wanted to be able to consider lived experience in the context of each individual's objective performance on face memory tests and FA's use of charts to analyse between-participant responses across multiple questions and measures is well suited to this task.

NVivo 14 (Lumivero, 2023) was used to code the open response data, identify a thematic framework and chart the data. I initially began by coding the open responses inductively and next developed some additional deductive codes. First, the survey questions themselves were used deductively to develop the charting; charting was then expanded to incorporate the key inductive themes that were developed from the coding of individual survey responses. Some quotes in the text have been edited for clarity e.g. to correct spelling mistakes and to change names or other details where these might have identified participants. Quantitative data were explored via descriptive statistics, independent samples *t* tests (or non-parametric alternatives) and correlational analysis using Jamovi (The Jamovi Project, 2023).

6.2.7 Techniques to enhance trustworthiness

FA is considered a type of 'codebook' thematic analysis for which inter-rater reliability and coder consensus are not considered appropriate quality indicators (Braun & Clarke, 2021). JL (a researcher of DP) conducted the primary coding, charting and thematic analysis and kept a reflexive log to track the development of codes and themes and to document key decisions and how potential bias was addressed. At each stage of the analysis, JL discussed coding and analysis decisions with LM, an experienced qualitative health psychology researcher with no experience of prosopagnosia research who thus brings a different perspective and has no prior expectations informed by current face processing theory. LM independently coded four responses; coding differences and

similarities were then discussed. From this the final coding framework was agreed and applied. Thematic decisions were reached through consensus within the author team that includes two additional face researchers.

6.2.8 Researcher characteristics and reflexivity

JL and AB have previously researched prosopagnosia and PH is an experienced face researcher. LM is an experienced health psychologist with research expertise in qualitative methods and the lived experience of health conditions. LM is not a face researcher and therefore brings no prior knowledge of prosopagnosia. None of the research team has prosopagnosia. The sampling frame for this study is individuals who contacted one of two prosopagnosia research groups because of severe face recognition difficulties who had given consent to be contacted about future research studies. JL previously conducted screening interviews by telephone or video call with some participants ($n = 12$) to confirm their eligibility for a prior experimental face recognition study and exchanged emails and administered online testing with the remainder ($n = 17$). No diagnosis was provided to participants although a small number of participants requested their test scores which were supplied after their survey responses were received.

6.3 Quantitative results

6.3.1 How many individual faces can people with DP typically recognise?

Participants were asked to make a private list of the individuals they believed they could reliably recognise *out of context* and to provide the total number of these identities. Estimates varied widely [range 0 -2500], the mode was zero, and the median seven. Overall, more than two thirds of participants (20/29) reported being able to recognise fewer than 10 familiar faces which is far below the estimate of an average of 5000 faces that a typical adult can recognise (Jenkins et al., 2018). Although these studies use different methods to arrive at these estimates, the difference is nonetheless striking.

6.3.2 Can DPs recognise close friends and family?

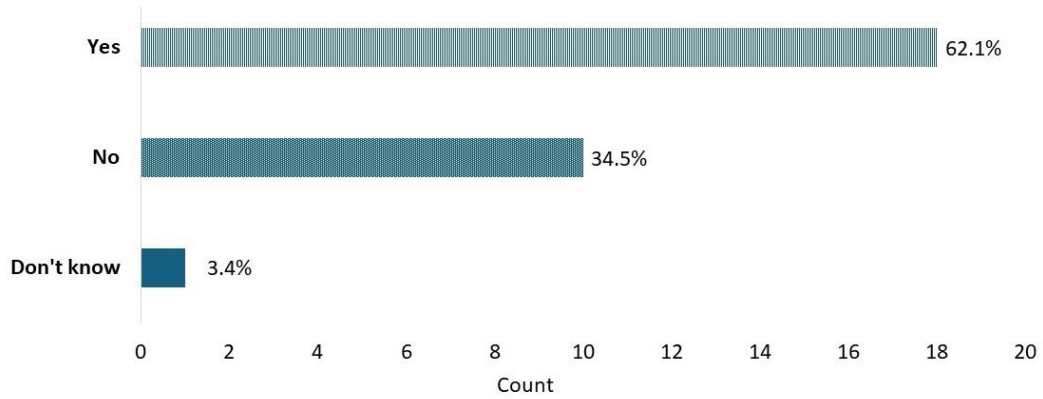
As shown in Figure 6.1 below, although the majority (62%) of participants reported being able to recognise their immediate family e.g. parent, partner or child, strikingly, around a third (35%) were unable to reliably recognise their immediate family

members *out of context*. Even fewer, under half, reported always being able to recognise their three closest friends when encountering them unexpectedly. Chi square or Fishers exact tests showed that these responses did not differ between genders, age groups or recruiting university (all $ps > .09$). Qualitative findings and illustrative examples of difficulties recognising family are discussed below in subtheme 3.3.

When asked how many times participants typically needed to meet someone before they might recognise them, responses again varied widely (Figure 6.2). At one extreme, one participant reported being able to recognise others after only one or two encounters, although they caveated this by stressing that recognition was only possible if they had been able to make an effort to memorise something distinctive about the individual and associate this with their name. This type of deliberative approach is very different to that used by typical perceivers. At the other extreme, two participants reported *never* being able to recognise a face, regardless of the number of encounters and the most common response was between six and ten times. Many participants noted that they found some individuals much easier to recognise again than others. Sometimes this depended on whether the person had distinguishing features, for others it was due to the “intensity” or “recency” of the encounter, or for reasons participants could not articulate.

Figure 6.1 Ability to recognise immediate family and close friends if meeting them unexpectedly out of context

Immediate family (e.g., partner, children, siblings)



Three closest friends

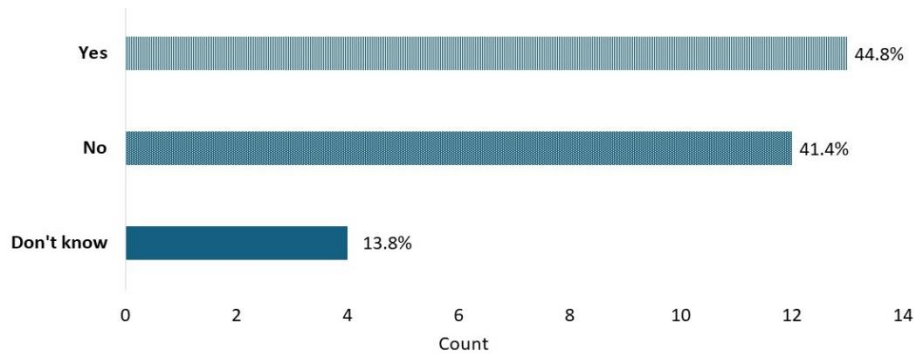
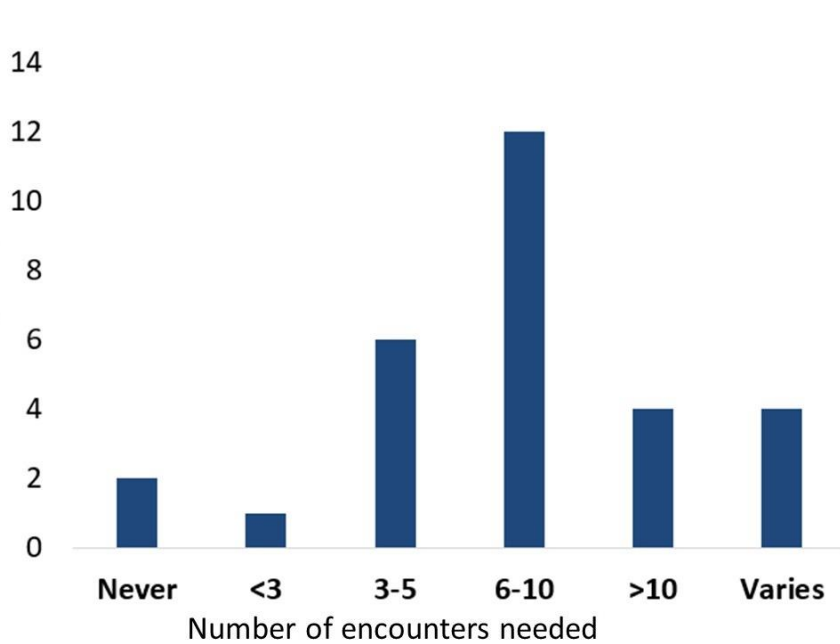


Figure 6.2 Number of encounters DPs needed to be able to recognise a person



6.3.3 Relationship between real life experience and test scores

Next, I asked whether there was a relationship between real life reported experiences e.g. the ability to recognise immediate family or the number of facial identities participants said they could recognise, and objective face memory measures or PI-20 scores. Independent *t* tests (see Appendix 4 Table S3) showed that objective face memory scores and PI-20 scores were not significantly different between the participants who reported they *could* and those who *could not* recognise immediate family, all *ps* > .172 and all Bayes Factors < 0.80). Notably, only 3/10 participants who reported *not* always recognising immediate family would be classified by researchers using a conservative neuropsychological criterion as having DP (i.e. participants who produced CFMT scores more than 2 *SD* below age matched typical controls [the most commonly used cut off for classifying DP]). Further, only 4/10 scored < -2 *SD* on the arguably more ecologically valid Famous Faces Test. These findings indicate that, contrary to what might be anticipated, very few of the subgroup who were unable to reliably recognise immediate family showed ‘major’ impairment on widely used laboratory measures of face memory despite experiencing face recognition difficulties in everyday life that most would certainly consider both severe and highly atypical.

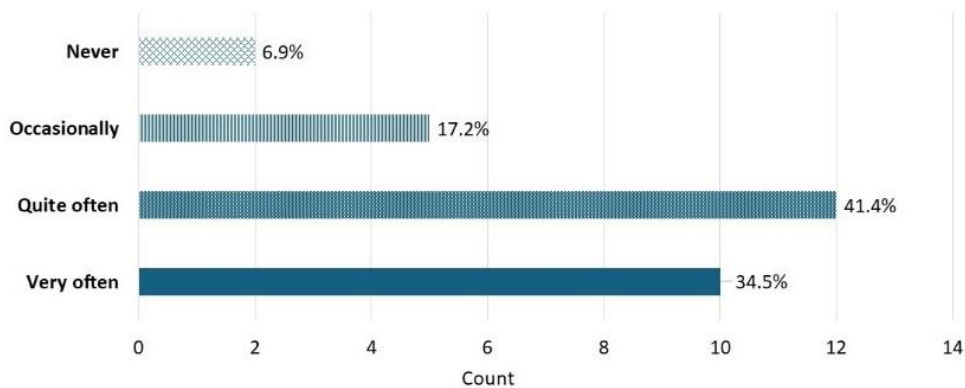
Spearman correlational analysis showed no significant associations between the number of familiar facial identities that participants estimated they could recognize in, or out of, context and any of the test or questionnaire scores (see Appendix 4 Table S2).

6.3.4 What types of face recognition errors do DPs make?

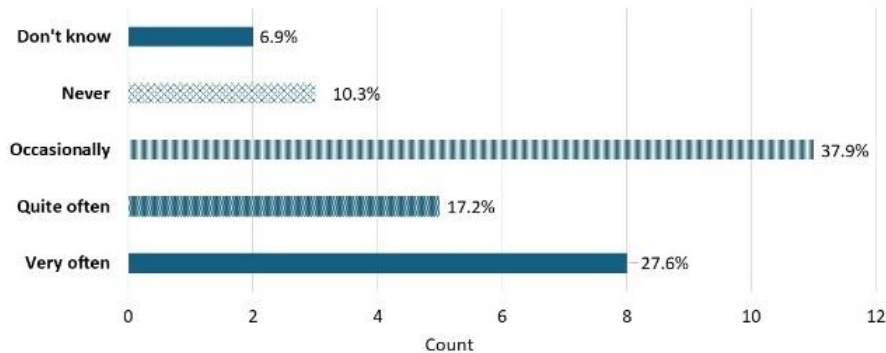
Within this sample, the types of errors that participants reported making varied widely (Figure 6.3). Forgetting previously recognised familiar faces was unsurprisingly a common, though not universal, experience. False alarms (mistaking a stranger for a familiar person) were relatively less common, though still widespread, with only a small number (*n* = 2) of participants reporting never making such errors.

Figure 6.3 Frequency of error types

I forget faces I could previously recognise if I haven't seen them for a while

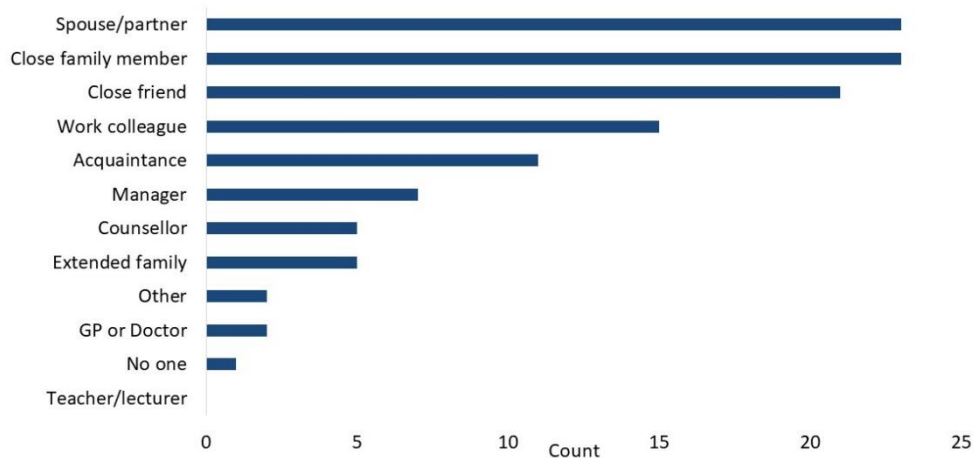


I have mistaken a stranger for someone I know



Note: Bars show the count of respondents (maximum 29) who selected each answer option, with the percentages shown at the end of each bar.

Figure 6.4 Who had participants disclosed their face recognition difficulties to?

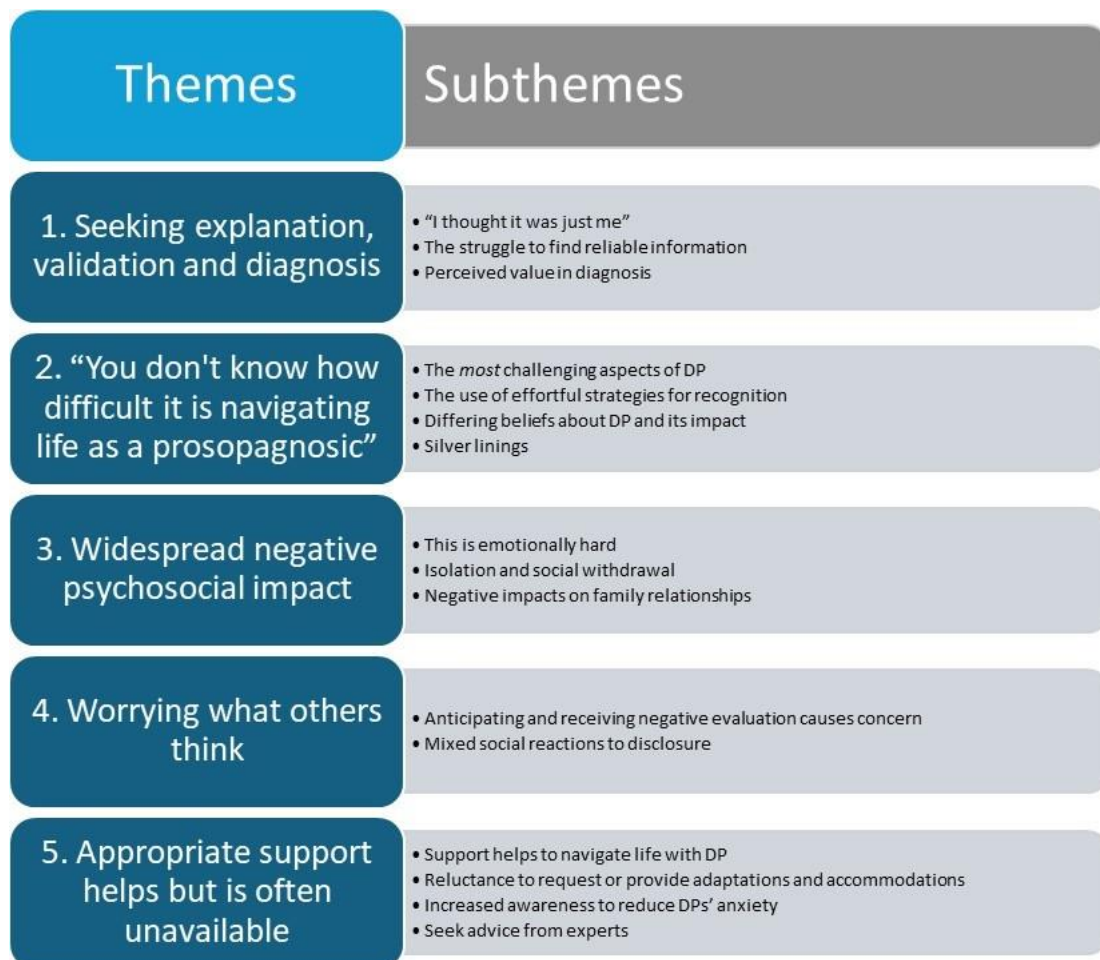


Note: Bars show the count of respondents (maximum 29) who selected each answer option. Participants could select more than one option.

6.4 Qualitative analysis of the written narratives of living with DP

Five themes were identified inductively and deductively from the open ended responses, and themes and sub themes are shown in Figure 6.5 below. A full list of the codes, mapped to themes is provided in Appendix 4 Table S5 Overview of themes, sub themes and associated codes.

Figure 6.5 Themes and sub themes



Theme 1: Seeking explanation, validation, and diagnosis

This theme describes the process by which participants came to realise their face recognition difficulties were atypical, and the existence of DP as a condition. It captures the sources of information participants sought out in order to understand or explain their difficulties, and their experiences of seeking a diagnosis (or classification) whether formal, informal or self-diagnosis and how this made them feel.

Sub theme 1.1 “I thought it was just me”

Although all participants were aware that they struggled to recognise people they knew and had poor face recognition, they had usually been unaware of the existence of DP as a condition for many years. Instead, they often believed their difficulties were “*my own fault*,” unique to them, or a result of “*some inadequacy*.” Several participants reported that their awareness that DP is a scientifically and, in some cases, medically recognised condition (NHS, 2024) came relatively late in life, for example, one participant was in their sixties and others were in their forties.

“I...found a few websites with interesting information, which helped me understand my difficulties were real, and not just something only I experienced, or imagined I experienced.” (AF009)

“I have only been aware of...face blindness properly for only a few years, I didn't know it was a thing.” (DA028)

Apart from one participant who was ‘diagnosed’ by fellow medical students during a lecture covering prosopagnosia, it was much more common for participants to discover the existence of DP or face blindness for themselves. Some described a “*light bulb*” moment when they heard, or read, about prosopagnosia and recognised their experiences in the description. Often this was through media coverage, a podcast, or - less commonly - as a result of studying psychology or reading a novel featuring a character with face blindness (Niven, 2016). Most participants described using the internet both to seek an explanation for their face recognition difficulties (“*Dr Google*”) and to seek further information about DP once they were aware of what to search for.

Sub theme 1.2 The struggle to find reliable information

Almost all ($n = 27$) participants had tried to find information and advice on DP. We enquired about sources and particular sites that participants had used to find out more about DP once they knew about the condition. The internet (“*Dr Google*”) was the main source consulted but relatively few participants could recollect specific sites. Several participants mentioned Facebook self-help/support groups. The most mentioned information sources that participants could remember accessing belonged to university research groups, in particular faceblind.org, a site run by Prof. Brad Duchaine at Dartmouth College, USA and the support organisation Faceblind UK. Three participants

had looked for information on the NHS England website, however two reported they found the information provided there to be of limited value and “very basic.” (DA001).

Sub theme 1.3 Perceived value in diagnosis

Obtaining a formal diagnosis did not initially appear to be especially important to many of the participants, 62% had not sought testing or diagnosis. Possible explanations for the low number of participants ($n = 10$) who had sought a diagnosis are provided by the frequently-expressed perception that (in the UK) there are currently no treatment or support pathways for DP, and beliefs that there is limited awareness and expertise around developmental prosopagnosia within UK health services. As a result, some participants were uncertain about the value of a diagnosis and/or where to seek support:

“I have never discussed it with a medical professional (and my understanding is that it would be hard to access one who knew as much as I do about the condition).” (AF003)

“I didn’t feel like there was anywhere I could go. The NHS website just describes it but doesn’t say what to do next.” (DA013)

A minority of participants ($n = 7$) reported that family or friends had suggested they get tested for prosopagnosia, but it was not clear from responses whether these suggestions were offered after participants shared their own suspicions or concerns about their face recognition abilities, or whether this was unprompted advice.

Seven participants knew of close family members who also had very poor face recognition and one respondent believed her young daughter “might also have issues with faces.” Two participants mentioned that they only become aware of their own parent’s face recognition difficulty after raising their own. Given that many participants described researching prosopagnosia extensively online, it might have been expected that they were also aware that DP often has a hereditary link, so it was initially surprising to us that many participants reported not knowing whether other family members were affected. This lack of knowledge of family members’ difficulties, despite information online about the hereditary nature of DP, is consistent with the finding that relatively few (5/29) reported having told any of their extended family that they had difficulty recognising familiar faces (Figure 6.4) indicating that the topic is not something most participants appeared to feel comfortable or able to bring up, even among family, perhaps due to perceived stigma or self-stigma.

Despite my sampling method, only one participant had received a formal diagnosis (from a researcher) and three more had been told by researchers they met the classification for DP, for them this constituted a diagnosis. A minority were keen to be formally diagnosed themselves; they felt diagnosis could be helpful for achieving some recognition of the challenges they experience as well as providing an explanation to others that, crucially, would be accepted. Even when not formally diagnosed, self-diagnosis through completion of an online test or simply recognising oneself in a description of DP allowed participants to give a name to their difficulties. Learning of the existence of DP as a condition appeared to be very important to many participants from whom the realisation that their difficulties were due to a recognised condition brought relief, validation and a welcome explanation for why they found recognising people so difficult. Only one participant had approached a GP to request testing for possible developmental prosopagnosia, they reported the GP “*Shrugged [their] shoulders*” and said it was “*not possible to test for it and not much you can do about it.*”

We enquired whether participants had disclosed their difficulties to others (separate from seeking a diagnosis). As shown in Figure 6.4, only a small minority (6.9%) had discussed their face recognition difficulties with a doctor, and only 24% had told their manager. Even more strikingly, around 20% (6/29) had not disclosed their difficulties even to immediate family (a similar proportion had not told an intimate partner although it is possible some participants have not been in a relationship since they became aware of DP as a condition). No participant reported discussing their difficulties with a teacher, even though at least two were keenly aware of their difficulties while still at school.

Theme 2: “You don’t know how difficult it is navigating life as a prosopagnosic”

Theme two captures depictions of what it’s like to live with poor face recognition, including whether this changed over time – for example whether certain life stages were more difficult and if the condition gets more, or less, challenging with age. It describes the types of settings and circumstances that participants said they found the hardest and captures the techniques participants use in their everyday lives to try to recognise people as well as the extent to which they find these strategies helpful. We did not want to assume that life with DP is always or necessarily difficult, so we enquired about any perceived benefits or advantages of the condition. It captures beliefs participants had about their DP and how these affected their self-identity.

Sub theme 2.1 The *most* challenging aspects of DP

We asked participants what they found most challenging about having poor face recognition. Responses tended to highlight either situational factors, discussed here, or the emotional impacts which are reported below. Meeting people out of context was commonly reported as challenging, as were situations where large numbers of people were present, for example conferences, parties, and large social events:

“My daughter started school last year and having to meet 30 new kids and 60 new parents was pretty much my worst nightmare!” (DA001)

“Large rooms are hard, it’s a bit Where’s Wally. Everyone merges into non people.” (DA012)

For most participants, DP did not solely affect their ability to recognise acquaintances or individuals they had only met briefly. Rather, it also caused considerable difficulties recognising people with whom they had close relationships and/or saw very frequently including office colleagues, close friends, neighbours and family:

“I could easily walk past pretty much anyone I knew except possibly my husband and children. It is isolating and exhausting living in a world of strangers.” (AF125)

“Sometimes when I am off work for a week and come back it’s really hard to figure out who is who.” (AF017)

“If there are lots of people around it is particularly challenging to even find a familiar face. For instance, in a shopping centre or airport with my husband I will ask him to wait in a specific place if I visit the loo so that I can find him when I come out...” (AF009)

Work was frequently mentioned as a particularly challenging setting. Reasons given included the need to interact with relatively large numbers of people, for example in universities, schools, hospitals or business consultancies where teams and clients changed regularly. In some work settings, uniforms were worn thus removing cues that participants often relied on such as clothing and accessories. For others, the high

cognitive demands of work prevented the use of the deliberative, associative face learning and recognition strategies they usually relied on.

“If my brain is busy and I’m focusing on multiple other things at the time of meeting someone I may struggle to recognise them later that same day” ...If I’ve been introduced to multiple people at one time I might struggle to identify them later on. This is particularly severe if...trying to retain other information relating to the job I’m doing or working on something particularly complicated.” (AF004)

Although most participants provided detailed accounts of settings, or circumstances, they found particularly difficult, a small minority ($n = 2$) initially reported (in response to a direct question) that their face recognition difficulties cause no major challenges in their life. Despite this, both later reported embarrassment during social encounters, with one going as far as to say that when they do not recognise someone they know “Through the eyes of person who recognises me it most likely devalues what they considered to be a close relationship and can actually change the way in which they think of you but in a negative way” (DA014). This comment suggests that it is possible some participants may simply prefer not to dwell on the difficulties their FIR problems cause, rather than that they do not experience such challenges.

Subtheme 2.2 The use of effortful strategies for recognition

I enquired about the strategies people used to recognise others. Spontaneous responses are categorised in Table 6.2 below alongside examples of how and when strategies can fail. After participants provided spontaneous responses, they were then provided with a list of strategies that DPs had previously described using and asked to indicate if they had ever tried them, and, if so, to rate how useful they found each strategy (Figure 6.3 above).

Table 6.2 Compensatory strategies (spontaneous report)

	Examples of strategy use	Examples of strategy failures
Situational context (e.g. where a person sits at work or in an orchestra)	<i>“Think about who I expect to be present and work by process of elimination.”</i> <i>“Context and expectation is everything...if I know someone would be there, then I stand a much better chance of recognising them.”</i>	<i>“It's hardest when the situation is entirely out of context - school mum in supermarket.”</i> <i>“...at a wedding or similar multi-generational event - where some people are my friends, some are friends of parents that I've known from growing up...there's no context around anyone...”</i>
Context derived from conversational cues	<i>“When I met someone that it is clear I should recognise but don't, I tend to play along with generalised comments and fish for clues as to context and identity.”</i> <i>“Listening to what they are saying in the hope that this will give an indication of how I know them.”</i>	
Voice	<i>“Accents help”</i> <i>“Voice - I'm good at recognising voices and am usually fine once someone speaks.”</i>	<i>“Voice is no good as a recognition cue, because I don't recognise voices any better than faces”</i> <i>“If the person is not speaking, I lose another clue that I use to recognise people”.</i> <i>“Starting a new job when I know I will meet many new people within a short space of time. I won't have had time to learn their voices or other things that would help me to tell them apart”</i>

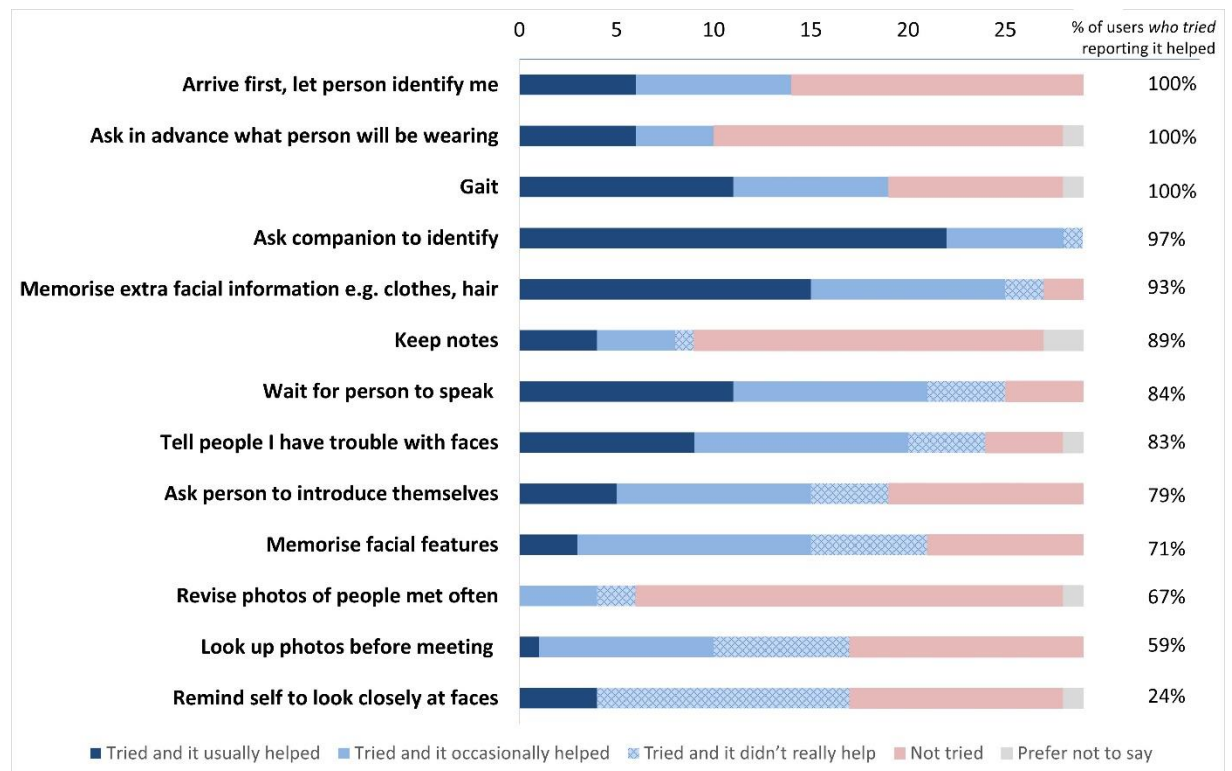
Body (gait, posture, shape, build)	<i>“Look at their overall body profile, height & build”</i> <i>“Some people have a distinctive walk”</i>	<i>“It’s...difficult if I have only seen that person over software such as Teams if they then come into the office it’s weird as I have never seen a body to attach to the face, i.e. .I know Pete is Pete as he is the tallest in the office.”</i> <i>“Meetings where people are seated are harder than standing-up occasions because it's harder to use e.g. gait clues</i>
Hair (including hair colour, style, length, facial hair or baldness)	<i>“I attempt to remember things like hair colour and style...”</i>	<i>“When someone changes something like their hair etc it can be a challenge.”</i> <i>“Sometimes a distinctive facial feature i.e. bald or have a beard. Though that one can backfire if they change their appearance.”</i>
Mannerisms	<i>“Oddly enough, in my year at school, we had a pair of identical twins. My friends told me they couldn't tell the twins apart, but I found it rather easy. I can't quite put into words why...I suspect it was down to their mannerisms. I guess my friends focused on other details than I did to ID the twins.”</i>	
Memorise facial features or style of make up	<i>“Memorising distinctive facial features, e.g. thick eyebrows, shape of chin etc.”</i> <i>“Distinctive unsightly facial features e.g. scars”</i>	<i>“Due to the nature of my work I'm often in dark venues or on dark stages and therefore unable to pick out particular features.”</i>
Memorise features extra-facial (glasses, accessories, tattoos, piercings, teeth, ears)	<i>“Distinctive features - tattoos or piercings are a good thing to look out for as they typically do not change”</i> <i>“By trying to remember something specific such as a piece of jewellery...”</i>	

Smell	<i>“Sometimes smell helps, which sounds weird but if someone has a strong aftershave or perfume it can help me to identify that person.”</i>	
Clothing or clothing style, including shoes	<i>“Clothes they characteristically wear...”</i> <i>“Remember clearly looking at people's shoes at school to know who they were.”</i>	<i>“I have trouble following plots in films because as soon as the scene changes, and the actor's clothes have changed, I struggle to recognise whether it is the same person or a new character.”</i> <i>“I work in an area where there are lots of people wearing the same attire and this can be especially difficult. Also as it is in healthcare and my patients change into gowns I can't recall who they are.”</i> <i>“I recognised one particular mum by a specific coat (a red puffer jacket) she wore every day in the winter. But when it came to the spring term, she no longer wore the coat so I couldn't recognise her. Then even worse... I found out that there were TWO mums with the same coat, and I'd been regularly calling one of them by the other's name, not realising they were two different people, aargh!”</i>
Things people typically have with them	<i>“They might also have a baby or toddler with them in a pram/buggy that I recognise”</i> <i>“Their dogs, cars, musical instruments”</i>	<i>“The other day I got chatting to a man whom I presumed I had met before because I knew his dog. It turned out it was his wife who had been walking the dog when I had met it several times before. I was surprised that I had forgotten the sex of the person as well as the face.”</i>

Wait for others to make the first move	<i>“When agreeing to meet a friend (e.g. for coffee), I often wait for them to greet me.”</i>	
	<i>“If I see someone I THINK I recognise, I don't acknowledge them or make eye contact until they make the first move.”</i>	
Use of written notes	<i>“I used to work in a large supermarket, so I made an excel spreadsheet of names/ hair colour/hair length/distinguishing features/their typical greeting/hobbies to tell people apart. I kept it on my phone so it was to hand when I needed it.”</i>	
	<i>“When teaching, I ask my class to introduce themselves the first time and draw a plan of who's sitting where and draw sketches of their bag while they're doing an activity later. They tend to sit in the same spaces, or next to the same people, in subsequent lessons.”</i>	
Name prompts e.g. name badges, online meeting technology	<i>“When I was working and people wore a security pass I tried to glance at that to see their name.”</i>	<i>“Name badges are really helpful but often small and difficult to read or really obvious that you're looking at them when you should know somebody.”</i>
	<i>“During the pandemic it was a lot easier, since on Zoom the names were always displayed, and I think I relied too much on these and focused even less on how people look.”</i>	<i>“Online meetings where... Zoom etc... would tell you who everyone was were great! I am really sad that so many meetings have gone back to being in person.”</i>
Prompts from others	<i>I have to rely on my children to prompt me and, again, that makes me feel stupid.</i>	<i>“I usually just ask my husband who they are. If he's not there, then I just act politely and keep the conversation vague and then feel bad afterwards for not remembering who it was.”</i>
	<i>“I ask my husband for help like I will whisper, ‘is that Chris?’”</i>	

	<i>“In a work context (e.g. meeting clients) I often struggle. When working with other members of my team I brief them in advance to introduce themselves if they haven't met the client before, so that I can have firm confirmation that it's the right person!”</i>	<i>“Asking my wife to prompt me when she knows I do know someone is best, if not always practical.”</i>
Other strategies	<i>I have severe difficulty identifying people out of context, when I generally can't say whether I know them or not; but I think I can often tell by the way they look at me. I cultivate a unique personal appearance, which makes people more likely to look and give me a clue as to whether they recognise me or not!</i> <i>“Quite often, person X will remind me of person Y and I'll remember that fact, which will happen reliably even if I don't know why...”</i> <i>“Going to the football is another time when I notice I have more difficulties than other people, I couldn't recognise the players I watched every week even when I was going to most games. I always had to rely on the shirt numbers and their position on the pitch to try and make an educated guess...except for players [who were] particularly distinctive.”</i>	<i>“(Of course that's only useful if I am not going to see X and Y in the same context! Unfortunately our new Head of School reminds me strongly of an existing colleague...)”</i>

Figure 6.6 Compensatory strategies rated for usefulness by participants



Note: Bars show count. The value shown at the end of each bar is the proportion of *those participants who had tried the strategy* who reported they found it helpful, at least some of the time. For each listed strategy participants selected one of five possible responses: ‘Tried and it usually helped’; ‘Tried and it occasionally helped’; ‘Tried and it didn’t really help’; ‘Haven’t tried’, ‘Prefer not to say’.

The complexity of strategies and approaches described highlights the sheer level of “conscious effort” that many participants were required to make to try to recognise familiar people, a task that those with typical face recognition ability can achieve rapidly and effortlessly. Participants were usually aware that some strategies were more reliable for them than others, and that particular approaches often failed (particularly reliance on clothing, situational context and hair). As a result, participants often adopted a staged approach whereby they developed a range of potential strategies and tried them out either in combination, or in turn:

“It’s more helpful to learn the things less likely to change often, e.g. voice > tattoos > body shape > hair colour > hairstyle > ...and then clothing etc. is less helpful in a long term sense.” (DA001)

This participant emphasised that sometimes it was the *combination* of cues that most aided recognition:

“...it's easier to recognise a couple (ah yes, she's the petite athletic-looking Asian lady, who has a very tall white husband with a beard, etc.) because the combination of features is more unique.” (DA001)

Although these atypical strategies could often assist with person identification, it was not always possible to employ them (see second column, Table 6.2). Furthermore, many participants reported that they found their strategies were much less helpful when meeting multiple people at the same time, perhaps explaining why so many participants said they found group settings particularly challenging. The strategies were also reported by many to be more helpful in the short, rather than the medium or longer term:

“I can do a sort of "emergency recognise", where I memorise all the shapes of the facial features in the same way you would with a memory game (e.g. memorising objects in a room etc.) It only lasts for an hour or two, and if I encountered them again in the future I'd have no better chance of recognising them than anyone else.” (AF006)

Participants' regular, embarrassing failures of face recognition were not due to lack of trying. All participants reported having tried at least four different strategies [median = 8, range 4 – 13] with varying degrees of success (Table 6.6). For example, the use of written notes was mentioned by 28% of participants and found helpful by 100% of those who used it; this took the form of making detailed notes about appearance, mannerisms, characteristics, typical greetings and hobbies, as well as or sketching bags or typical seating plans. Additionally, several participants described actively “revising” their notes in advance of expected encounters.

Most participants described using conscious, deliberative strategies to assist with identification. However, one participant reported that their strategy use was relatively unconscious, at least initially *“I think I've always used these strategies even before I knew about prosopagnosia.” (AF010)* and another wrote that they had not realised that they were using any strategies until they read the list of possible strategies that others had reported using.

Less common, but interesting, recognition techniques included memorising, or keeping a record of, the typical greetings or idiosyncratic phrases that people use and a person's smell (e.g. a distinctive fragrance worn). The increasing use of online meetings in both social and work contexts was mentioned by several participants as helpful, both because names appear beneath each face and because virtual meetings provide an opportunity to take screenshots of faces with names attached that can be used for future reference.

Sub theme 2.3 Differing beliefs about DP and its impact

Several participants highlighted how, with hindsight, their face recognition difficulties have made certain life stages relatively easier or more difficult for them. A few participants felt that their condition affected them more as they aged, either because as *“a youngster I was more frequently going out with friends and now realize I used to rely on them to “remind” me who someone was”* (AF099) or because the number of people one is *“expected to”* recognise increases with age (AF003). The natural decline of memory in older adulthood was also highlighted as a factor by two participants. For one, this was because poorer general memory made it more difficult to implement their previously-successful strategy of memorising features. By contrast, another felt they were more likely to be misjudged as *“suffering from dementia (now that I’m a bit older)”* (DA018).

Interestingly, one participant felt her face recognition abilities might have a hormonal link:

“This might sound slightly crazy, but ...I find that I’m slightly better at recognising people during certain times of the month. Also whilst I was pregnant I went through a couple of months where I thought I recognised absolutely EVERYONE! That strong “I’m sure I know you from somewhere” feeling. (It felt like a super power, even though I’m sure I was 99% wrong, haha).” (DA001)

Over time, some participants reflected that they gained some acceptance either of the condition itself as *“a limitation that I can do very little about”* (DA020); or of the discomfort it causes during social interactions:

“...learning to accept the awkwardness of that second meeting where I am unaware of who I am talking to and just plodding on with the conversation anyway has also helped, as I will often remember who the person is eventually...” (AF099)

Although neither the participant information, nor the survey questions, framed DP as a disability or a form of neurodiversity or neurodivergence, it was notable that around a quarter of participants used these terms themselves when describing their experience, indicating that they had considered whether DP should be categorised in these terms. These participants expressed mixed views on whether they themselves considered DP a disability, but frequently commented that, despite the severe challenges DP presents, UK society does not currently recognise DP as a disability.

“I tend not to think of it as a disability, more an eccentricity.” (DA033)

“I don't have a recognised disability but, also, I can't do what other people can do so I'm at a huge disadvantage...” (AF125)

“I like to think of it as a superpower rather than a disability. But if we class it as a disability, it's a hidden disability. People don't realise you're faceblind unless you tell them, and it's usually not on their radar...” (DA012)

Others described DP, both directly and indirectly, as a form of neurodiversity. One participant expressed the wish that DP could be *“ recognised by HR professionals as on a par with neurodiversity”* (DA006) with another stating that *“It needs to be seen as part of a spectrum of human abilities and not a disease.”* (DA022)

Sub theme 2.4 Silver linings

We did not want to make the assumption that the experience of living with developmental process recognises was entirely negative so we enquired about what positives, if any, DP brought. One participant reported meeting his now wife as a result of mistaking her for someone else *“something I am always teased about”* (DA014)

Fascinatingly, three separate participants (DA012, DA015, AF019) noted that could easily tell identical twins apart when their friends could not. Other participants

worked with clients in sensitive confidential settings and found some occasional benefits to having DP.

“It is actually weirdly handy for my work role. I work face to face with vulnerable clients in the community and I know I will unlikely have that awkward moment of bumping into service users out and about.” (AF017)

An academic reported a similar benefit since, for example, they would never remember which student they had needed to tell off. A small number of participants felt that they had been able to turn some difficulties that DP presented into positives. For example, attentive listening (due to the need to seek conversational cues to identity), an ability to “pivot conversations” easily (when suddenly realising the person they were speaking to was not the person they had initially thought it was), and public speaking in front of strangers (all or most people look like a stranger to them). One participant wondered:

“I have a unique mind, somewhat gifted in music, mathematics and wordplay. I have a lot of empathy, with animals as well as humans. Maybe these abilities are a result of space freed up in my brain where most people have a database of faces?” (AF019)

However, despite these perceived silver linings to the condition, almost half of participants (13/29) could think of no benefits:

“None at all. It's rubbish.” (AF006)

Theme 3: Widespread negative psychosocial impact

This theme captures the negative social and emotional impacts of living with DP. Although we did not directly ask about the psychosocial impact, responses to numerous open-ended questions consistently highlighted that DP presented both cognitive challenges requiring considerable effort from participants to learn and recognise faces (see Theme 2), and also negatively impacted all participants emotionally and psychologically, at least to some extent. This impact was felt across multiple domains of participants' lives including social relationships, restricted social development, family life, school and career:

“This condition impacts every aspect of my life - my relationships, friendships, work, activities and social life. It deserves more attention as it can have a seriously detrimental impact on the physical and mental health of those with the condition.” (AF125)

Sub theme 3.1 This is emotionally hard

Participants commonly reported that knowing they were going to have to attempt to recognise others e.g., at a work meeting or socially made them feel “nervous”, “awkward”, “uncomfortable” or like “a rabbit in the headlights” and often resulted in feelings of “anxiety”.

“Any occasion, social or work, where one is expected to mingle and meet new people is hard; I can't tell whether someone is a new person or not, so I can't ever introduce myself to anyone (it might turn out we've worked together for 10 years).” (AF003)

For many, these feelings were much stronger than the “mild embarrassment” reported by one participant, and negatively affected many aspects of their social life. For example:

“I avoid groups of more than one or two people and feel very nervous in a group.... It is a shame as I love being with people... I rarely take classes or join groups with a hobby, because groups make me panic, it's a big amorphous blob with lots of heads.” (AF125)

Sub theme 3.2 Isolation and social withdrawal

“Social occasions are terrifying and exhausting. I am warm and friendly, just exhausted from trying so hard and very, very worried.” (AF125)

As a result of finding social situations so emotionally challenging, many participants reported often withdrawing, when possible, from social situations leading to a social life that was rather “stifled” as a result of their condition. Others said they deliberately restricted their social circle to a small number of close friends either because it took “monumental effort to learn a new friend” or because doing so limited the number of faces they were “expected to know” thus reducing the likelihood of “embarrassing” and “frustrating” failures of face recognition.

“To be honest, I usually avoid eye contact with a lot of people and definitely avoid starting conversation because that will make for an awkward meeting and me blushing and becoming embarrassed...I don't have very good strategies, mainly avoidance of interaction.” (DA005)

“I avoid crowds and parties.” (AF019)

“I keep a very small in-person network, which works well as a strategy for facial recognition. Though sometimes I do feel isolated.” (DA012)

This sense of isolation was reported by many, though certainly not all, participants and several remarked that it had become more apparent in mid to late adulthood as they reflected on their life:

“It is isolating and exhausting living in a world of strangers” (AF0125)

“When I'm thinking back now, I realise how isolating this has been. I would meet people, talk to them and maybe have a real connection with them, but then meeting them again, I'd talk to them as if I hardly knew them until I, maybe, managed to identify them.” (DA18)

Some participants believed that avoidance of social gatherings while growing up had also contributed to relatively underdeveloped social skills:

“I am unsure if I was always this bad, looking back I now believe this has been a big factor in my social skills which are rather weak at best.” (DA028)

“Because I've always avoided social situations, I haven't had enough practice at how to behave, so I'm socially very awkward and often say the wrong thing, or act inappropriately because I wasn't properly socialised as a teenager.” (AF006)

Although the vast majority of participants perceived that their conditions negatively affected their social and work relationships, such feelings were not unanimously expressed, one participant felt that that DP *“hasn't affected relationships with closer friends and immediate colleagues.” (DA009)*

Sub theme 3.3 Negative impacts on family relationships.

For some participants, the impact of DP extended beyond relationships with friends and colleagues to their family relationships.

“My husband who I have been married to for 20 years, surprised me at the airport and I didn’t recognise him...” (DA2)

“I was accused by an ex-boyfriend of being racist because I didn't recognise him when I bumped into him unexpectedly in the street, and he said "I suppose we all look the same to you. This was pre-diagnosis, I didn't know what prosopagnosia was, so I didn't have a reply, and just felt awful because I thought he must be right.” (AF006)

One participant wrote movingly about how DP meant they were unable to develop a close bond with their extended family, echoing the beliefs reported above that DP had limited some participants’ social relationships:

“My dad was from a large Irish family...and I spent summer holidays in Ireland as a kid. I could never tell my many relatives apart, which I think affected my development of relationships with them. Now my dad has passed away, I desperately want to feel connected with the family, I miss being a part of it, but don't feel I was ever able to get to know people properly so can't reconnect now in my 40s.” (DA001)

Furthermore, multiple qualitative responses confirmed the quantitative results that recognition of immediate and wider family can also be affected (see 6.3.2 Can DPs recognise close friends and family? above). For example:

“Whilst I see faces and their features, they do not stay in my mind and I cannot easily see their face in my mind until it has permanently imprinted on my brain. I can just about recall the faces of my children but that is all. Everyone else has a blank face in my mind.” (AF125)

“Find my own children hard to tell apart in baby photos.” (AF001)

“I have consistently struggled to recognise several people [on my own doorstep] including my sister who have turned up unexpectedly.” (DA013)

Another participant who, despite responding that they *could* recognise their immediate family, later described an instance when they had not recognised their son when he came unexpectedly to meet them in a local shop. It is possible therefore that the proportion of DPs who can reliably recognise immediate family is lower than reported, either because of faulty recall or because being unable to recognise one's own family is perceived to be socially unacceptable as the quote below illustrates:

"I would never dare to tell my husband, who now knows about my problem, or my sons that there are times when I have not recognised them..." (DA018)

The extent to which participants reported receiving emotional and practical support from family members was again very mixed. Many reported that their partners and family were "understanding", often quietly reminded them who people were, and made efforts to research the condition:

"My family are supportive and never make me feel uncomfortable about me not being able to recognise faces." (DA005)

However, around a third of participants also reported being "teased" about their difficulties recognising familiar faces and family members were not universally supportive:

"My close family sometimes tease me about it but are generally supportive." (DA022)

"Close family not really interested and tend to think I'm making a fuss about nothing, apart from my dad (now passed away) who was delighted to find we had it in common. Partner initially thought I was being overly melodramatic, but I think is gradually coming to understand how distressing it is to realise after chatting to someone that it wasn't who you thought it was, and you've been talking about things they have no clue about." (AF006)

Theme 4: Worrying what others think

Themes two and three capture just how difficult, effortful, and emotionally draining the vast majority participants found the business of navigating everyday life.

Analysis of responses revealed that these challenges often appeared to stem from pervasive worries, usually founded in prior experience, about being negatively judged by others. Theme four captures how extremely important the positive regard of others is to DPs. It contains examples of the widespread and prominent concerns that almost all DPs expressed about potential and actual negative evaluation from others, a deep-seated worry that others would misinterpret their difficulty recognizing others for rudeness, being stuck up, cold or uncaring. Or, perhaps worse, others might make generalisations from their face memory problems and infer that DPs were “incapable”, “stupid” or even suffering from dementia. This theme captures how participants *anticipated* negative reaction and judgments, i.e. guessing how others might interpret and judge their difficulties and also *direct experienced* other’s negative reactions both in relation to observed instances of failures to recognise familiar people as well as experience of disclosure. The latter may well inform the former.

Subtheme 4.1 Anticipating and receiving negative evaluation causes concern

The majority of participants perceived a strong social expectation that they should be able to recognise and therefore acknowledge people they had previously met and were aware that their condition meant they often violated these expectations. For many participants, it was feeling “*pressured to recognise people I’m supposed to know*”(AF125), rather than the cognitive difficulties associated with DP, which presented their greatest difficulties:

“I would say the most challenging [aspect of having DP] for me is when people seem put out by the fact I don’t recognise them and they think I should.” (AF009)

“I hate the beginning of parties when I’m introduced to someone...I try to memorise their name and what they are wearing so I don’t offend them later on by not recognising them” (DA001)

This sub theme was one of the most widespread, more than two thirds of participants expressed worries that their face recognition difficulty would be misinterpreted leading to negative evaluation from others. For example:

“If I think I recognize someone I always face a dilemma - should I smile and say hello (and risk being taken for a weirdo) or ignore them (and risk being thought stand-offish)?” (AF099)

“...people think I am snooty as I don't say hello to them because I don't recognise them...” (DA033)

“I am naturally a sociable person, but I worry that I come across as stand-offish e.g. when I don't recognise people in the park or the playground.” (DA001)

A few participants directly reported that this concern that they would be misjudged as uncaring, stupid, or rude led to anxiety or reduced self-confidence in social contexts linking back to subtheme 3.1:

“I have anxiety in social situations because I constantly think there are people there thinking I'm rude because I don't approach anyone, ever.” (DA002)

“I worry that people will think I am rude or that I don't care about them. I feel anxious when I know I will meet a lot of new people e.g. new job/new social activity” (DA022)

“It massively affects my confidence when I'm in new situations as I know I will either come across as a bit dim or forgetful or cold and rude.” (AF125)

Additionally, one participant described how the worry itself exacerbated her face recognition difficult and this insight may explain why some participants can perform above their true ability on computerised laboratory tests where the possibility of immediate, direct, negative social evaluation is absent:

*“In a group setting, it's often useful to walk *away* from someone, even look at them from behind and from some distance away, because I often find that worrying about their face distracts me from using the strategies that do work for me - if I watch how someone walks, from behind, that will often settle it!” (AF003)*

Although most participants reported worrying about what others might think in the context of in person social interactions, this could at times also extend to online social communication:

“I worry that I over-compensate in non face-to-face situations, e.g. chatting on the WhatsApp group - paranoid that I talk too much, or come across as too dominant, as I'm really just overcompensating for those face to face interactions.” (DA001)

Interestingly one participant reflected that they believed they experienced *less* negative evaluation when they disclosed their face recognition difficulties to others:

“Mostly, I feel that it's a good idea to have told people, because then if later I fail to recognise them in a way which someone might find insulting, even if I have to remind them about my prosopagnosia, it goes better if they have a vague memory that I told them this before. (If you wait until someone is already feeling insulted and then mention prosopagnosia, I think it can just come over as an excuse.)” (AF003)

Sub theme 4.2 Mixed social reactions to disclosure

Initially, there appeared to be a tension between participants' reticence to openly disclose their condition (see Figure 6.4 Who had participants disclosed their face recognition difficulties to?) something that might be thought to reduce the widespread fear among participants that they would be misjudged (as illustrated by the previous quote). However, a possible explanation for this reluctance on the part of many participants to tell others about their face recognition difficulties may be rooted in their past experiences of disclosure. Participants who had disclosed their suspected or confirmed DP status condition were asked about the kind of reactions they received and they reported a very wide range of responses to disclosures of DP, whether officially diagnosed or not.

Overall, most participants had experienced a sympathetic and supportive response from at least *some* of those they told about their difficulties, however this was by no means a universal experience and many participants perceived that their disclosure was

met with a degree of suspicion and disbelief. Around one third experienced supportive responses ($n = 10$) and felt believed and affirmed ($n = 9$):

“Very surprised and interested in the condition but usually accepting and supportive.” (DA014)

“My family are supportive and never make me feel uncomfortable about...not being able to recognise faces” (DA005)

However, positive reactions to disclosure of DP was by no means a universal experience. Negative reactions were also fairly widely reported ($n = 8$), including from close family and many participants perceived that their disclosure was met with a degree of suspicion. Participants reported scepticism or even “disbelief” both that such a condition exists, and/or that the participant has it. Further, some participants ($n = 2$) reported “very little response or acknowledgment” or a “lack of response” to their disclosure, suggesting that they had hoped for a more positive reaction, or at least some acknowledgement of their difficulty.

“Not sure that friends or work believe that it is a proper condition.” (AF010)

“My mother didn't believe me” (DA020)

“My manager at work was horrible and said impatiently “honestly! We ALL don't recognise people from time to time, it's nothing unusual...Partner initially thought I was being overly melodramatic, but I think is gradually coming to understand how distressing it is to realise after chatting to someone that it wasn't who you thought it was.” (AF006)

Others described mixed responses to their disclosure with some people being ‘sceptical’ or treating the difficulties as a “running joke.”

“Most people I have told have been interested and curious about my experience, with a few exceptions, ranging from dubious to dismissive...My youngest brother was very dismissive and convinced this is just in my head.” (AF009)

Telling others about their difficulty was clearly not always an easy decision, one participant talked about “*coming out*” to her friend suggesting perhaps some sense of shame and also an awareness that she had been deliberately concealing their condition for some time. Another reported that she usually only told people she’d known for some time and who she felt would believe her. Even when others had tried to understand, several participants reported feeling misunderstood due to “*misplaced sympathy*”, by which they meant that others attempted to show understanding but, in so doing, demonstrated that they had failed to fully comprehend the difficulties the participant was describing:

“Often, people don't understand at first (and have not heard of it) and think I am saying that I forget names.” (DA022)

Theme 5: Appropriate support helps, but is often unavailable

This theme captures participants’ recommendations for things that others can do (rather than strategies that individuals with DP can try themselves) that make navigating life as a DP easier and experiences of requesting or receiving such support. This includes informal support. Although some DPs were able to suggest sources of information and advice they had personally found useful, it was notable how many participants were simply unable to recommend any sources, suggesting a real lack of information and support for the condition. This theme also captures the other side of the coin from theme 4, instances of others being supportive and understanding of DPs’ difficulties and how valued this is.

Sub theme 5.1 Support helps to navigate life with DP

Most participants described instances when others had been supportive and understanding of their difficulties and emphasise how valued such support was:

“There's this one colleague who understands and says "Hello, it's [his name]" every single time we meet - and every single time I am grateful. This is too awkward for most people.” (AF003)

“My colleagues are good are introducing other people or telling me before a person comes up who it is...” (DA013)

“My best friend...now goes out of her way to let me know if she has done something different to her hair when I visit.”(AF017)

“There is one meeting I regularly go to, in which after I mentioned it, the administrator in charge asked if it would be useful if she emailed me a chart of who is sitting where, each time. It is fantastic - by ten minutes into the meeting, I have an ascii art chart of who everyone is.” (AF003)

Sub theme 5.2 Reluctance to request or provide adaptations and accommodations

Although many of the suggestions such as name badges or introductions would be simple and low cost to implement, participants noted a frequent unwillingness (with notable, greatly-appreciated, exceptions) on the part of colleagues, organisations and others to implement helpful adjustments:

I have been known to ask for, or ask the facilitator of a meeting to do a round of introductions, even when in principle everyone knows one another, just to help me. (But there is always the odd colleague who just says "oh everyone knows ME"!) (AF003)

I asked for name badges at meetings but no-one else was keen... it seemed a bit unnecessary for everyone else as they all knew who each other was (I had no clue who most of them were!) (AF125)

“It was treated as a joke. I was too embarrassed to pursue it.” (DA017)

Analysis also indicated a reluctance on the part of some DPs themselves to request accommodations, for example at work:

“I have not had the confidence to ask for any specific adaptations in my corporate job, other than sometimes asking team members to introduce themselves to clients first, so that I can hear the client's name....I requested that [not-for-profit organisation] provided name badges in large capitals for the annual...conference, which they did. The request was received well. As an organisation founded on principles of equality, diversity and human rights, I felt comfortable making this request. I probably would have felt less comfortable if it were a commercial organization.”(DA001)

“It doesn't seem widespread enough to change how people operate. It's not a visible disability so people can't adapt to accommodate you.” (AF075)

Some participants negatively contrasted this absence of reasonable adjustments with accommodations they observe being routinely implemented for neurodivergent colleagues and people with visual impairments.

“It would help if 1) People wore name badges, 2) People addressed each other by name, 3) People told me their names every time they spoke to me. 3 would be most effective, but unlikely to happen in normal life. I have experienced it in group meetings where blind people are present, and it helped me a lot.” (AF019)

Subtheme 5.3 Increased awareness to reduce anxiety

Top of the list of things that participants felt would make life with DP easier were greater public and professional awareness and understanding and “recognition of the condition particularly in schools and health care settings:

“Training for teachers, health professionals etc. to mention prosopagnosia would be useful - especially with some idea of prevalence, because I think it's a lot more common than most people (with or without it!) realise, because most people tend to keep quiet about it or even not know there's a name for their difficulties.” (AF003)

“Training for health professionals, wider public awareness...If other people learned about it, so that we could just say that I might not recognise you if we meet again, as I have Prosopagnosia - and they would understand. If they knew that it's not because I don't care about them or find them important and that I don't have dementia.” (DA018)

“Get it recognised by HR professionals as on a par with neurodiversity and provide advice on how they can get managers to understand the condition and what people need at work - and that they are not stupid or rude when not recognising anyone.” (DA006)

Subtheme 5.4 Seek advice from experts

Although some participants were able to suggest sources of information and advice they had personally found useful (see Appendix 4, S5 for a complete list), it was notable that the majority (19/29) felt unable to recommend any sources, suggesting a real lack of information and support for the condition.

“The majority of current resources I’ve found are inadequate. Many do not chime with my personal experience.” (DA012)

Among the participants who did recommend sources they had personally found useful, the importance of seeking advice from others with particular expertise and knowledge around face blindness was highlighted. This could be informal peer support from individuals with lived experience of DP unspecified Facebook self-help groups as well as from DP researchers and voluntary organisations. Several participants mentioned the support organisation Faceblind UK, the research website Faceblind.org as useful sources of information and advice for individuals with DP and their families, however not all participants were aware of such resources.

“The Facebook groups. Reading other peoples’ situations that I recognise so well. It has been quite liberating. Until a few years ago, I didn’t know there was such a thing as Prosopagnosia! I always thought that it was “my own fault” somehow. I would definitely recommend these kind of groups.” (DA019)

“I possibly would recommend an official diagnosis for someone young, as this may help with coping strategies.” (AF009)

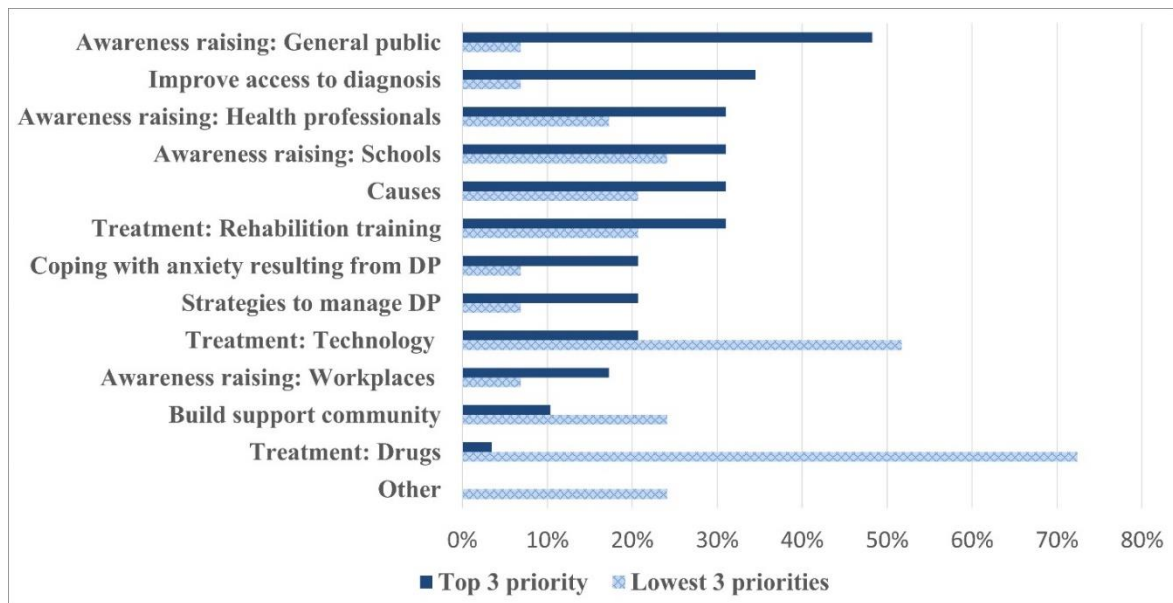
6.5 Research priorities

Finally, participants were asked to identify their priorities for future research. They first provided spontaneous suggestions and were then asked to prioritise their top three areas for future research from a list that participants had raised in screening interviews and that researchers drew from previous grey and published literature – or to select “Other” and provide their own (Figure 6.7 below). Participants were also asked which of the topics they considered to be low priority.

Participants' *unprompted* suggestions for future research included the association and co-occurrence of DP with other conditions and factors, for example navigation ability, object recognition, personality type, sexuality, Autism and ADHD and the ability to detect lying (see Appendix 4 table S4 for a full list). Most participants listed multiple avenues for future research they would like researchers to prioritise, only one participant did not feel further research was important, writing "*I don't think it's necessary. It's a minor quirk, causing some social embarrassment which can be countered with more public awareness*" (DA017).

However, few of the spontaneous suggestions were later rated by any participant as being one of their top three priorities (Figures 6.7). Improved awareness and understanding of DP among the general public, health and education professional and, to a lesser extent, workplaces were highlighted as important areas for future research, followed by improved access to diagnosis. The latter perhaps reflects the difficulties that many participants had experienced having their difficulties acknowledged and accepted (see Theme 4). Additional areas participants rated as important for future research included understanding the cause(s) of DP and developing training to help improve face recognition. A couple of participants noted that they had not selected certain topics, for example developing treatment to improve face recognition, due to a belief that improvement was not possible. Others noted that they felt all of the topics were important priorities and they had struggled to choose only three.

Figure 6.7 Participants' priorities for future research.



Note: Participants selected their top three and bottom three priorities for future DP research from a closed choice list, or provided alternatives priorities under “other”.

6.6 Discussion

Twenty-nine adults with DP completed an online survey comprising closed-ended questions describing their experience of living with poor face recognition, how this manifests in their day-to-day life, the psychosocial and practical consequences of their condition and their priorities for future research in this area. Findings were that the condition affected participants in quite different ways. Some participants were slow to learn faces, needing many encounters before having any chance of recognising the person again but once a face because ‘imprinted’ in memory it was recognised reasonably reliably quite well thereafter. Others were sometimes able to recognise faces after relatively few encounters, usually as a result of adopting effortful strategies, but could experience very rapid fading of face memories, after a matter of hours – or even minutes. Around half of participants reported that were unable to reliably recognise their three closest friends out of context, and around a third reported being unable to recognise immediate family thus highlighting that DP commonly affects the recognition of highly familiar faces with whom individuals have a close emotional relationships. Furthermore, participants who *could* and *could not* recognise their immediate family showed no significant difference in objective face memory ability.

6.6.1 Lived experiences of DP are extremely varied

The qualitative findings provided strong support for the idea that DP is a highly heterogeneous condition. Variability in the types of difficulties and errors that participants reported making could be due to bias or recall errors but cannot account for all heterogeneity since some participants reported directly opposing difficulties. For example, 13/29 participants said they made false alarms, i.e. mistook a stranger for someone they know, “very” or “quite” often, by contrast 3/29 said they *never* made this type of error, and such false alarms are easy to detect. Most participants said they forgot familiar faces that they had *previously been able to recognise* if they hadn’t seen them for a while, however 2/29 said this *never* happened to them and 7/29 reported that this happened only infrequently suggesting that in these latter cases the face memory traces, once formed, were relatively robust. In terms of wider difficulties, many participants relied heavily on voice to recognise others, suggesting (relatively) unimpaired voice recognition, but some found it difficult to recognise others by voice as well as by face. Together, the diverse range of types of errors, and the lack of consistent symptom presentation across the sample support the idea that DP is indeed a heterogeneous condition, and that the heterogeneity widely reported in the literature cannot solely be explained by methodological factors.

6.6.2 Face perception can be affected

Previous research has investigated whether certain aspects of face perception such as sex categorisation (Ariel & Sadeh, 1996; Esins et al., 2016; Jones & Tranel, 2001) are affected in DP. Several responses indicated that face perception can indeed also be affected in some cases of DP. For example, one participant described a difficulty categorising facial sex, another reported difficulty matching the face of a familiar colleague with the photo on their ID badge. Other participants described how groups of people appeared as ‘an amorphous blob with lots of heads’ or how “everyone merges into non people” suggesting a difficulty with face discrimination as well as face identity recognition. Although reports of face perception difficulties were not widespread, perhaps because the questions I asked primarily interrogated face identity recognition, it was interesting to note that three separate participants reported being extremely good at telling identical twins apart, even when their friends could not. One case of an ability to discriminate identical twins has previously been reported (Portch et al., 2023) and our

finding that 3/29 participants spontaneously reported this particular strength suggests this unusual ability might be more widespread in DP than previously recognised. This suggests that the atypical techniques they employ (e.g., memorizing features or mannerisms) are effective to some extent. It was not clear from the responses whether any of the participants was able to identify either of the twins if bumping into them unexpectedly, or whether these participants were only able to discriminate the twins when seen together and in context.

6.6.3 Theoretical implications

Consistent with previous research (Murray et al., 2018), many participants reported finding large gatherings or busy places to be the most challenging settings, and several participants described attempting to locate the person they were with as being like playing the visual search task “Where's Wally?” or described “losing” the face of their husband in a busy shop. Others mentioned a particular difficulty learning multiple new faces together, for example at a party. Together these findings suggest that a particular characteristic of DP is a difficulty learning or recognising faces when they are surrounded by other faces. However, the most widely used tests of face memory and face learning require participants to learn one face at a time, presented on an uncluttered background which may aid face learning. Similarly, at test, commonly used face recognition tasks typically use a single face (famous faces tests), two faces (old/new paradigms) or a three alternative paradigm e.g., Cambridge Face Memory Test (Dalrymple et al., 2012; Duchaine & Nakayama, 2006) which would not detect this specific difficulty. Two face matching paradigms, the Benton Face Recognition Task (Benton et al., 1983; Rossion & Michel, 2018) and the 1 in 10 task (Bruce et al., 1999) do require participants to identify a target face from a choice of either 6 or 10, but these tests are relatively rarely used in the DP literature (Nørkær et al., 2024). Interestingly a recent study identified the Benton Face Recognition Task as particularly sensitive for detection of DP (Mishra et al., 2021) and our findings suggest that this may be because the task taps this specific difficulty that appears to be fairly widespread among DPs.

A particular issue that has been highlighted by several recent studies (Bate, Bennetts, et al., 2019; Burns, 2024; Burns et al., 2022; Lowes et al., 2024a) is the high proportion of DPs who presented to research groups yet appeared to show no objective face recognition difficulties. It has been suggested that using a limited number of tasks

which potentially lack ecological validity (Burns, 2023a; Burns, 2024; Ramon et al., 2019) for classification of DP may miss particular types of difficulty and our findings provide further possible explanations for this. For example, the use of an “emergency recognise” whereby a feature memorisation technique allows a memory to be created that can last a few hours before rapidly decaying is a strategy that could allow that participant to achieve what appears to be normal face recognition scores on tests of short-term face memory such as an old new task or the CFMT.

Interestingly one participant perceived that she was better able to recognise faces as familiar at certain phases in her menstrual cycle and that during pregnancy she felt like she had a facial recognition “superpower.” Oxytocin is a neuropeptide whose levels fluctuate during menstrual cycle (Engel et al., 2019) and can increase three to four fold during pregnancy (Uvnäs-Moberg et al., 2019). Face recognition ability in autism (Skuse et al., 2014) and DP (Cattaneo et al., 2016) have been associated with variations in the oxytocin receptor gene. Intra nasal inhalation of oxytocin has been reported to improve performance on face tests in people with typical face recognition (Rimmele et al., 2009; Savaskan et al., 2008) and, temporarily, in individuals with DP (Bate et al., 2014).

6.6.4 Quality of life and workplace experience – a case for DP to be included in the definition of neurodivergence.

With a small number of exceptions, our participants considered DP to be a lifelong condition that causes significant adverse impacts on their social and work interactions. In their personal lives participants developed bespoke ways of navigating personal and social situations, but many reported that interacting with others was a source of anxiety or even panic. Raising awareness in the general population was one of the highest research priorities listed by our participants. Campaigns raising awareness of the condition may thus be helpful in alleviating stigma or perceptions of ‘rudeness’ by people who, in fact, simply fail to recognise others. In work settings, participants struggled remembering colleagues and clients, large meetings were problematic, and some participants reported dismissive attitudes from their employers. DPs’ difficulties were particularly highlighted when work demands made it impossible to engage with compensatory strategies. The DPs’ pattern of ‘masking’ failures of face recognition strikes us as resembling masking reported by participants with Autism (Miller et al., 2021) or ADHD (Kosaka et al., 2019 and Watson et al., under review). Masking is an exhausting and costly cognitive effort, often leading to negative mental-health

consequences (Bradley et al., 2021; Chapman et al., 2022) and burnout (Arnold et al., 2023; Raymaker et al., 2020).

In contrast, participants reported significant easing of pressure related to their DP by requesting adjustments such as large, capitalised name badges, or by some colleagues spontaneously helping them by repeatedly introducing themselves, or by secretaries sending them a seating plan with names. Crucially, while some participants wished that DP were recognised on par with other types of neurodivergence such as Autism and ADHD, recognised in the UK by The Equality Act 2010, only one (DA001) reported that they felt comfortable requesting workplace adjustments, and even then, not in their main corporate role – only in a non-for-profit organisation which espoused strong inclusivity values. It is thus clear that DP can disadvantage people in work settings (see also Fine, 2012) and that people with DP, should they wish to disclose it, should be afforded reasonable adjustments and be protected from possible discrimination. ACAS (2024), the UK-based independent (but government-funded) public body providing advice on employment rights and best practice and policies includes Autism, ADHD, dyslexia and dyspraxia in their neurodiversity definition. I believe that DP should be included in this, and similar, list of forms of neurodivergence owing to its high occurrence as a standalone condition affecting 2-4 % of the population and potentially as high as 8%, (for a review see DeGutis et al., 2023) and frequent co-occurrence with Autism (Minio-Paluello et al., 2020). While not all DPs may opt for asking for workplace adjustments, some would clearly benefit from reasonably easily implemented changes to practices.

6.6.5 Limitations and directions for future research

A limitation of this study is its reliance on self-report which may be subject to bias. Where possible I attempted to mitigate this, for example by cross checking responses with objective test data. Thus, the participant who reported difficulties remembering the sex of a person did indeed score significantly below average on a task of gender categorisation. However, relevant objective data were not always available.

The findings show that individuals with DP have a particular difficulty learning, and recognising, faces when multiple faces are encountered at the same time. A promising avenue for future research could be to investigate whether a face recognition paradigm that presents the target face alongside distractors, and tests recognition among a higher number of distractor face might be a more sensitive measure for detecting DP than current tests. The findings also

Chapter 7: General discussion

7.1 Summary of chapters

In this thesis, I investigated objective and subjective experiences of developmental prosopagnosia (DP). To assess objective abilities, participants first completed an initial cognitive screening battery which including a matrix reasoning task the MaRs-IB (Chierchia et al., 2019). The key finding from the screening battery (chapter 3) was that DPs adopted different speed accuracy trade-offs to typical controls and therefore future research must consider reaction times (RTs), as well as accuracy when classifying DP. This trade-off was observed for a fluid reasoning task, a bicycle memory task (chapter 3), and face memory tasks (chapter 4).

Chapter 4 investigated which face memory measures, or combination of measures, best discriminated 24 participants with self-reported face recognition difficulties from typical perceivers. Participants completed the CFMT, a test of novel face recognition, a matched bicycle recognition task (CBMT), an old new faces recognition task with a 2-minute delay between study and test (Old New Faces) and a famous face recognition test (FFT). To control for the group differences in speed-accuracy trade-offs which were observed in Chapter 3, I decided to directly compare how well accuracy and balanced integration scores (BIS), calculated for the *same* tasks completed by the *same* participants, could classify DP. Logistic regression results showed that individual BIS measures were more sensitive than accuracy alone correctly classifying 68.2% vs 59.1% respectively of potential DPs with minimal difference in specificity (BIS measures correctly classified 94.9% of typical controls vs 97.4% for accuracy measures). Scores were averaged across the four face measures to compute global face memory scores (BIS and accuracy). At individual level, global face memory BIS classified a higher proportion of participants with subjective face recognition difficulties than global face memory accuracy (83% vs 58% respectively) and produced larger between group effect sizes ($d = -1.78$ vs -1.36 for global face memory BIS and accuracy respectively). Self-reported face recognition ability (measured by PI-20, Shah et al., 2015 or an adapted version of the above in the case of children) was a significant predictor of objective global face memory BIS ($p < .001$), explaining around 38% - 50% of the variance.

Having classified 20 individuals with DP (Chapter 4), I next systematically tested multiple aspects of face perception in these individuals and 80 age-matched controls to determine which stages of face processing are impaired in DP (Chapter 5). An important pre-registered research question was whether the underpinning deficits are the same in all cases of DP. Single case analysis, group analysis and intraclass correlations showed that face perception ability (and face processing more generally) is highly heterogeneous in DP. The stages at which most DPs showed impairment were face detection and face identity perception which were affected in 50% of participants, though importantly not always the same 50%. Inspection of results at individual case level showed evidence for two sub types of DP; a *mnemonic* subtype (in which face memory is impaired, but face perception is within typical range ($> -1 SD$)) and an *apperceptive* subtype, in which both face perception and face memory are impaired. Most apperceptive cases showed impairment on at least two separate face perception tasks, although a small number ($n = 3$) showed impairment *only* on the CFPT, thus providing tentative evidence for a possible third subtype of DP, affecting face *identity* perception and face memory, but not earlier stages of face perception.

Finally, I explored subjective experiences of living with DP (Chapter 6) using an online mixed methods survey. Responses provided further evidence of the heterogeneity of the condition and highlighted the types of difficulties that DPs ($n = 29$) experience in everyday life. Two notable findings were first that DPs found large groups a particularly challenging setting, both from the perspective of learning new faces and of recognising previously encountered faces. A second striking finding was that being unable to recognise even close friends and family is a common experience in DP. It has sometimes been proposed that difficulties recognising immediate family and closest friends represent “severe cases” of prosopagnosia (Burns, 2023b; Cook & Biotti, 2016) however chapter 6 suggests that this experience is widespread. Furthermore, a comparison of the CFMT scores of the subgroups of DPs who could – and could not- reliably recognise immediate family showed they did not differ and thus lower CFMT scores were not associated with the likelihood of being able to recognise one's nearest and dearest.

7.2 The present chapter

The findings of each empirical chapter (chapters three, four, five and six) are discussed at the end of the relevant chapter so the present general discussion chapter considers additional insights and implications for future research and policy.

7.3 Individuals with DP show metacognitive insight

7.3.1 DPs have insight into their difficulties, even when unaware of DP as a condition

Results from the objective test battery (chapters four and five) showed strong evidence that individuals with DP are aware that they have severe problems recognising faces. Subjective difficulties, as measured by scores on a standardised self-report Prosopagnosia Index questionnaire (PI-20, Shah et al., 2015) were strong predictors of objective face recognition ability across four separate different face memory tests. Despite participants' awareness that they struggled with face recognition in everyday life, survey responses (chapter 6) highlighted that the DPs often did not realise that their difficulties were the result of a recognised condition, sometimes even until late adulthood. This finding is consistent with previous anecdotal reports (e.g., Diaz, 2008; Fine, 2012) and a larger scale study by Murray et al. (2018). However, Murray and colleagues argued that not becoming aware of the name of the condition until adulthood was evidence that DPs have "limited insight into their difficulties" whereas this thesis suggests an alternative interpretation, namely that participants had limited insight into the existence of DP as a *condition*, or an explanation for their difficulties, but nonetheless all were aware of the fact that they struggled greatly to recognise faces. In some cases, this realisation came at an early age, including while at still at school. Further, the qualitative results highlighted how others' dismissive or disbelieving responses (see also Burns, 2023b) often caused participants to doubt their own belief that they had a genuine face recognition difficulty. The commonly reported downplaying or outright disbelieving of the considerable difficulties that DPs experience in everyday life could conceivably contribute to the late diagnosis/awareness of DP that has been widely reported here and by others (Murray et al., 2018).

7.3.2 PI-20 scores predicted objective face memory scores

The question of whether DPs *have* (Burns, 2023; Burns et al., 2022; Gray et al., 2017; Livingston & Shah, 2018; Stollhoff et al., 2011; Tsantani et al., 2021; Ventura et al., 2018) or *do not* have (DeGutis, Bahierathan, et al., 2023; Murray & Bate, 2019; Palermo et al., 2017) good metacognitive insight into their face recognition abilities has been a matter of some debate in the literature. Chapter 4 results provided evidence that DPs do have insight into their face recognition difficulties and that these difficulties appear to be well captured by the PI-20 questionnaire (Shah et al., 2015, or an adapted parental version in the case of children with poor face recognition). Linear regression modelling showed that self-report scores significantly predicted objective face memory ability (for both accuracy and BIS, both $ps < .001$), as operationalised by an age-standardised global face memory score that was calculated by averaging performance on: i) a test of novel face learning (one of three age-appropriate version of the CFMT, Dalrymple, Garrido, et al., 2014; Dalrymple & Palermo, 2016; Duchaine & Nakayama, 2006)), ii) an old new faces task with approximately two minute delay between learning and test (Old New Faces, Dalrymple et al., 2014); iii) a famous faces test measuring long term face memory (Lowes et al., 2024a) and iv) a contrast measure which is the difference between face (CFMT) and bicycle recognition ability (CBMT, Dalrymple, Garrido, et al., 2014).

The findings reported in chapter 4, emphasise the need to consider both RT and accuracy when assessing face recognition ability in DP. This approach may help resolve the disagreements about DPs' metacognitive abilities. To date, research has mainly focused on whether self-report, using the standardised PI-20 questionnaire (Shah, Gaule, Sowden, et al., 2015) or a similar instrument such as the Cambridge Face Memory Questionnaire (CFMQ, Arizpe et al., 2019), is related to face memory. This literature has focused most notably on discussions of the relationship, if any, between self-report and *accuracy* on the CFMT, a task which is currently considered to be the “gold standard” (Mishra et al., 2021; Murray & Bate, 2020; Stantic et al., 2022) for detecting prosopagnosia. This thesis shows that accuracy alone is a less sensitive measure than balanced integration score (BIS), which adjusts accuracy to take account of RTs and it controls for the different speed-accuracy trade-off strategies that were observed between DPs and typical controls (Chapters 2 and 3). Therefore, self-report may show a stronger

relationship with a measure like BIS, or with a more robust global score computed from the average of several face memory measures, than with CFMT accuracy alone. As discussed above, although chapter 4 shows that self-reported scores significantly predicted CFMT accuracy ($p < .001$), explaining 17% - 25% of the variability in CFMT accuracy scores, it also showed that self-report explained a far higher proportion of variability (44% to 57%) on the more sensitive *face memory BIS* individual measures that also consider RT.

Why then might the global face memory BIS measure show a closer relationship with subjective report than CFMT scores, aside from better controlling for the speed-accuracy strategy differences (discussed in chapters three and four) that can mask impairment? Major advantages of the CFMT are that the stimuli are tightly controlled, hair is obscured, and all the identities are initially novel thus controlling for differences in pre-test familiarity. However, CFMT tests the ability to recognise *newly-learned* faces over a *short* retention interval, whereas the self-report questionnaires interrogate *long-term* recognition of *familiar* faces. The latter is arguably better tested using paradigms involving familiar faces (such as FFT) and/or a face recognition task with a longer retention interval between learning and test than that CFMT (for example the revised Old New Faces test used here). It has previously been shown that that memory for newly-learned (CFMT) and famous faces was dissociable in 63/165 adults with DP (Bate, Bennetts, et al., 2019). It is possible then that some DPs can learn and retain faces in memory over a relatively *short* duration, thus achieving typical CFMT accuracy scores, but that their impairment is in *longer term* face recognition which is not tapped by the standard version of the CFMT used here (though see CFMT-Aus (McKone et al., 2017) which has a 24 hour test condition). Consistent with this idea, Pertzov et al. (2020) concluded that DP is characterised by *rapid forgetting* of upright faces and that DPs showed greater recognition impairment vs controls when tested at 6 s vs 1 s (the only retention intervals tested). Similarly, work that tested memory for novel faces after one year found that shorter term and longer term face memory were dissociable subtypes of DP (Stollhoff et al., 2011) and concluded that DP should be assessed using measures that specifically tap long-term face recognition memory. The findings of Chapter 4 support these recommendations.

This thesis allowed DPs' metacognition to be tested in an additional novel way. Rich qualitative and extensive quantitative data were collected from the same participants meaning that subjective reports could be directly compared with objective scores at an individual level. Turning to the qualitative data (chapter 6), to help answer whether some cases of DP can be explained by rapid forgetting, whereas others appear to be underpinned by a difficulty learning novel faces (needing to sample faces more often or for longer than average in order to later recognise them), survey responses supported this idea that *some* individuals with DP are able to learn faces without too much difficulty but that their memory traces fade rapidly (Pertzov et al., 2020). In their qualitative accounts of living with DP (chapter 6), some individuals spontaneously described how they were able to recognise some, though by no means all, faces quite reliably once some kind of "threshold" was passed, and one described how, if needed, they could force themselves to do an "emergency recognise" and hold a face in memory for a couple of hours. In this case, and most others, the ability to recognise a particular face usually did not last long: 10/29 participants reported that they forgot faces they had previously been able to recognise "very often" and 12/29 said this happened "quite often", suggesting the ability to form an accurate facial percept at least some of the time, and ability to maintain and access that facial representation memory in the short, but not long, term. This could point to issues with consolidating short-term memories into long-term storage and implies a particular deficit in the processing involved in stabilising and storing facial memories over extended periods.

Taken together, thesis results suggest that in *some* cases of DP the impairment is primarily in long-term face recognition memory. If so, impairment may not be detected by the CFMT or similar paradigms which require only very short retention intervals yet the CFMT has often been used as the sole test of face memory (e.g., Bylemans et al., 2020; Gerlach et al., 2020; Tsantani et al., 2020; Tsantani & Cook, 2020). One important consequence of the possibility that *long term* face memory may be the main deficit in some DPs is that those participants would be excluded by DP studies which use conservative DP inclusion criteria e.g. CFMT *accuracy* scores < -2 SD below control mean since the CFMT tests short term face memory. This is despite the fact all individuals in the studies presented here showed clear impairments on ≥ 2 objective face memory measures (CFMT, CMT difference, Old New Faces or FFT) though not necessarily on

CFMT accuracy. To address this, the current guidelines (Dalrymple & Palermo, 2016) should be updated to recommend DP categorisation using a combination of self-report (e.g. PI-20) and two objective face memory tests, including at least one measure of *long term* face memory.

Although the PI-20 may have some redundant items (Nørkær et al., 2023) and more precise wording of some items might improve sensitivity even further (discussed below in the subsection 7.4 *DP is characterised by the frequency and type of recognition errors*) overall, the objective and subjective findings in this thesis support the view that the PI-20 is a good measure of prosopagnosic traits in line with previous work (Burns, 2023; Burns et al., 2022; Gray et al., 2017; Livingston & Shah, 2018; Tsantani et al., 2021; Ventura et al., 2018). Furthermore, the findings strongly suggest that the PI-20 has strong construct validity.

7.4 DP is characterised by the frequency and type of recognition errors

Of course, typical perceivers can also make various types of face recognition errors from time to time, but this thesis provides evidence that what appears to differentiate DPs from the typical population are the *type* and *frequency* of the errors made. When typical perceivers report that they did not recognise someone they know, this can often mean they did not *immediately* and effortlessly recognise the person, rather than that recognition did not occur at all. In Young et al.'s seminal diary study (1985) more than half (58% - 65%) of recognition failures were in fact resolved by the diarist spontaneously recalling who the person was, i.e. without being told or using a non-facial cue. Although some instances of spontaneous, if delayed, familiar face recognition have recently been reported in DP (Portch et al., 2023), the authors concluded that such instances were typically the result of compensatory strategies, rather than true spontaneous recognition and the survey data (chapter 6) are consistent with their interpretation. The examples of recognition failures described by the survey participants were rarely, if ever, resolved by spontaneous recognition. Instead, despite their use of elaborate strategies and heavy reliance on non-face cues such as hair, gait, distinguishing features, or context, it was nonetheless common for participants to simply have no idea who the person they had failed to recognise was.

Notwithstanding the frequent, complete failures of face identity recognition, the thesis results suggest that face identity recognition in DP is not an “all-or-nothing” ability. Consistent with Portch et al.’s finding (2023) of *some* degree of preserved recognition in DP, survey responses (chapter 6) showed that most of the DP participants reported that they were able to recognise a *small number* of individuals out of context. The median number of faces participants estimated that they could reliably recognise was seven, with two thirds saying they could recognise fewer than ten faces. These estimates, whilst undoubtedly imprecise, are nevertheless far below the estimated average of 5000 faces that typical adults can recognise (Jenkins et al., 2018). Although this latter estimate is likely also imprecise, the vast difference in these ballpark estimates is striking, suggesting that true ranges (DPs vs controls) are indeed very different. Notably, the estimates provided by DPs were informed by their real-life experience and, although participants were asked to make estimates based on recognition of the face alone, in practice it is difficult to know when converging non-facial identity cues are also being drawn upon. Outside laboratory, cues such as body shape, hair and gait would often be available to participants in addition to faces. As such, DPs may have overestimated their ability to recognise familiar people based on their faces alone meaning that the *true* number of faces that DPs can typically recognise may in fact be even lower than the estimates provided. However, access to non-facial cues in real world settings was *still* insufficient to enable participants to recognise even highly familiar others.

7.4.1 Inability to recognise immediate family and close friends is common in DP

Previous studies (e.g., Dalrymple, Fletcher, et al., 2014; Yardley et al., 2008) have noted that DPs can have difficulty even recognising family and close friends. This thesis provides the first estimates of how prevalent such difficulties may be in DP. Findings were that, excluding those who said they did not know, almost half (48%) reported being unable to recognise their three closest friends and over one third (35.7%) said they could not always recognise immediate family members. Although control data are not available for comparison, the inability to recognise the faces of close family (parent, sibling, child, or partner) and friends who are highly familiar is extremely uncommon among typical perceivers, and this may therefore be one, albeit not universal, characteristic of DP.

It may be possible to detect this characteristic through minor revision of the widely used PI-20 questionnaire (Shah, Gaule, Sowden, et al., 2015). Item 18 of the PI-20 states “at family gatherings I sometimes confuse individual family members.” However, the thesis findings suggest that an item that enquires directly about recognition of *immediate* family may be more indicative of DP than the present wording because an emphasis on *immediate* family would differentiate individuals who do not recognise siblings, partners or children from others who might have initially struggled to recognise a long-lost cousin or a great aunt they may previously have only met once or twice at a large wedding. The latter may be something that typical perceivers also find difficult and thus this item may not discriminate DP. In sum, it appears that certain *types* of face recognition errors involving close friends, family and colleagues one sees regularly, and the *frequency* of these errors in everyday life, may be particularly characteristic of DP and could form a useful future line of research.

7.5 The Underpinnings of DP are heterogeneous

7.5.1 Inter and intra individual differences in behavioural patterns of face perception and face memory

The finding that DP is heterogeneous is not new (for a review see Corrow et al., 2016) but this thesis is the first to systematically analyse behavioural performance across such a large number of cognitive face processing measures at both individual and group level. Additionally, it tests the heterogeneity claim across these multiple measures at group level using intraclass correlation. The choice of tasks to be included in the neurocognitive battery has a strong theoretical basis and was informed by a systematic review of tests of upright face processing (Robotham & Starrfelt, 2018) that categorised the many face processing paradigms in the literature according to eight levels of processing, or subprocesses. Robotham and Starrfelt (2018) additionally mapped the paradigms onto the three main stages of Bruce and Young’s model (Bruce & Young, 1986; Young & Bruce, 2011), namely face *perception*/encoding, followed by face *recognition* leading to *identification*. Although different patterns of deficits were observed across the DP sample, these broadly clustered into two sub types, one *apperceptive* subtype characterised by widespread deficits in face perception in addition to impairment in face memory and a second *mnemonic* subtype affecting only face

memory. Of course, face memory paradigms such as the CFMT also require encoding of faces and are not therefore pure measures of face memory. However, if participants show typical performance on multiple face perception tasks, and are impaired only on face *memory* tasks, this indicates that the deficit is primarily in memory for faces.

Recent work using eye tracking methods (Peterson et al., 2019) also provides preliminary evidence of two distinct phenotypes of DP similar to those identified through behavioural methods here; one affecting face memory only (associated with looking more at the upper face), and a second larger subgroup with impaired face perception *and* face memory that was associated with looking lower on the face. Neurotypical controls and those DPs who looked high on the face performed best when faces were presented near their preferred fixation location, but this was not true for lower-looking DPs. Lower-looking DPs did not benefit from their gaze patterns in the same way, leading to poorer face processing. Peterson et al. (2019) posit that in cases of apperceptive DP, the deficit might not be due to how DPs look at faces, or how their visual system processes faces in general, but rather is driven by a misalignment between where they are looking and where their face processing is most effective, they call this the Mismatched Tuning Hypothesis.

7.5.2 Caution needed when generalising from the CFPT.

The CFPT (Duchaine, Germine, et al., 2007) is the probably the most commonly used test of face perception in DP and has been used in 54 separate studies of DP (Nørkær et al., 2024). Participants must sort six faces according to their resemblance to a target face shown at the top of the screen and thus measures *face identity perception*. Although some studies have interpreted CFPT performance as indicative of face *perception* ability more broadly (e.g. Bate et al., 2019; Dalrymple, Garrido, et al., 2014), this thesis casts doubt on whether such generalisations are possible based on a single test. It presents converging evidence that different perceptual sub processes can be affected in different sub types of DP (chapter 6). Importantly this held true at both group and individual level. While CFPT performance was at least mildly impaired in 10/20 DPs, a similar proportion to previous reports e.g. Dalrymple, Garrido, et al.. (2014), notably 30% of participants (3/10) were *only* impaired on the CFPT and showed typical performance on all the other five face perception tasks despite clear face memory deficits. These results demonstrate that CFPT performance cannot be generalised to make broader claims about other sub processes of face perception.

Evidence that different perceptual sub processes have distinct neural correlates (for an overview see Duchaine & Yovel, 2015; Manippa et al., 2023)) means that it is possible any of these might theoretically be selectively impaired in DP. Theoretically, the differential patterns of perceptual deficits observed are consistent with the neuronal migration error account of DP (Ramus, 2004), which posits that neurons in the developing brain migrate abnormally and can therefore end up in the wrong place. Crucially, the neurons can end up in slightly different places in different individuals and therefore the structural abnormalities can also differ between individuals, resulting in heterogeneous neurodevelopmental disorders. The findings presented here highlight that future work investigating face perception should test more than one stage of face perception before generalising findings from one perceptual task to infer ability on other aspects of face perception. Robotham and Starrfelt's (2018) levels of processing framework appears useful in this regard.

7.5.3 Inter individual differences in subjective experiences of face perception and face memory

DPs' qualitative descriptions of their face processing difficulties provide novel converging evidence that certain stages of face perception can be differentially affected in DP. At one end of the perceptual ability spectrum were three participants (10% of the sample) who each spontaneously reported a surprising ability to individuate identical twins. Although they may achieve this through the use of atypical processing strategies, for example by focusing on individual face features or characteristic facial movements (we do not know if they were able to discriminate the twins from photos), this ability nevertheless suggesting that these individuals are able to differentiate people whose faces are very similar. By contrast, apparently at the other end of the face perception ability spectrum were other participants who told how all faces looked the same to them; one described how looking at a group of people was like seeing "a big amorphous blob with lots of heads" suggesting a severe difficulty individuating physically dissimilar faces. Subjective accounts therefore also point to heterogeneous perceptual abilities and difficulties within the DP sample.

Descriptions of other aspects of face processing difficulties also revealed intra-individual differences. For example, several participants reported that faces often looked familiar but that they were unable to identify the individual. The sense of familiarity allowed them

to *pretend* to recognise the person while keeping the conversation sufficiently vague to avoid disclosing their inability to identify them, or by stringing the conversation along in the hope that contextual, voice, or conversational cues would eventually lead to identification. In contrast, others said that they got no such sense of familiarity, in one example even when unexpectedly bumping into a colleague whom they worked closely with five days a week. Another participant noted that “most people feel slighted on second meeting because I will ignore them and not know I've previously met them,” also indicating a lack of familiarity signal. Previous work has reported that DPs have “intact familiarity recognition” (Avidan & Behrmann, 2008; Bate et al., 2008; Stumps et al., 2020) but these thesis results suggests that, if so, findings from more controlled laboratory settings may not generalise to the real world. Alternatively, it may be that some, but not all, DPs have intact recognition familiarity (Bentin et al., 2007) and the relatively small sample size typically found in DP research (32 DPs in the Stumps study) are not representative of the wider DP population.

7.6 The use of group level statistics alone to characterise DP can be misleading

This thesis addressed limitations of previous studies that either analysed results at group level only (Bennetts et al., 2024; Biotti et al., 2019), involved a very small number of participants, (Ulrich et al., 2017), used only one or two perceptual tests (Dalrymple, Garrido, et al., 2014; Dalrymple & Duchaine, 2016) or tested a single type of perceptual ability e.g. face matching (Mishra et al., 2021) by thoroughly investigating face perception at an individual and at group level. Further, it used qualitative and behavioural quantitative approaches and a range of pre-registered statistical methods to seek converging evidence for or against perceptual impairment in DP.

Results show that analysing results at individual level *in addition to* group level can generate valuable insights in neuropsychological and special populations. In studies with special populations, sample sizes are often small and a small number of extreme scores can disproportionately influence groups means thus creating a potentially misleading picture (e.g. a study of super recognisers Phillips et al., 2018). For example, on the CFMT, at group level, after correction for multiple comparison, DPs were significantly less accurate than typical controls. Superficially, the group difference for

CFMT BIS ($p < .001$, $d = -1.03$) looks very similar to CFMT accuracy ($p < .001$, effect size $d = -1.08$), yet at individual level clear differences emerged. Here 12/24 participants showed a major impairment (< -1.7 SD) on CFMT BIS vs only 5/24 on CFMT accuracy (chapter 4).

Similarly, significant group differences were observed on the faces among non faces task, the CFPT, gender categorisation and configural featural faces task but the *proportion* of individual DPs who showed major impairment on these tasks ranged from 35% (configural featural faces BIS) to 50% (faces among non faces RT and CFPT accuracy). Different DPs showed different impairment and participants' overall perceptual ability only became clear when assessing results across the entire perceptual battery.

If then, as the results of this thesis appear to confirm, DP is indeed a heterogeneous condition, then the search for a measure that shows robust and replicable *group* differences that differentiate DPs from typical perceivers may be the wrong approach because the presence, or absence of, significant group differences will depend in large part on the make-up of any given DP sample. Inspection of results at *individual case level*, in addition to group level, can help to identify measures that can better differentiate heterogeneous DP samples from typical controls. Further, as others have suggested, investigations of DP should employ multiple tasks (Bennetts et al., 2024; Biotti et al., 2019) and state clearly which particular stage of face processing is being investigated (Robotham & Starrfelt, 2018).

7.7 Other neurodevelopmental conditions also show spiky cognitive profiles

Accepting that DP is a heterogeneous condition, with different aspects, may be considered problematic by some since there is currently no genetic test or biomarker for the condition and thus diagnosis, or classification, is usually made on the basis of subjective report and below average performance on one or more behavioural tests (chapter 4). Heterogeneous performance makes the choice of test(s), measures and cut offs, challenging and at times controversial (see Burns, 2023a; Burns et al., 2022; DeGutis et al., 2023; Gerlach et al., 2024). It would be desirable to identify a sensitive test that can identify DP in all, or at least most, cases. However, such an aim may not be

possible. Heterogeneous behavioural profiles are common, and not considered an artefact of behavioural testing, in many other neurodevelopmental conditions (Idei & Yamashita, 2024) including those identified by genetic tests such as Prader Willi-Syndrome (Copet et al., 2010; Louveau et al., 2023; Whittington et al., 2004) and Williams Syndrome (WS). WS is a genetic condition that presents a spiky visuospatial cognitive profile, with severe impairment in some aspects of spatial representation but relative strength in others (Landau & Ferrara, 2013) illustrating that even in neurodevelopmental conditions with clear genetic origins (meaning there is no need to rely on behavioural tests for diagnosis), behavioural performance in a single domain can be highly heterogeneous. This thesis provides evidence that the same is true in DP.

7.8 Limitations

This thesis presents an in-depth analysis of the underpinnings of DP from multiple perspectives. However, despite collecting data from testing a total of 41 participants with severe face recognition difficulties¹⁴, the sample size of each study is still relatively small, which limits the generalizability of the findings given the clear evidence that DP is a heterogeneous condition. In chapters three, four and five, 24 participants with self-reported DP completed the objective test battery; of these 12 eligible participants, that is adults with confirmed objective face recognition difficulties living in the UK, agreed to complete the mixed method survey alongside an independent sample of 17 DPs recruited from Prof. Brad Duchaine's database. To mitigate the small sample sizes which can lead to underpowered studies, Bayesian analysis (involving 5000 iterations) was used in addition to frequentist methods, and this provided an assessment of the strength of the evidence for or against each of the pre-registered hypotheses in chapters three, four and five. Very few ($n = 4$) of the participants in this thesis had previously been tested by other research groups meaning that this is largely a new sample of DPs, which addresses a more general limitation of the DP literature namely that the same sample is often repeatedly tested and the same individuals reported in multiple papers (without authors clearly stating that this is the case, allowing for linking individual profiles). Recruiting DPs into research is challenging, but future research should always specify *which* participants, if

¹⁴ 24 in chapter 4, of which 20 were retained in the analyses in chapter 5. 12 of the 20 also completed the survey (chapter 6) in addition to 17 new participants.

any, have previously been reported in other studies and give a clear explanation of how any subsequent sub samples have been selected, to avoid the appearance of fishing for results or hypothesising after results were known.

A further limitation was the small number of children I was able to recruit. I had hoped to recruit 12 child participants but, despite strong initial interest from parents, a high proportion of the children were unfortunately ineligible to participate due to co-occurring conditions or visual impairment or were unaware of their difficulties leading to ethical concerns. I am therefore unable to say anything about the developmental nature of DP, which remains a major gap in the literature. One positive indication for such future research is that parents of children as young as four years of age contacted me, suggesting that general awareness of face recognition difficulties and DP may be increasing.

7.9 Future directions

7.9.1 Where in development does DP occur?

As already discussed, little is known about the developmental nature of DP, and indeed whether it is present from birth, and this is an important future line of research. One challenge of detecting DP in young children is the widely reported variability of face processing ability in typically developing children (Bennetts, Murray, et al., 2017; Dalrymple, 2012; Dalrymple, Garrido, et al., 2014). One way to address this would be to use non-invasive methods such as eye tracking and pupillometry. Eye tracking was one of the methods I had originally intended to use but unfortunately COVID-19 restrictions prevented in person data collection during the first two years of my research when the majority of data were collected. To the best of my knowledge there have been no longitudinal studies of face processing development beginning in early childhood and this could also be a very fruitful research direction that could help to elucidate the developmental nature of DP.

7.9.2 Individualised training

The neuropsychological test battery used, combined with rich qualitative accounts of the nature of face processing difficulties in DP, provide fine grained information at an individual level about the underpinnings of DP. In future these findings can be used to

develop personalised training that targets particular difficulties and also builds on the relative strengths of participants.

7.9.3 Awareness and changes to public policy (priority for this cohort of DPs) has potential to attenuate the difficulties DPs experience

A new avenue of research, and work, is around measuring and raising awareness of DP in the general population, as well as among health and education professions. This thesis is the first research to ask individuals with DP about their priorities for future research. Individuals living with the condition in the UK rated improved awareness, followed by training to improve face recognition, as the highest priorities for future DP research. However, survey participants were not a representative sample of all individuals with DP so the priority setting exercise would benefit from being extended to a larger sample from other research groups.

7.9.4 Other future directions

This thesis generated several other future testable hypotheses. For example, is long term face memory impairment one potential sub type of DP? Greater use of delayed face recognition tasks in DP research could help to elucidate this. Second, it could be interesting to test whether the differing subjective reports of the presence - or absence - of face familiarity signals could be detected objectively, for example using EEG. Third, given the high observed prevalence of early-stage face processing impairments in this cohort, for example face detection and gender categorisation 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. More generally, future work could compare the nature of the underpinnings of face processing impairments in individuals with “pure” DP as reported here with those observed in individuals with DP and co-occurring neurodevelopmental conditions such as autism.

8.0 Conclusion

Due to different speed accuracy trade-offs between DPs and typical controls I found that BIS, an integrated measure that adjusts accuracy for response time, was a more sensitive measure for classifying DPs than accuracy alone. This finding has important applied value for researchers who must identify and classify DP; it can allow a better

understanding of the underpinnings of DP and avoid unnecessary exclusion of participants whose impairments may be masked by speed-accuracy trade-offs.

Face processing comprises many separate subprocesses and the precise underpinnings of DP varied between individual cases. I showed 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 categorising facial gender suggesting early-stage perceptual difficulties. The DP and control groups both showed clear intra and inter individual differences across tests. Two subtypes of DP were identified. The largest DP subgroup showed widespread face perception *and* face memory deficits (the *apperceptive* subtype). A second subgroup showed fairly typical face perception yet a distinct face memory deficit (the *mnemonic* subtype). 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. I therefore recommend the use of multiple tasks assessing different stages of face perception as well as group and individual level analysis.

With a small number of exceptions, participants considered DP to be a lifelong condition that caused significant adverse impacts on their social and work interactions. Strikingly, one third could not reliably recognise faces of immediate family members and over 40% could not recognise their three closest friends highlighting that DP commonly impairs the ability to recognise highly familiar faces. In their personal lives participants developed bespoke ways of navigating personal and social situations, but many reported that interacting with others was a source of anxiety or even panic. It is clear that DP can also disadvantage people in work settings and that people with DP, should they wish to disclose it, should be afforded reasonable adjustments and be protected from possible discrimination. I believe that DP should be recognised as a form of neurodivergence by employers and governments. While not all DPs may opt for asking for workplace adjustments, some would clearly benefit from reasonably easily implemented changes to practices. Raising awareness of DP was the highest research priority listed by participants closely followed with improving access to diagnosis.

References

- ACAS. (2024, January 8). *Definition of disability—What disability means by law*. Acas.
<https://www.acas.org.uk/what-disability-means-by-law>
- Adams, A., Hills, P. J., Bennetts, R. J., & Bate, S. (2020a). Coping strategies for developmental prosopagnosia. *Neuropsychological Rehabilitation*, 30(10), 1996–2015. <https://doi.org/10.1080/09602011.2019.1623824>
- Adams, A., Hills, P. J., Bennetts, R. J., & Bate, S. (2020b). Coping strategies for developmental prosopagnosia. *Neuropsychological Rehabilitation*, 30(10), 1996–2015. <https://doi.org/10.1080/09602011.2019.1623824>
- Albonico, A., & Barton, J. (2019). Progress in perceptual research: The case of prosopagnosia version 1; peer review: 2 approved]. *F1000Research*, 8(765).
<https://doi.org/10.12688/f1000research.18492.1>
- 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](https://doi.org/10.1016/S0010-9452(96)80048-7)
- Arizpe, J., Saad, E., Douglas, A., Germine, L., Wilmer, J., & DeGutis, J. (2019). Self-reported face recognition is highly valid, but alone is not highly discriminative of prosopagnosia-level performance on objective assessments. *Behavior Research Methods*, 51. <https://doi.org/10.3758/s13428-018-01195-w>
- Arnold, S. R., Higgins, J. M., Weise, J., Desai, A., Pellicano, E., & Trollor, J. N. (2023). Towards the measurement of autistic burnout. *Autism*, 27(7), 1933–1948.
<https://doi.org/10.1177/13623613221147401>

- Avidan, G., & Behrmann, M. (2008). Implicit familiarity processing in congenital prosopagnosia. *Journal of Neuropsychology*, 2(1), 141–164.
<https://doi.org/10.1348/174866407X260180>
- 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>
- Baker, K. A., Laurence, S., & Mondloch, C. J. (2017). How does a newly encountered face become familiar? The effect of within-person variability on adults' and children's perception of identity. *Cognition*, 161, 19–30.
- Balas, B., & Saville, A. (2015). N170 face specificity and face memory depend on hometown size. *Neuropsychologia*, 69, 211–217.
<https://doi.org/10.1016/j.neuropsychologia.2015.02.005>
- Barton, J. J. S., Cherkasova, M., & O'Connor, M. (2001). Covert recognition in acquired and developmental prosopagnosia. *Neurology*, 57(7), 1161–1168.
<https://doi.org/10.1212/WNL.57.7.1161>
- Barton, J. J. S., Press, D. Z., Keenan, J. P., & O'Connor, M. (2002). Lesions of the fusiform face area impair perception of facial configuration in prosopagnosia. *Neurology*, 58(1), 71–78. <https://doi.org/10.1212/WNL.58.1.71>
- Bate, S. (2011). Review of Understanding facial recognition difficulties in children: Prosopagnosia management strategies for parents and professionals. *Journal of Psychiatric and Mental Health Nursing*, 18(6), e17–e18.
<https://doi.org/10.1111/j.1365-2850.2011.01759.x>
- Bate, S. (2013). *Face recognition and its disorders*. Palgrave Macmillan.

- Bate, S., Adams, A., Bennetts, R., & Line, H. (2019). Developmental prosopagnosia with concurrent topographical difficulties: A case report and virtual reality training programme. *Neuropsychological Rehabilitation*, 29(8), 1290–1312. <https://doi.org/10.1080/09602011.2017.1409640>
- Bate, S., & Bennetts, R. J. (2014). The rehabilitation of face recognition impairments: A critical review and future directions. *Frontiers in Human Neuroscience*, 8(Journal Article), 491–491. <https://doi.org/10.3389/fnhum.2014.00491>
- 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., 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., Cook, S. J., Duchaine, B., Tree, J. J., Burns, E. J., & Hodgson, T. L. (2014). Intranasal inhalation of oxytocin improves face processing in developmental prosopagnosia. *Cortex*, 50, 55–63. <https://doi.org/10.1016/j.cortex.2013.08.006>
- 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., Haslam, C., Tree, J. J., & Hodgson, T. L. (2008). Evidence of an eye movement-based memory effect in congenital prosopagnosia. *Cortex: A Journal Devoted to the Study of the Nervous System and Behavior*, 44(7), 806–819. <https://doi.org/10.1016/j.cortex.2007.02.004>
- Bate, S., Murray, E., & Bennetts, R. J. (2024). Familial Transmission of Developmental Prosopagnosia: New Case Reports from an Extended Family and Identical Twins. *Brain Sciences*, 14(1), Article 1. <https://doi.org/10.3390/brainsci14010049>
- Bate, S., & Tree, J. J. (2017). The definition and diagnosis of developmental prosopagnosia. *The Quarterly Journal of Experimental Psychology*, 70(2), 193–200. <https://doi.org/10.1080/17470218.2016.1195414>
- Bateson, P., & Horn, G. (1994). Imprinting and recognition memory: A neural net model. *Animal Behaviour*, 48(3), 695–715. <https://doi.org/10.1006/anbe.1994.1289>
- Behrmann, M., Avidan, G., Marotta, J., & Kimchi, R. (2005). Detailed Exploration of Face-related Processing in Congenital Prosopagnosia: 1. Behavioral Findings. *Journal of Cognitive Neuroscience*, 17, 1130–1149. <https://doi.org/10.1162/0898929054475154>
- Behrmann, M., & Geskin, J. (2018). Over time, the right results will emerge. *Cognitive Neuropsychology*, 35(1–2), 102–111. <https://doi.org/10.1080/02643294.2018.1447917>
- 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 (Methodological)*, 57(1), 289–300.
- 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>
- Bentin, S., DeGutis, J. M., D’Esposito, M., & Robertson, L. C. (2007). Too Many Trees to See the Forest: Performance, Event-related Potential, and Functional Magnetic Resonance Imaging Manifestations of Integrative Congenital Prosopagnosia. *Journal of Cognitive Neuroscience*, 19(1), 132–146. <https://doi.org/10.1162/jocn.2007.19.1.132>
- Benton, A. L., Hamsher, K. d, Varney, N. R., & Spreen, O. (1983). Benton test of facial recognition. *New York, NY: New York, NY.*
- Bindemann, M., & Hole, G. J. (2020). Understanding face identification through within-person variability in appearance: Introduction to a virtual special issue. *Quarterly Journal of Experimental Psychology* (2006), 73(12), NP1–NP8. <https://doi.org/10.1177/1747021820959068>

- Bindemann, M., & Lewis, M. B. (2013). Face detection differs from categorization: Evidence from visual search in natural scenes. *Psychonomic Bulletin & Review*, 20(6), 1140–1145. <https://doi.org/10.3758/s13423-013-0445-9>
- Biotti, F., Gray, K. L. H., & Cook, R. (2017). Impaired body perception in developmental prosopagnosia. *Cortex*, 93, 41–49. <https://doi.org/10.1016/j.cortex.2017.05.006>
- Biotti, F., Gray, K. L. H., & Cook, R. (2019). Is developmental prosopagnosia best characterised as an apperceptive or mnemonic condition? *Neuropsychologia*, 124, 285–298. <https://doi.org/10.1016/j.neuropsychologia.2018.11.014>
- Biotti, F., Wu, E., Hua Yang, Guo Jiahui, Duchaine, B., & Cook, R. (2017). Normal composite face effects in developmental prosopagnosia. *Cortex*, 95(Journal Article), 63–76. <https://doi.org/10.1016/j.cortex.2017.07.018>
- 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*. <https://doi.org/10.1016/j.cortex.2023.05.018>
- Boutet, I., Nelson, E. A., Watier, N., Cousineau, D., Béland, S., & Collin, C. A. (2021). Different measures of holistic face processing tap into distinct but partially overlapping mechanisms. *Attention, Perception, & Psychophysics*, 83(7), 2905–2923. <https://doi.org/10.3758/s13414-021-02337-7>
- 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>

- Bradley, L., Shaw, R., Baron-Cohen, S., & Cassidy, S. (2021). Autistic Adults' Experiences of Camouflaging and Its Perceived Impact on Mental Health. *Autism in Adulthood*, 3(4), 320–329. <https://doi.org/10.1089/aut.2020.0071>
- Braun, V., & Clarke, V. (2021). One size fits all? What counts as quality practice in (reflexive) thematic analysis? *Qualitative Research in Psychology*, 18(3), 328–352. <https://doi.org/10.1080/14780887.2020.1769238>
- Bruce, V., Burton, M. A., & Dench, N. (1994). What's Distinctive about a Distinctive Face? *The Quarterly Journal of Experimental Psychology Section A*, 47(1), 119–141. <https://doi.org/10.1080/14640749408401146>
- Bruce, V., Henderson, Z., Greenwood, K., Hancock, P. J. B., Burton, A. M., & Miller, P. (1999). Verification of face identities from images captured on video. *Journal of Experimental Psychology: Applied*, 5(4), 339–360. <https://doi.org/10.1037/1076-898X.5.4.339>
- Bruce, V., & Valentine, T. (1986). Semantic Priming of Familiar Faces. *The Quarterly Journal of Experimental Psychology Section A*, 38(1), 125–150. <https://doi.org/10.1080/14640748608401588>
- Bruce, V., & Young, A. (1986). Understanding face recognition. *British Journal of Psychology*, 77(Journal Article), 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>
- Bruyer, R., & Brysbaert, M. (2011). Combining Speed and Accuracy in Cognitive Psychology: Is the Inverse Efficiency Score (IES) a Better Dependent Variable

- than the Mean Reaction Time (RT) and the Percentage Of Errors (PE)?
Psychologica Belgica, 51(1), Article 1. <https://doi.org/10.5334/pb-51-1-5>
- Burns, E. (2023a). *A symptom guided diagnosis of prosopagnosia is valid: A commentary on DeGutis et al.(2023)*. <https://psyarxiv.com/9aqdf/>
- Burns, E. (2023b, March 23). *People with faceblindness aren't believed by their doctors – here's what needs to change*. The Conversation. <http://theconversation.com/people-with-faceblindness-arent-believed-by-their-doctors-heres-what-needs-to-change-200859>
- Burns, E. J. (2024). Improving the DSM-5 approach to cognitive impairment: Developmental prosopagnosia reveals the need for tailored diagnoses. *Behavior Research Methods*. <https://doi.org/10.3758/s13428-024-02459-4>
- Burns, E. J., Gaunt, E., Kidane, B., Hunter, L., & Pulford, J. (2022). A new approach to diagnosing and researching developmental prosopagnosia: Excluded cases are impaired too. *Behavior Research Methods*. <https://doi.org/10.3758/s13428-022-02017-w>
- Burton, A. M., Bruce, V., & Johnstone, Robert. A. (1990). Understanding face recognition with an interactive activation model. *British Journal of Psychology*, 81(3), 361. <https://doi.org/10.1111/j.2044-8295.1990.tb02367.x>
- Bushnell, I. W. R. (2001). Mother's face recognition in newborn infants: Learning and memory. *Infant and Child Development*, 10(1–2), 67–74. <https://doi.org/10.1002/icd.248>
- Busigny, T., Van Belle, G., Jemel, B., Hosein, A., Joubert, S., & Rossion, B. (2014). Face-specific impairment in holistic perception following focal lesion of the right

- anterior temporal lobe. *Neuropsychologia*, 56, 312–333.
<https://doi.org/10.1016/j.neuropsychologia.2014.01.018>
- Butterworth, B., Varma, S., & Laurillard, D. (2011). Dyscalculia: From brain to education. *Science (New York, N.Y.)*, 332(6033), 1049–1053.
<https://doi.org/10.1126/science.1201536>
- Bylemans, T., Vrancken, L., & Verfaillie, K. (2020). Developmental Prosopagnosia and Elastic Versus Static Face Recognition in an Incidental Learning Task. *Frontiers in Psychology*, 11. <https://doi.org/10.3389/fpsyg.2020.02098>
- Carbon, C.-C., Grüter, T., Weber, J. E., & Lueschow, A. (2007). Faces as objects of non-expertise: Processing of thatcherised faces in congenital prosopagnosia. *Perception*, 36(11), 1635–1645.
- Carey, S., & Diamond, R. (1977). From piecemeal to configurational representation of faces. *Science (New York, N.Y.)*, 195(4275), 312–314.
<https://doi.org/10.1126/science.831281>
- Carey, S., & Diamond, R. (1994). Are faces perceived as configurations more by adults than by children? *Visual Cognition*, 1(2–3), 253–274.
<https://doi.org/10.1080/13506289408402302>
- Cattaneo, Z., Daini, R., Malaspina, M., Manai, F., Lillo, M., Fermi, V., Schiavi, S., Suchan, B., & Comincini, S. (2016). Congenital prosopagnosia is associated with a genetic variation in the oxytocin receptor (OXTR) gene: An exploratory study. *Neuroscience*, 339, 162–173. <https://doi.org/10.1016/j.neuroscience.2016.09.040>
- Cattaneo, Z., Vecchi, T., Monegato, M., Pece, A., Merabet, L. B., & Carbon, C.-C. (2013). Strabismic amblyopia affects relational but not featural and Gestalt

- processing of faces. *Vision Research*, 80, 19–30.
<https://doi.org/10.1016/j.visres.2013.01.007>
- Chapman, L., Rose, K., Hull, L., & Mandy, W. (2022). “I want to fit in... but I don’t want to change myself fundamentally”: A qualitative exploration of the relationship between masking and mental health for autistic teenagers. *Research in Autism Spectrum Disorders*, 99, 102069.
<https://doi.org/10.1016/j.rasd.2022.102069>
- Chatterjee, G. (2012). *Perceptual and cognitive processes in developmental prosopagnosia* (2012-99200-568; Issues 4-B) [ProQuest Information & Learning]. APA PsycInfo.
<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>
- Chierchia, G., Fuhrmann, D., Knoll, L. J., Pi-Sunyer, B. P., Sakhardande, A. L., & Blakemore, S.-J. (2019). The Matrix Reasoning Item Bank (MaRs-IB): Novel, open-access abstract reasoning items for adolescents and adults. *Royal Society Open Science*, 6(10), 190232. <https://doi.org/10.1098/rsos.190232>
- Condon, D. M., & Revelle, W. (2014). The international cognitive ability resource: Development and initial validation of a public-domain measure. *Intelligence*, 43, 52–64. <https://doi.org/10.1016/j.intell.2014.01.004>
- Connolly, H. L., Young, A. W., & Lewis, G. J. (2019). Recognition of facial expression and identity in part reflects a common ability, independent of general intelligence

- and visual short-term memory. *Cognition and Emotion*, 33(6), 1119–1128.
<https://doi.org/10.1080/02699931.2018.1535425>
- Cook, R., & Biotti, F. (2016). Developmental prosopagnosia. *Current Biology: CB*, 26(8), R312–313. <https://doi.org/10.1016/j.cub.2016.01.008>
- Copet, P., Jauregi, J., Laurier, V., Ehlinger, V., Arnaud, C., Cobo, A.-M., Molinas, C., Tauber, M., & Thuilleaux, D. (2010). Cognitive profile in a large French cohort of adults with Prader-Willi syndrome: Differences between genotypes. *Journal of Intellectual Disability Research: JIDR*, 54(3), 204–215.
<https://doi.org/10.1111/j.1365-2788.2010.01251.x>
- Corrow, S. L., Dalrymple, K. A., & Barton, J. J. (2016). Prosopagnosia: Current perspectives. *Eye and Brain*, 8, 165. <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*, 31.
- 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>
- Croydon, A., Pimperton, H., Ewing, L., Duchaine, B. C., & Pellicano, E. (2014). The Cambridge Face Memory Test for Children (CFMT-C): A new tool for measuring

- face recognition skills in childhood. *Neuropsychologia*, 62(Journal Article), 60–67. <https://doi.org/10.1016/j.neuropsychologia.2014.07.008>
- Dalrymple, K. A. (2012). Developmental prosopagnosia in childhood. *COGNITIVE NEUROPSYCHOLOGY*, 29(5–6), 393.
- 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., Elison, J. T., & Duchaine, B. (2017). Face-specific and domain-general visual processing deficits in children with developmental prosopagnosia. *Quarterly Journal of Experimental Psychology*, 70(2), Article 2.
- Dalrymple, K. A., Fletcher, K., Corrow, S., das Nair, R., Barton, J. J. S., Yonas, A., & Duchaine, B. (2014). ‘A room full of strangers every day’: The psychosocial impact of developmental prosopagnosia on children and their families. *Journal of Psychosomatic Research*, 77(2), 144–150. <https://doi.org/10.1016/j.jpsychores.2014.06.001>
- 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(Journal Article), 10–20. <https://doi.org/10.1016/j.dcn.2014.07.003>
- Dalrymple, K. A., & Palermo, R. (2016). Guidelines for studying developmental prosopagnosia in adults and children. *Wiley Interdisciplinary Reviews-Cognitive Science*, 7(1), 73–87. <https://doi.org/10.1002/wcs.1374>

- Dalrymple, K., Gomez, J., & Duchaine, B. (2012). CFMT-Kids: A new test of face memory for children. *Journal of Vision*, 12(9), 492–492. <https://doi.org/10.1167/12.9.492>
- Davies-Thompson, J., Pancaroglu, R., & Barton, J. (2014). Acquired prosopagnosia: Structural basis and processing impairments. *Frontiers in Bioscience-Elite*, 6(1), Article 1. <https://doi.org/10.2741/E699>
- Davis, J., Mckone, E., Dennett, H., O'Connor, K., O'Kearney, R., & Palermo, R. (2011). Individual Differences in the Ability to Recognise Facial Identity Are Associated with Social Anxiety. *PloS One*, 6, e28800. <https://doi.org/10.1371/journal.pone.0028800>
- De Haan, E. H. F. (1999). A Familial Factor in the Development of Face Recognition Deficits. *Journal of Clinical and Experimental Neuropsychology*, 21(3), 312–315. <https://doi.org/10.1076/jcen.21.3.312.917>
- De Haan, E., Young, A., & Newcombe, F. (1991). Covert and overt recognition in prosopagnosia. *Brain*, 114, 2575–2591. <https://doi.org/10.1093/brain/114.6.2575>
- 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](https://doi.org/10.1016/S0010-9452(13)80125-6)
- 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(Journal Article), 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*. <https://doi.org/10.31234/osf.io/wtvxd>
- 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>
- DeGutis, J. M., Chiu, C., Grosso, M. E., & Cohan, S. (2014). Face processing improvements in prosopagnosia: Successes and failures over the last 50 years. *Frontiers in Human Neuroscience*, 8, 561.
<https://doi.org/10.3389/fnhum.2014.00561>
- DeGutis, J., Yosef, B., Lee, E. A., Saad, E., Arizpe, J., Song, J. S., Wilmer, J., Germine, L., & Esterman, M. (2023). The rise and fall of face recognition awareness across the life span. *Journal of Experimental Psychology: Human Perception and Performance*, 49, 22–33. <https://doi.org/10.1037/xhp0001069>

- Diamond, R., & Carey, S. (1986). Why faces are and are not special: An effect of expertise. *Journal of Experimental Psychology: General*, 115(2), 107–117.
<https://doi.org/10.1037/0096-3445.115.2.107>
- Diaz, A. L. (2008). Do I Know You? A Case Study of Prosopagnosia (Face Blindness). *Journal of School Nursing*, 24(5), 284–289.
<https://doi.org/10.1177/1059840508322381>
- 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](https://doi.org/10.1016/S0010-9452(08)70501-X)
- Draheim, C., Mashburn, C. A., Martin, J. D., & Engle, R. W. (2019). Reaction time in differential and developmental research: A review and commentary on the problems and alternatives. *Psychological Bulletin*, 145(5), 508–535.
<https://doi.org/10.1037/bul0000192>
- Duchaine, B. (2008). Comment on prevalence of hereditary prosopagnosia (HPA) in Hong Kong Chinese population. *American Journal of Medical Genetics Part A*, 146A(22), 2860–2862. <https://doi.org/10.1002/ajmg.a.32548>
- 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., 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>

- 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., & Yovel, G. (2015). A Revised Neural Framework for Face Processing. *Annual Review of Vision Science*, 1(1), 393–416. <https://doi.org/10.1146/annurev-vision-082114-035518>
- Duchaine, B., Yovel, G., & Nakayama, K. (2007). No global processing deficit in the Navon task in 14 developmental prosopagnosics. *Social Cognitive and Affective Neuroscience*, 2, 104–113. <https://doi.org/10.1093/scan/nsm003>
- Eimer, M. (2018). What do associations and dissociations between face and object recognition abilities tell us about the domain-generalty of face processing? *Cognitive Neuropsychology*, 35(1–2), 80–82. <https://doi.org/10.1080/02643294.2017.1414691>
- Ellis, A. W., Young, A. W., Flude, B. M., & Hay, D. C. (1987). Repetition priming of face recognition. *The Quarterly Journal of Experimental Psychology Section A*, 39(2), 193–210. <https://doi.org/10.1080/14640748708401784>
- Engel, S., Klusmann, H., Ditzen, B., Knaevelsrud, C., & Schumacher, S. (2019). Menstrual cycle-related fluctuations in oxytocin concentrations: A systematic review and meta-analysis. *Frontiers in Neuroendocrinology*, 52, 144–155. <https://doi.org/10.1016/j.yfrne.2018.11.002>

- Epihova, G., & Astle, D. E. (2024). What is developmental about developmental prosopagnosia? *Cortex*. <https://doi.org/10.1016/j.cortex.2024.02.006>
- Equality Act 2010 (2010). <https://www.legislation.gov.uk/ukpga/2010/15/contents>
- Esins, J., Schultz, J., Stemper, C., Kennerknecht, I., & Bühlhoff, I. (2016). Face Perception and Test Reliabilities in Congenital Prosopagnosia in Seven Tests. *I-Perception*, 7. <https://doi.org/10.1177/2041669515625797>
- Esins, J., Schultz, J., Stemper, C., Kennerknecht, I., Wallraven, C., & Bühlhoff, I. (2015). ‘Do congenital prosopagnosia and the other-race effect affect the same face recognition mechanisms?’: Corrigendum. *Frontiers in Human Neuroscience*, 9. <https://doi.org/10.3389/fnhum.2015.00294>
- Faceblind UK. (n.d.). *History*. Face Blind UK. Retrieved 16 December 2020, from <https://www.faceblind.org.uk/information/history/>
- Fine, D. R. (2012). A life with prosopagnosia. *Cognitive Neuropsychology*, 29(5–6), 354–359. <https://doi.org/10.1080/02643294.2012.736377>
- Freire, L. de A., Negrão, J. V. R. de T., Venâncio, T. S., Araújo, B. M. F. de, & Kasahara, N. (2022). Face memory deficit in visually impaired children who live in a developing country. *Applied Neuropsychology: Child*, 0(0), 1–7. <https://doi.org/10.1080/21622965.2022.2108710>
- Fysh, M. C., & Ramon, M. (2022). Accurate but inefficient: Standard face identity matching tests fail to identify prosopagnosia. *Neuropsychologia*, 165, 108119. <https://doi.org/10.1016/j.neuropsychologia.2021.108119>
- 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>
- Gauthier, I., & Tarr, M. J. (2002). Unraveling mechanisms for expert object recognition: Bridging brain activity and behavior. *Journal of Experimental Psychology-Human Perception and Performance*, 28(2), 431–446. <https://doi.org/10.1037//0096-1523.28.2.431>
- Geldart, S., Mondloch, C. J., Maurer, D., De Schonen, S., & Brent, H. P. (2002). The effect of early visual deprivation on the development of face processing. *Developmental Science*, 5(4), 490–501. <https://doi.org/10.1111/1467-7687.00242>
- Gelman, A., Hill, J., & Yajima, M. (2012). Why We (Usually) Don't Have to Worry About Multiple Comparisons. *Journal of Research on Educational Effectiveness*, 5(2), 189–211. <https://doi.org/10.1080/19345747.2011.618213>
- Gerlach, C., Nørkær, E., & Starrfelt, R. (2024). Class A, Class B. Is that the only chemistry? A commentary on DeGutis et al. (2023): *What is the prevalence of developmental prosopagnosia? An empirical assessment of different diagnostic cutoffs*. *Cortex*. <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*, 1–23. <https://doi.org/10.1080/02643294.2020.1869709>
- Gerlach, C., Susilo, T., Barton, J. J. S., Albonico, A., Malaspina, M., & Starrfelt, R. (2020). *Contrasting shared- and specific-mechanism accounts of developmental prosopagnosia: A new approach*. PsyArXiv. <https://doi.org/10.31234/osf.io/zjymv>

- Germine, L. T., Duchaine, B., & Nakayama, K. (2011). Where cognitive development and aging meet: Face learning ability peaks after age 30. *Cognition*, 118(2), 201–210. <https://doi.org/10.1016/j.cognition.2010.11.002>
- 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>
- Gobbini, M. I., & Haxby, J. V. (2007). Neural systems for recognition of familiar faces. *Neuropsychologia*, 45(1), 32–41. <https://doi.org/10.1016/j.neuropsychologia.2006.04.015>
- Gray, K. L. H., Bird, G., & Cook, R. (2017). Robust associations between the 20-item prosopagnosia index and the Cambridge Face Memory Test in the general population. *Royal Society Open Science*, 4(3), 160923–160923. <https://doi.org/10.1098/rsos.160923>
- Gray, K. L. H., & Cook, R. (2018). Should developmental prosopagnosia, developmental body agnosia, and developmental object agnosia be considered independent neurodevelopmental conditions? *Cognitive Neuropsychology*, 35(1–2), 59–62. <https://doi.org/10.1080/02643294.2018.1433153>
- Griffin, J. W., Bauer, R., & Scherf, K. S. (2021). A quantitative meta-analysis of face recognition deficits in autism: 40 years of research. *Psychological Bulletin*, 147(3), 268–292. <https://doi.org/10.1037/bul0000310>
- Grueter, M., Grueter, T., Bell, V., Horst, J., Laskowski, W., Sperling, K., Halligan, P. W., Elli, H. D., & Kennerknecht, I. (2007). Hereditary prosopagnosia: The first case series. *Cortex*, 43(6), 734–749.

- Grüter, T., & Grüter, M. (2007). Prosopagnosia in biographies and autobiographies. *Perception*, 36(2), 299–301.
- Hancock, P. J. B. (2012). Unfamiliar face recognition. In C. Wilkinson & C. Rynn (Eds.), *Craniofacial Identification* (pp. 11–23). Cambridge Univ Press.
- Haxby, J. V., Hoffman, E. A., & Gobbini, M. I. (2000). The distributed human neural system for face perception. *Trends in Cognitive Sciences*, 4(6), 223–233.
[https://doi.org/10.1016/S1364-6613\(00\)01482-0](https://doi.org/10.1016/S1364-6613(00)01482-0)
- Haxby, J. V., Hoffman, E. A., & Gobbini, M. I. (2002). Human neural systems for face recognition and social communication. *Biological Psychiatry*, 51(1), 59–67.
[https://doi.org/10.1016/S0006-3223\(01\)01330-0](https://doi.org/10.1016/S0006-3223(01)01330-0)
- Huis in 't Veld, E., Van den Stock, J., & de Gelder, B. (2012). Configuration perception and face memory, and face context effects in developmental prosopagnosia. *Cognitive Neuropsychology*, 29(5–6), 464–481.
<https://doi.org/10.1080/02643294.2012.732051>
- Idei, H., & Yamashita, Y. (2024). Elucidating multifinal and equifinal pathways to developmental disorders by constructing real-world neurorobotic models. *Neural Networks*, 169, 57–74. <https://doi.org/10.1016/j.neunet.2023.10.005>
- JASP Team. (2023). *JASP* (Version 0.18.0) [Computer software]. <https://jasp-stats.org>
- Jenkins, R., Dowsett, A. J., & Burton, A. M. (2018). How many faces do people know? *Proceedings of the Royal Society B: Biological Sciences*, 285(1888), 20181319.
<https://doi.org/10.1098/rspb.2018.1319>
- Jenkins, R., White, D., Van Montfort, X., & Mike Burton, A. (2011). Variability in photos of the same face. *Cognition*, 121(3), 313–323.
<https://doi.org/10.1016/j.cognition.2011.08.001>

- Jiahui, G., Yang, H., & Duchaine, B. (2018). Developmental prosopagnosics have widespread selectivity reductions across category-selective visual cortex. *Proceedings of the National Academy of Sciences of the United States of America*, 115(28), E6418–E6427. <https://doi.org/10.1073/pnas.1802246115>
- Johnen, A., Schmukle, S. C., Hüttenbrink, J., Kischka, C., Kennerknecht, I., & Döbel, C. (2014). A family at risk: Congenital prosopagnosia, poor face recognition and visuoperceptual deficits within one family. *Neuropsychologia*, 58, 52–63. <https://doi.org/10.1016/j.neuropsychologia.2014.03.013>
- Johnson, M. H., Senju, A., & Tomalski, P. (2015). The two-process theory of face processing: Modifications based on two decades of data from infants and adults. *Neuroscience & Biobehavioral Reviews*, 50, 169–179. <https://doi.org/10.1016/j.neubiorev.2014.10.009>
- 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>
- Kanwisher, N., McDermott, J., & Chun, M. M. (1997). The Fusiform Face Area: A Module in Human Extrastriate Cortex Specialized for Face Perception. *Journal of Neuroscience*, 17(11), 4302–4311. <https://doi.org/10.1523/JNEUROSCI.17-11-04302.1997>
- Kennerknecht, I., Grueter, T., Welling, B., Wentzek, S., Horst, J., Edwards, S., & Grueter, M. (2006). First report of prevalence of non-syndromic hereditary prosopagnosia (HPA). *American Journal of Medical Genetics Part A*, 140A(15), 1617–1622. <https://doi.org/10.1002/ajmg.a.31343>

- 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, 2863–2870. <https://doi.org/10.1002/ajmg.a.32552>
- Kho, S. K., Leong, B. Q. Z., Keeble, D. R. T., Wong, H. K., & Estudillo, A. J. (2023). A new Asian version of the CFMT: The Cambridge Face Memory Test – Chinese Malaysian (CFMT-MY). *Behavior Research Methods*. <https://doi.org/10.3758/s13428-023-02085-6>
- 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>
- Klargaard, S. K., Starrfelt, R., Petersen, A., & Gerlach, C. (2016). Topographic processing in developmental prosopagnosia: Preserved perception but impaired memory of scenes. *Cognitive Neuropsychology*, 33(7–8), 405–413. <https://doi.org/10.1080/02643294.2016.1267000>
- 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>
- Kosaka, H., Fujioka, T., & Jung, M. (2019). Symptoms in individuals with adult-onset ADHD are masked during childhood. *European Archives of Psychiatry and Clinical Neuroscience*, 269(6), 753–755. <https://doi.org/10.1007/s00406-018-0893-3>

- 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>
- Kruschke, J. K. (2010). What to believe: Bayesian methods for data analysis. *Trends in Cognitive Sciences*, 14(7), 293–300. <https://doi.org/10.1016/j.tics.2010.05.001>
- Landau, B., & Ferrara, K. (2013). Space and language in Williams syndrome: Insights from typical development. *WIREs Cognitive Science*, 4(6), 693–706. <https://doi.org/10.1002/wcs.1258>
- Laurence, S., & Mondloch, C. J. (2016). That’s my teacher! Children’s ability to recognize personally familiar and unfamiliar faces improves with age. *Journal of Experimental Child Psychology*, 143, 123–138. <https://doi.org/10.1016/j.jecp.2015.09.030>
- 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>
- 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>
- Leong, B. Q. Z., Estudillo, A. J., & Hussain Ismail, A. M. (2023). Holistic and featural processing’s link to face recognition varies by individual and task. *Scientific Reports*, 13(1), 16869. <https://doi.org/10.1038/s41598-023-44164-w>

- 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*. <https://doi.org/10.3758/s13428-022-01843-2>
- Lissauer, H. (1890). Ein Fall von Seelenblindheit nebst einem Beitrage zur Theorie derselben. *Archiv für Psychiatrie und Nervenkrankheiten*, 21(2), 222–270. <https://doi.org/10.1007/BF02226765>
- Liu, J., Harris, A., & Kanwisher, N. (2002). Stages of processing in face perception: An MEG study. *Nature Neuroscience*, 5(9), 910–916. <https://doi.org/10.1038/nn909>
- Livingston, L. A., & Shah, P. (2018). People with and without prosopagnosia have insight into their face recognition ability. *QUARTERLY JOURNAL OF EXPERIMENTAL PSYCHOLOGY*, 71(5), 1260–1262. <https://doi.org/10.1080/17470218.2017.1310911>
- 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>
- Louveau, C., Turtuluci, M.-C., Consoli, A., Poitou, C., Coupaye, M., Krebs, M.-O., Chaumette, B., & Iftimovici, A. (2023). Prader-Willi syndrome: Symptoms and

- topiramate response in light of genetics. *Frontiers in Neuroscience*, 17, 1126970.
<https://doi.org/10.3389/fnins.2023.1126970>
- 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*, 0(0), 1–12. <https://doi.org/10.1080/13506285.2024.2359743>
- Lumivero. (2023). NVivo (Version 14) [Computer software]. Lumivero. www.lumivero.com
- Manippa, V., Palmisano, A., Ventura, M., & Rivolta, D. (2023). The Neural Correlates of Developmental Prosopagnosia: Twenty-Five Years on. *Brain Sciences*, 13(10), Article 10. <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. <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](https://doi.org/10.1016/s1364-6613(02)01903-4)
- McConachie, H. R. (1976). Developmental prosopagnosia: A single case report. *Cortex: A Journal Devoted to the Study of the Nervous System and Behavior*, 12(1), 76–82. [https://doi.org/10.1016/S0010-9452\(76\)80033-0](https://doi.org/10.1016/S0010-9452(76)80033-0)

- 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>
- McKone, E., Stokes, S., Liu, J., Cohan, S., Fiorentini, C., Pidcock, M., Yovel, G., Broughton, M., & Pelleg, M. (2012). A Robust Method of Measuring Other-Race and Other-Ethnicity Effects: The Cambridge Face Memory Test Format. *PLOS ONE*, 7(10), e47956. <https://doi.org/10.1371/journal.pone.0047956>
- McKone, E., Wan, L., Robbins, R., Crookes, K., & Liu, J. (2017). Diagnosing prosopagnosia in East Asian individuals: Norms for the Cambridge Face Memory Test–Chinese. *Cognitive Neuropsychology*, 34(5), 253–268. <https://doi.org/10.1080/02643294.2017.1371682>
- McKone, E., & Yovel, G. (2009). Why does picture-plane inversion sometimes dissociate perception of features and spacing in faces, and sometimes not? Toward a new theory of holistic processing. *Psychonomic Bulletin & Review*, 16(5), 778–797. <https://doi.org/10.3758/PBR.16.5.778>
- Megreya, A. M., & Burton, A. M. (2006). Unfamiliar faces are not faces: Evidence from a matching task. *Memory & Cognition*, 34(4), 865–876. <https://doi.org/10.3758/BF03193433>
- Meissner, C. A., & Brigham, J. C. (2001). Thirty years of investigating the own-race bias in memory for faces: A meta-analytic review. *Psychology, Public Policy, and Law*, 7(1), 3–35. <https://doi.org/10.1037/1076-8971.7.1.3>
- Meyer, K., Schmitz, F., Wilhelm, O., & Hildebrandt, A. (2018). Perceiving faces: Too much, too fast?—face specificity in response caution. *Journal of Experimental*

- Psychology: Human Perception and Performance*, 45(1), 16.
<https://doi.org/10.1037/xhp0000582>
- Miller, D., Rees, J., & Pearson, A. (2021). “Masking Is Life”: Experiences of Masking in Autistic and Nonautistic Adults. *Autism in Adulthood: Challenges and Management*, 3(4), 330–338. <https://doi.org/10.1089/aut.2020.0083>
- Minear, M., & Park, D. C. (2004). A lifespan database of adult facial stimuli. *Behavior Research Methods, Instruments, & Computers*, 36(4), 630–633. <https://doi.org/10.3758/BF03206543>
- Minio-Paluello, I., Porciello, G., Pascual-Leone, A., & Baron-Cohen, S. (2020). Face individual identity recognition: A potential endophenotype in autism. *Molecular Autism*, 11, 81. <https://doi.org/10.1186/s13229-020-00371-0>
- 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>
- Mishra, M. V., Fry, R. M., Saad, E., Arizpe, J. M., Ohashi, Y.-G. B., & DeGutis, J. M. (2021). Comparing the sensitivity of face matching assessments to detect face perception impairments. *Neuropsychologia*, 163, 108067. <https://doi.org/10.1016/j.neuropsychologia.2021.108067>
- Mondloch, C. J., Segalowitz, S. J., Lewis, T. L., Dywan, J., Le Grand, R., & Maurer, D. (2013). The effect of early visual deprivation on the development of face detection. *Developmental Science*, 16(5), 728–742. <https://doi.org/10.1111/desc.12065>

- Morton, J., & Johnson, M. H. (1991). CONSPEC and CONLERN: A two-process theory of infant face recognition. *Psychological Review*, 98(2), 164–181. <https://doi.org/10.1037/0033-295x.98.2.164>
- Murray, E., & Bate, S. (2019). Self-ratings of face recognition ability are influenced by gender but not prosopagnosia severity. *Psychological Assessment*, 31(6), 828–832. <https://doi.org/10.1037/pas0000707>
- Murray, E., & Bate, S. (2020). Diagnosing developmental prosopagnosia: Repeat assessment using the Cambridge Face Memory Test. *Royal Society Open Science*, 7(9), 200884. <https://doi.org/10.1098/rsos.200884>
- Murray, E., Bennetts, R., Tree, J., & Bate, S. (2022). An update of the Benton Facial Recognition Test. *Behavior Research Methods*, 54(5), 2318–2333. <https://doi.org/10.3758/s13428-021-01727-x>
- Murray, E., Hills, P. J., Bennetts, R. J., & Bate, S. (2018). Identifying hallmark symptoms of developmental prosopagnosia for non-experts. *Scientific Reports*, 8(Journal Article), 1690–1690. <https://doi.org/10.1038/s41598-018-20089-7>
- Nelson, C. A. (2001). The development and neural bases of face recognition. *Infant and Child Development*, 10(1–2), 3–18. <https://doi.org/10.1002/icd.239>
- Németh, K., Zimmer, M., Schweinberger, S. R., Vakli, P., & Kovács, G. (2014). The Background of Reduced Face Specificity of N170 in Congenital Prosopagnosia. *PLoS ONE*, 9(7), e101393. <https://doi.org/10.1371/journal.pone.0101393>
- NHS. (2024, July 19). *Prosopagnosia (face blindness)*. Nhs.Uk. <https://www.nhs.uk/conditions/face-blindness/>
- Niven, J. (2016). *Holding up the universe* (First edition). Alfred A. Knopf.

- Nørkær, E., Gobbo, S., Roald, T., & Starrfelt, R. (2024). Disentangling Developmental Prosopagnosia: A scoping review of terms, tools and topics. *Cortex*.
<https://doi.org/10.1016/j.cortex.2024.04.011>
- 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), Article 2.
<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>
- Nussenbaum, K., Scheuplein, M., Phaneuf, C. V., Evans, M. D., & Hartley, C. A. (2020). Moving Developmental Research Online: Comparing In-Lab and Web-Based Studies of Model-Based Reinforcement Learning. *Collabra: Psychology*, 6(1), 17213. <https://doi.org/10.1525/collabra.17213>
- Palermo, R., Rossion, B., Rhodes, G., Laguesse, R., Tez, T., Hall, B., Albonico, A., Malaspina, M., Daini, R., Irons, J., Al-Janabi, S., Taylor, L. C., Rivolta, D., & McKone, E. (2017). Do people have insight into their face recognition abilities? *Quarterly Journal of Experimental Psychology* (2006), 70(2), 218–233.
<https://doi.org/10.1080/17470218.2016.1161058>
- Pascalis, O., de Haan, M., & Nelson, C. A. (2002). Is face processing species-specific during the first year of life? *Science (New York, N.Y.)*, 296(5571), 1321–1323.
<https://doi.org/10.1126/science.1070223>

- Pellicano, E., & Rhodes, G. (2003). Holistic Processing of Faces in Preschool Children and Adults. *Psychological Science*, 14(6), 618–622.
https://doi.org/10.1046/j.0956-7976.2003.psci_1474.x
- Pertsov, Y., Krill, D., Weiss, N., Lesinger, K., & Avidan, G. (2020). Rapid forgetting of faces in congenital prosopagnosia. *Cortex*, 129, 119–132.
<https://doi.org/10.1016/j.cortex.2020.04.007>
- Peterson, M. F., Zaun, I., Hoke, H., Guo Jiahui, Duchaine, B., & Kanwisher, N. (2019). Eye movements and retinotopic tuning in developmental prosopagnosia. *Journal of Vision*, 19(9), 7–7. <https://doi.org/10.1167/19.9.7>
- Phillips, P. J., Yates, A. N., Hu, Y., Hahn, C. A., Noyes, E., Jackson, K., Cavazos, J. G., Jeckeln, G., Ranjan, R., Sankaranarayanan, S., Chen, J.-C., Castillo, C. D., Chellappa, R., White, D., & O’Toole, A. J. (2018). Face recognition accuracy of forensic examiners, superrecognizers, and face recognition algorithms. *Proceedings of the National Academy of Sciences*, 115(24), 6171–6176.
<https://doi.org/10.1073/pnas.1721355115>
- Picozzi, M., Cassia, V. M., Turati, C., & Vescovo, E. (2009). The effect of inversion on 3- to 5-year-olds’ recognition of face and nonface visual objects. *Journal of Experimental Child Psychology*, 102(4), 487–502.
<https://doi.org/10.1016/j.jecp.2008.11.001>
- Portch, E., Wignall, L., & Bate, S. (2023). Why can people with developmental prosopagnosia recognise some familiar faces? Insights from subjective experience. *PeerJ*, 11, e15497. <https://doi.org/10.7717/peerj.15497>
- Putzar, L., Hötting, K., & Röder, B. (2010). Early visual deprivation affects the development of face recognition and of audio-visual speech perception.

- Restorative Neurology and Neuroscience*, 28(2), 251–257.
<https://doi.org/10.3233/RNN-2010-0526>
- Pyett, P. M. (2003). Validation of Qualitative Research in the “Real World”. *Qualitative Health Research*, 13(8), 1170–1179. <https://doi.org/10.1177/1049732303255686>
- R Core Team. (2021). *R: The R Project for Statistical Computing* [Computer software]. R Foundation for Statistical Computing. <https://www.r-project.org/>
- Ramon, M. (2021). Super-recognizers—A novel diagnostic framework, 70 cases, and guidelines for future work. *Neuropsychologia*, 158. <https://doi.org/10.1016/j.neuropsychologia.2021.107809>
- Ramon, M., Bobak, A. K., & White, D. (2019). Towards a ‘manifesto’ for super-recognizer research. *British Journal of Psychology*, 110(3), 495–498. <https://doi.org/10.1111/bjop.12411>
- Ramus, F. (2004). Neurobiology of dyslexia: A reinterpretation of the data. *Trends in Neurosciences*, 27(12), 720–726. <https://doi.org/10.1016/j.tins.2004.10.004>
- Raymaker, D. M., Teo, A. R., Steckler, N. A., Lentz, B., Scharer, M., Delos Santos, A., Kapp, S. K., Hunter, M., Joyce, A., & Nicolaidis, C. (2020). “Having All of Your Internal Resources Exhausted Beyond Measure and Being Left with No Clean-Up Crew”: Defining Autistic Burnout. *Autism in Adulthood*, 2(2), 132–143. <https://doi.org/10.1089/aut.2019.0079>
- 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>
- Rezlescu, C., Susilo, T., Wilmer, J. B., & Caramazza, A. (2017). The inversion, part-whole, and composite effects reflect distinct perceptual mechanisms with varied relationships to face recognition. *Journal of Experimental Psychology: Human Perception and Performance*, 43(12), 1961–1973. <https://doi.org/10.1037/xhp0000400>
- Rhodes, G., & Tremewan, T. (1993). The Simon Then Garfunkel Effect: Semantic Priming, Sensitivity, and the Modularity of Face Recognition. *Cognitive Psychology*, 25(2), 147–187. <https://doi.org/10.1006/cogp.1993.1004>
- Richler, J. J., & Gauthier, I. (2014). A Meta-Analysis and Review of Holistic Face Processing. *Psychological Bulletin*, 140(5), 1281–1302. <https://doi.org/10.1037/a0037004>
- Riddoch, M. J., & Humphreys, G. W. (1993). *Birmingham object recognition battery*. Lawrence Erlbaum Associates.
- Rimmele, U., Hediger, K., Heinrichs, M., & Klaver, P. (2009). Oxytocin makes a face in memory familiar. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, 29(1), 38–42. <https://doi.org/10.1523/JNEUROSCI.4260-08.2009>
- Ritchie, K. L., & Burton, A. M. (2017). Learning faces from variability. *Quarterly Journal of Experimental Psychology*, 70(5), 897–905. <https://doi.org/10.1080/17470218.2015.1136656>

- Rivolta, D., Schmalzl, L., Coltheart, M., & Palermo, R. (2010). Semantic information can facilitate covert face recognition in congenital prosopagnosia. *Journal of Clinical and Experimental Neuropsychology*, 32(9), 1002–1016. <https://doi.org/10.1080/13803391003662710>
- Robotham, R. J., & Starrfelt, R. (2018). Tests of whole upright face processing in prosopagnosia: A literature review. *Neuropsychologia*, 121(Journal Article), 106–121. <https://doi.org/10.1016/j.neuropsychologia.2018.10.018>
- Rosenthal, G., & Avidan, G. (2018). A possible neuronal account for the behavioural heterogeneity in congenital prosopagnosia. *Cognitive Neuropsychology*, 35(1–2), 74–77. <https://doi.org/10.1080/02643294.2017.1417248>
- Rossion, B. (2018). Prosopdysgnosia? What could it tell us about the neural organization of face and object recognition? *Cognitive Neuropsychology*, 35(1–2), 98–101. <https://doi.org/10.1080/02643294.2017.1414778>
- Rossion, B., & Michel, C. (2018). Normative accuracy and response time data for the computerized Benton Facial Recognition Test (BFRT-c). *Behavior Research Methods*, 50(6), 2442–2460. <https://doi.org/10.3758/s13428-018-1023-x>
- 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>
- Sagiv, N., & Bentin, S. (2001). Structural Encoding of Human and Schematic Faces: Holistic and Part-Based Processes. *Journal of Cognitive Neuroscience*, 13(7), 937–951. <https://doi.org/10.1162/089892901753165854>
- Sasson, N. J. (2006). The development of face processing in autism. *Journal of Autism and Developmental Disorders*, 36(3), 381–394. <https://doi.org/10.1007/s10803-006-0076-3>
- Savaskan, E., Ehrhardt, R., Schulz, A., Walter, M., & Schächinger, H. (2008). Post-learning intranasal oxytocin modulates human memory for facial identity. *Psychoneuroendocrinology*, 33(3), 368–374.
<https://doi.org/10.1016/j.psyneuen.2007.12.004>
- Schmalzl, L., Palermo, R., & Coltheart, M. (2006). Congenital prosopagnosia: A family study. *Australian Journal of Psychology*, 58(Journal Article), 89–89.
- Schmalzl, L., Palermo, R., & Coltheart, M. (2008). Cognitive heterogeneity in genetically based prosopagnosia: A family study. *Journal of Neuropsychology*, 2(Journal Article), 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>

- Scott, L. S., Pascalis, O., & Nelson, C. A. (2007). A Domain-General Theory of the Development of Perceptual Discrimination. *Current Directions in Psychological Science*, 16(4), 197–201. <https://doi.org/10.1111/j.1467-8721.2007.00503.x>
- 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>
- Skuse, D. H., Lori, A., Cubells, J. F., Lee, I., Conneely, K. N., Puura, K., Lehtimäki, T., Binder, E. B., & Young, L. J. (2014). Common polymorphism in the oxytocin receptor gene (OXTR) is associated with human social recognition skills. *Proceedings of the National Academy of Sciences of the United States of America*, 111(5), 1987–1992. <https://doi.org/10.1073/pnas.1302985111>
- Sorensen, T. A., & Overgaard, M. S. (2018). Prosopagnosia or Prosopdysgnosia Facing up to a change of concepts. In *2018 10th International Conference on Knowledge and Smart Technology (kst 2018)—Cybernetics in the Next Decades* (pp. 260–263). Ieee.
- Sorger, B., Goebel, R., Schiltz, C., & Rossion, B. (2007). Understanding the functional neuroanatomy of acquired prosopagnosia. *Neuroimage*, 35(2), 836–852. <https://doi.org/10.1016/j.neuroimage.2006.09.051>

- Spencer, J., & Ritchie, L. (1994). Qualitative data analysis for applied policy research. In *Analyzing Qualitative Data*. Routledge.
- Stantic, M., Brewer, R., Duchaine, B., Banissy, M. J., Bate, S., Susilo, T., Catmur, C., & Bird, G. (2022). The Oxford Face Matching Test: A non-biased test of the full range of individual differences in face perception. *Behavior Research Methods*, 54(1), 158–173. <https://doi.org/10.3758/s13428-021-01609-2>
- Stantić, M., Pounder, Z., Bate, S., Susilo, T., Catmur, C., & Bird, G. (2022). Individuals with developmental prosopagnosia show independent impairments in face perception, face memory and face matching. *Cortex*, 157, 266–273. <https://doi.org/10.1016/j.cortex.2022.09.012>
- Steeves, J. K. E., Culham, J. C., Duchaine, B. C., Pratesi, C. C., Valyear, K. F., Schindler, I., Humphrey, G. K., Milner, A. D., & Goodale, M. A. (2006). The fusiform face area is not sufficient for face recognition: Evidence from a patient with dense prosopagnosia and no occipital face area. *Neuropsychologia*, 44(4), 594–609. <https://doi.org/10.1016/j.neuropsychologia.2005.06.013>
- 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–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>

- Sunday, M. A., Patel, P. A., Dodd, M. D., & Gauthier, I. (2019). Gender and hometown population density interact to predict face recognition ability. *Vision Research*, 163, 14–23. <https://doi.org/10.1016/j.visres.2019.08.006>
- Susilo, T., & Duchaine, B. (2013). Advances in developmental prosopagnosia research. *Current Opinion in Neurobiology*, 23(3), 423–429. <https://doi.org/10.1016/j.conb.2012.12.011>
- Susilo, T., McKone, E., Dennett, H., Darke, H., Palermo, R., Hall, A., Pidcock, M., Dawel, A., Jeffery, L., Wilson, C. E., & Rhodes, G. (2010). Face recognition impairments despite normal holistic processing and face space coding: Evidence from a case of developmental prosopagnosia. *Cognitive Neuropsychology*, 27(8), 636–664. <https://doi.org/10.1080/02643294.2011.613372>
- Svart, N., & Starrfelt, R. (2022). Is It Just Face Blindness? Exploring Developmental Comorbidity in Individuals with Self-Reported Developmental Prosopagnosia. *Brain Sciences*, 12(2), Article 2. <https://doi.org/10.3390/brainsci12020230>
- Tanaka, J. W., & Farah, M. J. (1993). Parts and wholes in face recognition. *The Quarterly Journal of Experimental Psychology A: Human Experimental Psychology*, 46A(2), 225–245. <https://doi.org/10.1080/14640749308401045>
- The jamovi project* (Version 2.3). (2023). [Computer software]. <https://jamovi.org>
- 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>

- Tsantani, M., & Cook, R. (2020). Normal recognition of famous voices in developmental prosopagnosia. *Scientific Reports*, 10. <https://doi.org/10.1038/s41598-020-76819-3>
- Tsantani, M., Gray, K. L. H., & Cook, R. (2020). Holistic processing of facial identity in developmental prosopagnosia. *Cortex; a Journal Devoted to the Study of the Nervous System and Behavior*, 130, 318–326. <https://doi.org/10.1016/j.cortex.2020.06.003>
- Tsantani, M., Vestner, T., & Cook, R. (2021). The Twenty Item Prosopagnosia Index (PI20) provides meaningful evidence of face recognition impairment. *Royal Society Open Science*, 8(11), 202062. <https://doi.org/10.1098/rsos.202062>
- Turati, C., Sangrigoli, S., Ruelly, J., & de Schonen, S. (2004). Evidence of the Face Inversion Effect in 4-Month-Old Infants. *Infancy: The Official Journal of the International Society on Infant Studies*, 6(2), 275–297. https://doi.org/10.1207/s15327078in0602_8
- 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>
- Uvnäs-Moberg, K., Ekström-Bergström, A., Berg, M., Buckley, S., Pajalic, Z., Hadjigeorgiou, E., Kotłowska, A., Lengler, L., Kielbratowska, B., Leon-Larios, F., Magistretti, C. M., Downe, S., Lindström, B., & Dencker, A. (2019). Maternal plasma levels of oxytocin during physiological childbirth – a systematic review with implications for uterine contractions and central actions of oxytocin. *BMC*

Pregnancy and Childbirth, 19(1), 285. [https://doi.org/10.1186/s12884-019-2365-](https://doi.org/10.1186/s12884-019-2365-9)

9

- Ventura, P., Livingston, L. A., & Shah, P. (2018). Adults have moderate-to-good insight into their face recognition ability: Further validation of the 20-item Prosopagnosia Index in a Portuguese sample. *Quarterly Journal of Experimental Psychology*, 71(12), 2677–2679. <https://doi.org/10.1177/1747021818765652>
- Verfaillie, K., Huysegems, S., De Graef, P., & Van Belle, G. (2014). Impaired holistic and analytic face processing in congenital prosopagnosia: Evidence from the eye-contingent mask/window paradigm. *Visual Cognition*, 22(3–4), 503–521. <https://doi.org/10.1080/13506285.2014.881446>
- 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*, 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>
- Volfart, A., & Rossion, B. (2024). The neuropsychological evaluation of face identity recognition. *Neuropsychologia*, 198, 108865. <https://doi.org/10.1016/j.neuropsychologia.2024.108865>
- White, D., Rivolta, D., Burton, A. M., Al-Janabi, S., & Palermo, R. (2017). Face matching impairment in developmental prosopagnosia. *The Quarterly Journal of Experimental Psychology*, 70(2), 287–297. <https://doi.org/10.1080/17470218.2016.1173076>
- Whittington, J., Holland, A., Webb, T., Butler, J., Clarke, D., & Boer, H. (2004). Cognitive abilities and genotype in a population-based sample of people with

- Prader-Willi syndrome. *Journal of Intellectual Disability Research: JIDR*, 48(Pt 2), 172–187. <https://doi.org/10.1111/j.1365-2788.2004.00556.x>
- Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L., François, R., Golemund, 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>
- Wilmer, J. B. (2017). Individual Differences in Face Recognition: A Decade of Discovery. *Current Directions in Psychological Science*, 26(3), 225–230. <https://doi.org/10.1177/0963721417710693>
- Wilmer, J. B., Germine, L., Chabris, C. F., Chatterjee, G., Williams, M., Loken, E., Nakayama, K., & Duchaine, B. (2010). Human face recognition ability is specific and highly heritable. *Proceedings of the National Academy of Sciences of the United States of America*, 107(11), 5238–5241. <https://doi.org/10.1073/pnas.0913053107>
- Yardley, L., McDermott, L., Pisarski, S., Duchaine, B., & Nakayama, K. (2008). Psychosocial consequences of developmental prosopagnosia: A problem of recognition. *Journal of Psychosomatic Research*, 65(5), 445–451. <https://doi.org/10.1016/j.jpsychores.2008.03.013>

- 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. (1998). Covert face recognition in prosopagnosia. In A. W. Young (Ed.), *Face and Mind* (p. 0). Oxford University Press. <https://doi.org/10.1093/acprof:oso/9780198524205.003.0011>
- 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>
- Young, A. W., Hay, D. C., & Ellis, A. W. (1985). The faces that launched a thousand slips: Everyday difficulties and errors in recognizing people. *British Journal of Psychology* (London, England: 1953), 76 (Pt 4), 495–523. <https://doi.org/10.1111/j.2044-8295.1985.tb01972.x>
- Young, A. W., Hellawell, D., & Hay, D. C. (1987). Configurational Information in Face Perception. *Perception*, 16(6), 747–759. <https://doi.org/10.1068/p160747>
- Yovel, G., & Duchaine, B. (2006). Specialized Face Perception Mechanisms Extract Both Part and Spacing Information: Evidence from Developmental Prosopagnosia. *Journal of Cognitive Neuroscience*, 18, 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(Journal Article), 151–161. <https://doi.org/10.1016/j.neuroimage.2017.12.023>

- Zhen, Z., Fang, H., & Liu, J. (2013). The Hierarchical Brain Network for Face Recognition. *PLOS ONE*, 8(3), e59886. <https://doi.org/10.1371/journal.pone.0059886>
- Zhu, Q., Song, Y., Hu, S., Li, X., Tian, M., Zhen, Z., Dong, Q., Kanwisher, N., & Liu, J. (2010). Heritability of the specific cognitive ability of face perception. *Current Biology: CB*, 20(2), 137–142. <https://doi.org/10.1016/j.cub.2009.11.067>

Appendices

Appendix 1: Chapter 3 supplementary materials

1. Methods: Task instructions

The full wording of the on-screen participant instructions was as follows:

“In this task you will be shown a 3 x 3 grid of patterns. The last one in the bottom right corner will be missing. You need to select which of the four possible patterns along the bottom fits into the gap. Try to be as fast and accurate as you can. If you cannot solve the puzzle then you should guess. You will not be penalised for an incorrect answer. You will have 30 seconds to complete each puzzle”.

2. Normality

Mean accuracy and median RT data were inspected for normality. Both were distributed normally in each group (Shapiro Wilk $p = .21$ control group and $p = .45$ in the DP group. Skew and Kurtosis were also acceptable (± 0.76).

3. Individual DP performance

Except for CF059, AF017 and AF001, the remaining 18/21 DPs (85.7%) were *more accurate* than age matched controls.

Except for CF059, AF016, AF017 and AF001, the remaining 17/21 DPs (81.0%) were *slower* to respond than controls.

The clear group differences observed are therefore clearly *not* driven by a small number of DPs with extreme scores.

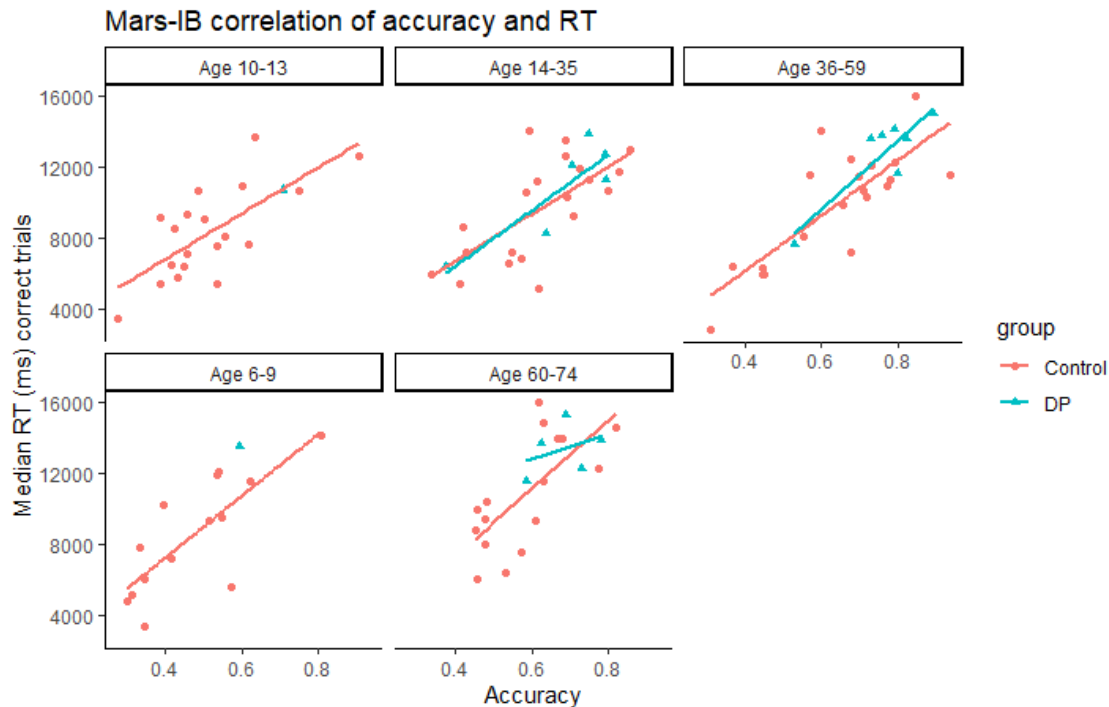
Table S1 DP individual level performance vs the relevant age group control means

ID	Age	Gender	Accuracy (proportion correct)	Age matched control accuracy		Median RT (ms)	Age matched control median RT (ms)	
				Mean	SD		Mean	SD
CF007	8	Female	0.59	0.47	0.15	12,131	7,417	3,260
CF008	10	Male	0.71	0.52	0.14	8,526	6,691	2,385
CF059	16	Female	0.19	0.62	0.15	2,071	8,593	2,925
CF005	17	Male	0.79	0.62	0.15	12,424	8,593	2,925
AF016	23	Female	0.64	0.62	0.15	8,299	8,593	2,925
AF002	25	Female	0.71	0.62	0.15	13,853	8,593	2,925
AF017	29	Female	0.38	0.62	0.15	6,103	8,593	2,925
AF004	31	Male	0.75	0.62	0.15	11,750	8,593	2,925
AF007	31	Female	0.79	0.62	0.15	11,800	8,593	2,925
AF009	42	Male	0.79	0.63	0.16	12,814	8,873	3,231
AF013	44	Female	0.76	0.63	0.16	14,060	8,873	3,231
AF075	45	Female	0.80	0.63	0.16	9,678	8,873	3,231
AF018	49	Female	0.73	0.63	0.16	9,961	8,873	3,231
AF001	51	Female	0.53	0.63	0.16	5,546	8,873	3,231
AF006	54	Female	0.82	0.63	0.16	14,599	8,873	3,231
AF003	55	Female	0.89	0.63	0.16	16,077	8,873	3,231
AF060	61	Female	0.63	0.58	0.12	12,676	9,958	3,231
AF022	64	Female	0.73	0.58	0.12	11,804	9,958	3,231
AF098	67	Male	0.59	0.58	0.12	10,025	9,958	3,231
AF099	70	Male	0.78	0.58	0.12	14,013	9,958	3,231
AF008	71	Female	0.69	0.58	0.12	17,389	9,958	3,231

4. Do DPs and age matched controls show similar speed accuracy trade-offs?

4.1 Age group level

Figure S1. Correlation between accuracy and RT by group and age group.



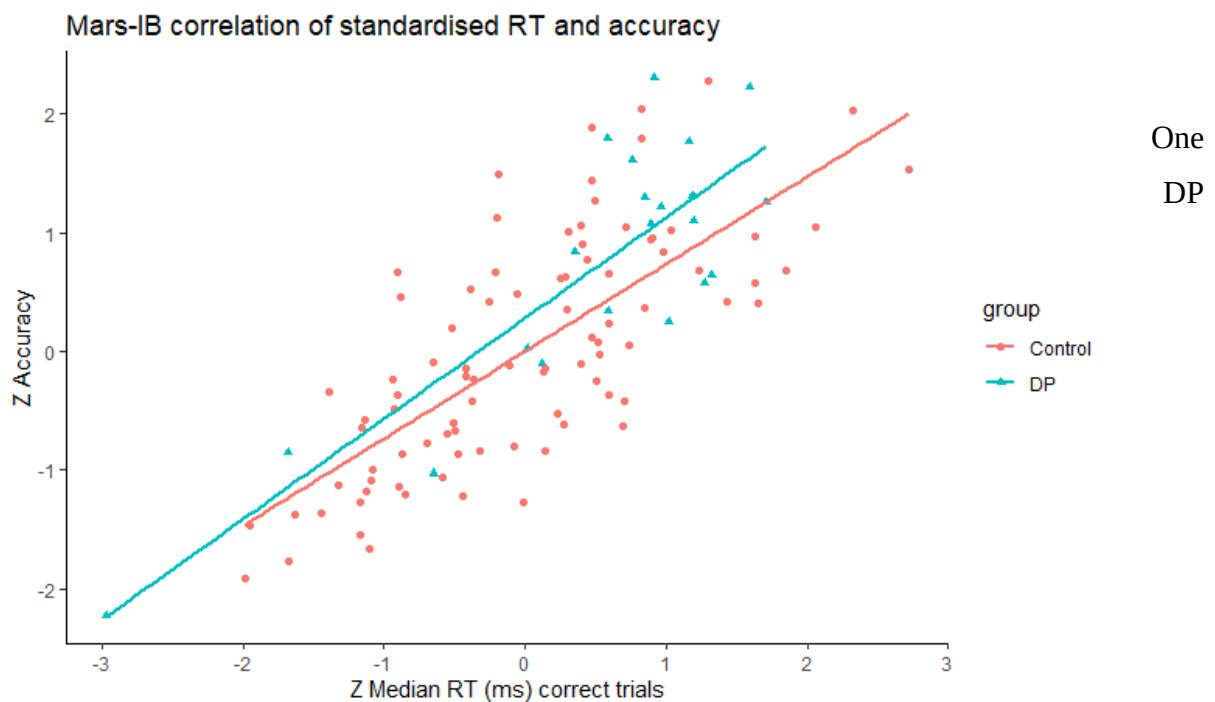
Note: One DP participant (age 16) who had very low accuracy and very fast RT has been removed for correlational analysis because they were strongly influencing the correlation

- The key point of interest here is that, at least in the adult age groups where there was more than one DP per group, the majority of DPs appear to cluster in the top right of each plot (higher than average accuracy and RT) and therefore show a strikingly different performance pattern vs age matched controls despite overlap in scores/RT between groups.
- Speed accuracy trade-offs were highly variable at the individual level and also varied across age group thus confirming the importance of considering performance at the age group rather than simply overall sample level (by centring z scores on age group control means as we have done).

4.2 At the overall group level

- To enable comparison across age groups, accuracy and RT were converted to z scores (using *age matched* control means and *SD*) .
- As shown in Figure S2 below, the correlation was stronger in the DP group $r(19) = .843$ [.65, .93], $p < .001$ than in the control group $r(88) = .735$ [.62, .82], $p < .001$, but the difference was not significant, Fisher's $Z = -1.13$, $p = 0.26$. However, importantly, note that at the group level the different patterns across age groups (seen in Figure 1 above) is obscured.

Figure S2. Correlation of standardised accuracy and RT



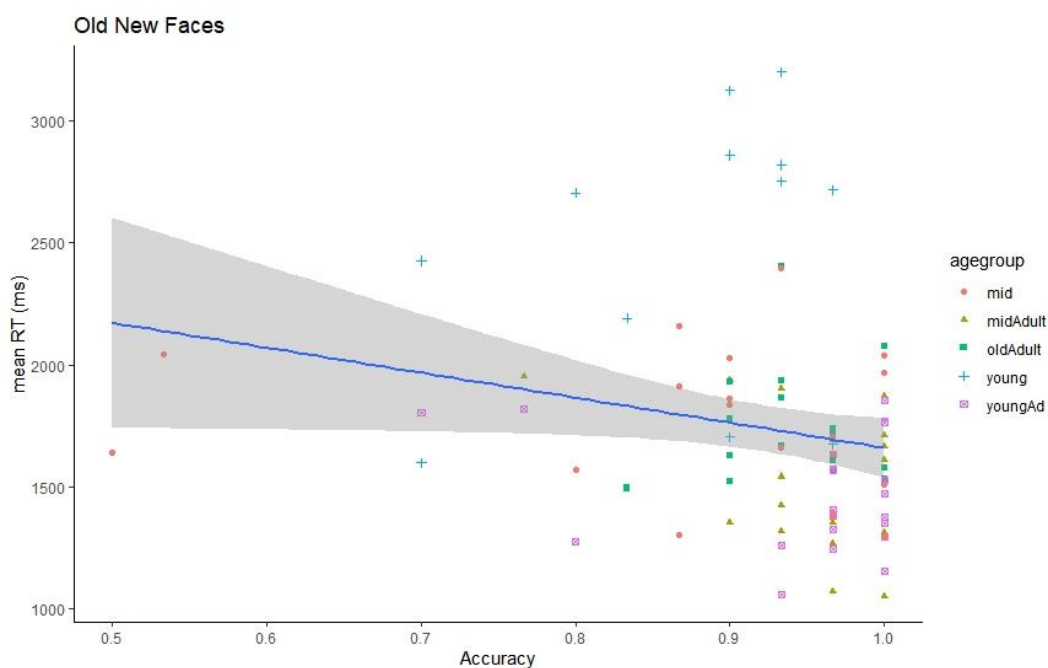
achieved very low accuracy but had very fast RT. When this DP was removed the DP group correlation $r(18) = .737$ [.47, .89], $p < .001$ was almost identical to controls ($r = .735$).

Appendix 2: Chapter 4 supplementary materials

1. Correlational analysis control accuracy and mean RT (correct trials only)

1.1 Old New Faces

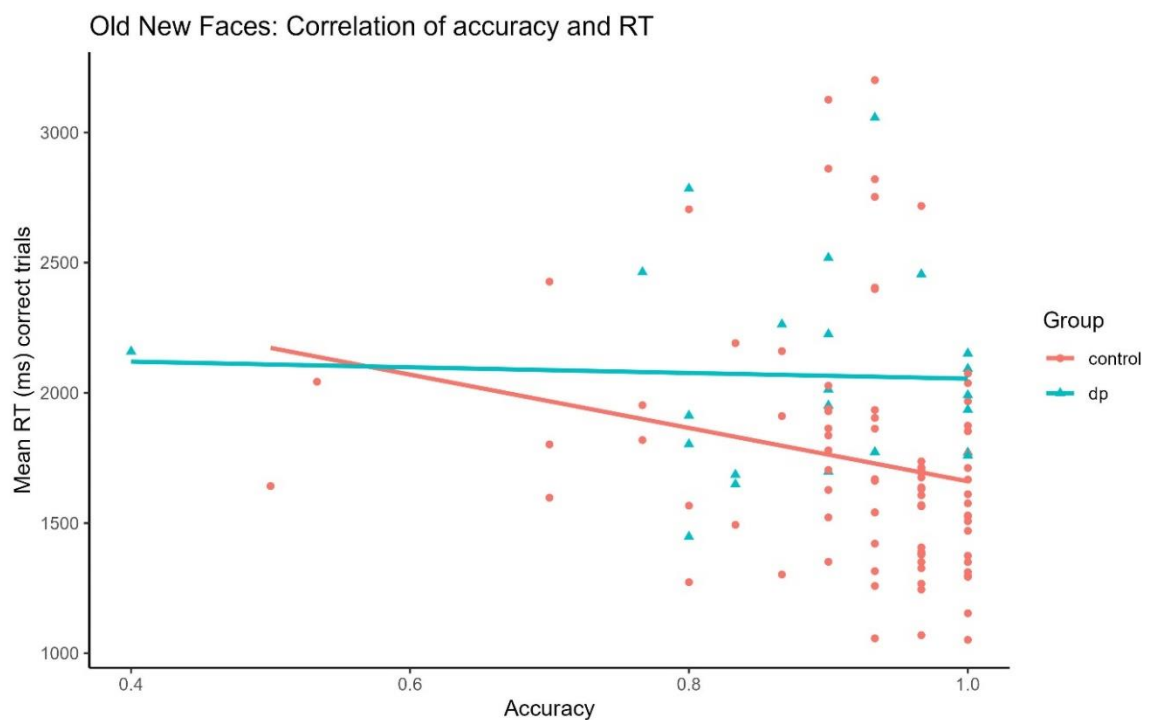
- At an overall group level controls showed a significant negative correlation between accuracy and RT $r(85) = -.216, p = .044$. In other words, as accuracy increased then RT decreased with more accurate participants being also quicker to respond. On this task, no speed-accuracy trade-offs were observed among controls, except in the young children age group (“young”) as illustrated below



- Young child controls (6-9 years) showed a different pattern to all other age groups (here, accuracy *increased* with RT indicating speed-accuracy trade off). When removing the youngest age group, the effect size for the accuracy/RT correlation among controls increased to $r(73) = -.229, p = .048$
- Because the data were not linear, a quadratic regression was performed (squared age) and age explained 29.6% of the variance in RT, $F(2, 84) = 17.63, p = <.001$, multiple $R^2 = .296$
- The association between mean RT (correct trials) and accuracy was inspected in each control age group

Young 6-9 years	$r(10) = .327, p = .299$ (higher scorers were slower)
Mid 10-13 years	$r(16) = -.104, p = .682$ (higher scorers were faster)
Young Adults 14 -35 years	$r(19) = -.237, p = .301$ (higher scorers were faster)
Mid adults 36-59 years	$r(17) = -.366, p = .123$ (higher scorers were faster)
Old adults ≥ 60 years	$r(15) = .068, p = .795$ (higher scorers were marginally slower)

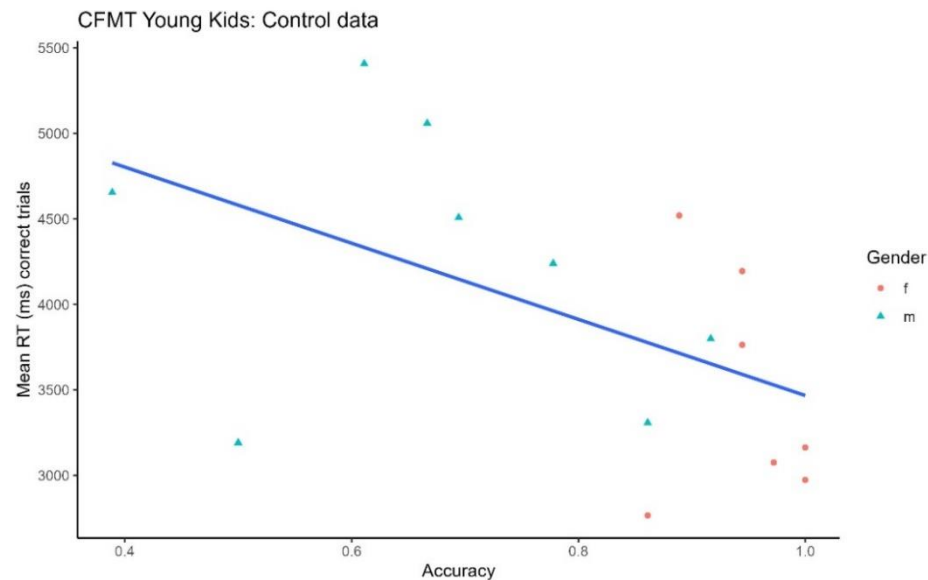
- Among DPs no relationship was observed between accuracy and RT $r(21) = -.04, p = .869$ – see below



1.2 Cambridge Face Memory Test

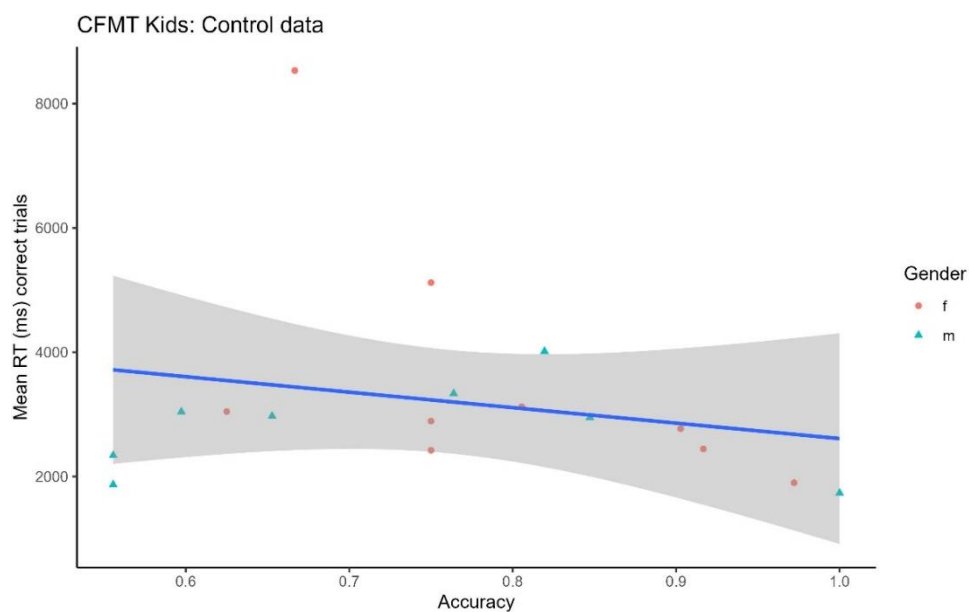
CFMT Young Kids: Children 6-9 years

- Among controls a strong significant negative relationship was observed between score and RT $r(13) = -.516, p = .049$. In other words, on average, more accurate participants were also quicker to respond.

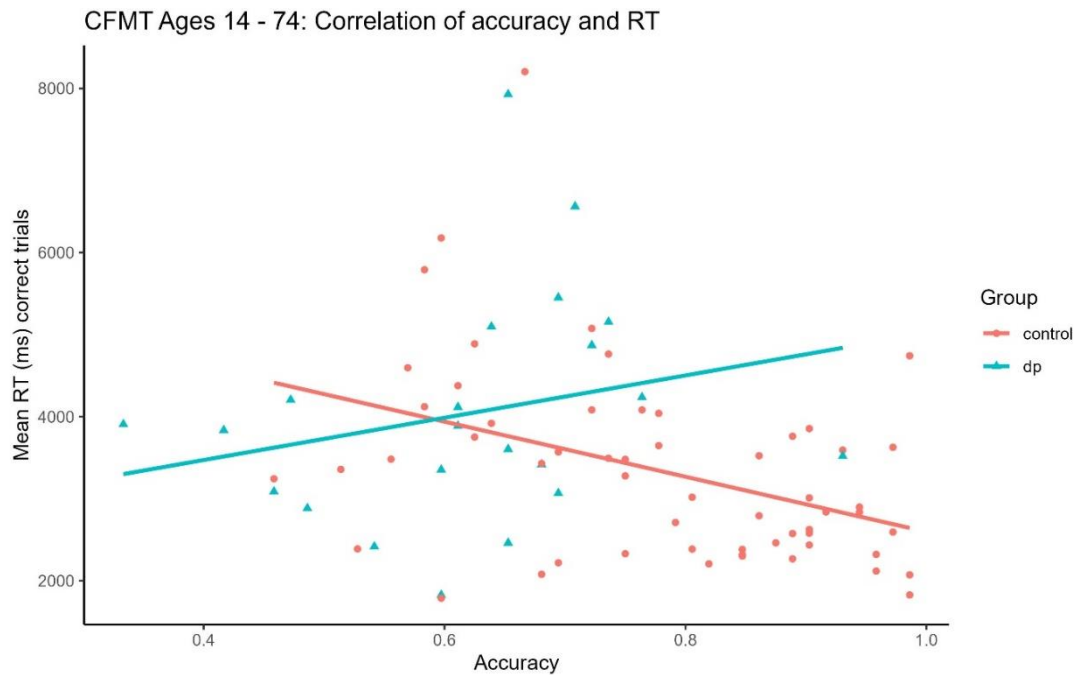


CFMT Kids: Children 10-13 years

- Among controls there was a moderate negative but non-significant correlation between score and RT, $r(15) = -.217, p = .403$. In other words, on average, more accurate participants were also quicker to respond.



CFMT Adults 14-74



- Among controls there was a significant negative correlation between accuracy and RT, $r(56) = -.406$, $p = .002$. In other words, on average more accurate participants were also quicker to respond indicating no overall speed accuracy trade off.
- Among DPs the opposite trend was observed (see figure above) with more accurate participants responding, on average, more slowly. In other words, at group level DPs appeared to be achieving higher accuracy scores as a result of taking longer to respond, $r(20) = 0.242$, $p = .278$.
- The relationship between accuracy and mean RT (correct trials) was inspected in each control age group

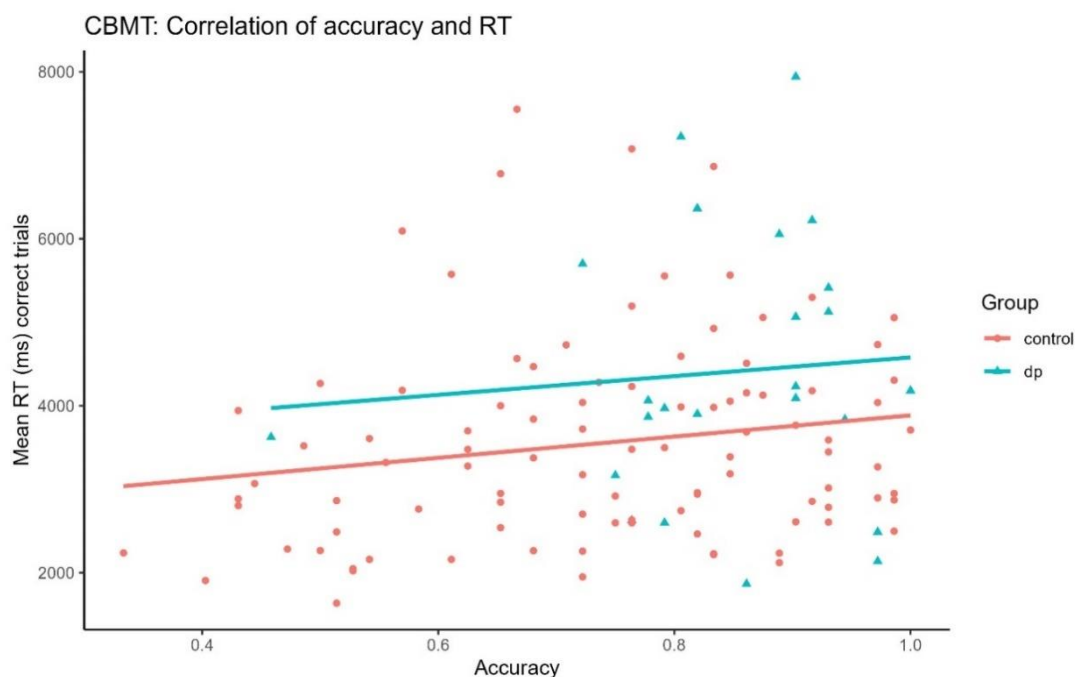
Young Adults 14 -35 years $r(19) = -.087$, $p = .709$ (higher scorers were faster)

Mid adults 36-59 years $r(18) = -.560$, $p = .010$ (higher scorers were faster)

Old adults ≥ 60 years $r(15) = -.327$, $p = .200$ (higher scorers were faster)

1.3 Cambridge Bicycle Memory Test

- Overall, on this task, no significant speed-accuracy trade-offs were observed among controls. At an overall group level controls showed a small but non-significant positive correlation between accuracy and RT $r(95) = .173$ $p = .09$. See below. In other words, as accuracy increased so did RT with more accurate participants being, on average, slower to respond.
- The same pattern was observed for the DP group, with a very small non-significant positive association observed between accuracy and RT $(22) = 0.080$, $p = .711$



- The association between mean RT (correct trials) and accuracy was inspected in each control age group

Young 6-9 years $r(13) = .475$, $p = .074$ (higher scorers were slower)

Mid 10-13 years $r(19) = .368$, $p = .101$ (higher scorers were slower)

Young Adults 14 -35 years $r(19) = -.236$, $p = .303$ (higher scorers were **faster**)

Mid adults 36-59 years $r(20) = -.127$, $p = .572$ (higher scorers were **faster**)

Old adults ≥ 60 years $r(16) = .242$, $p = .331$ (higher scorers were slower)

- Child and older adult controls showed a different pattern to all other age groups (here, accuracy *increased* with RT indicating speed-accuracy trade off).

Conclusion: Although in our study we controlled for different patterns of performance between age groups by centring all z scores on age group control means, these age group data reinforce the need to account for differences in speed accuracy trade-off between participants (here controls of different ages).

2. Effect of removing Subjective DPs: rerunning key analyses

In our main analysis we included all participants with subjective face recognition complaints. Here, as suggested during peer review, we present the key between-group analyses shown in Table 8, this time with all ten participants who showed no objective accuracy impairment (i.e. those participant classified as subjective DPs using the global face memory accuracy score) and including adult participants only. This revised sample comprises 79 participants (DPs $n = 13$, Controls $n = 66$).

Table S1 Revised comparison of effect sizes for the difference between DP and control group scores before (Accuracy) and after (BIS) accounting for RT

No subjective DPs or kids	Accuracy			Balanced Integration Score		
	<i>BF10</i>	<i>p</i>	<i>d</i>	<i>BF10</i>	<i>p</i>	<i>d</i>
Old New Faces	88.3	.064	-0.81	18,519	.005**	-1.29
CFMT	52,693	< .001***	-1.82	403	< .001***	-1.36
CBMT	0.44	.389	0.28	0.38	.455	0.23
CMT Difference	2,061	< .001***	-1.64	790	< .001***	-1.54
FFT Identification	3,741,613	< .001***	-1.91	-	-	-
Global Memory	1.48e+8	< .001***	-2.11	1.01e+8	< .001***	-2.20

Importantly, the pattern of results is similar to that seen when analysing the entire sample (Table 8). Other key points to note are first that the global face memory effect sizes remain larger for BIS than for accuracy. The second point of interest is that on the tests with the highest mean accuracy e.g. old New Faces, significant group differences are observed for BIS even when mean accuracy scores are not significantly different. In summary, removing all DPs who

showed no objective global face memory accuracy impairment (i.e. those who were able to achieve high accuracy likely due to extended RTs) does not change the pattern of results found in the main analyses.

3. Effect of removing Subjective DPs: Logistic regression

We also reran the logistic regression analyses (see Table S3 below) to assess which models best predicted group when considering this smaller sub sample from which children and subjective DPs had been removed. Findings were that once again, the pattern of results mirrored those from the full sample. As before, the model that best predicted group was Model 2, the model comprising the individual BIS measures. This finding confirms that, even when considering *only* those DPs who show clear accuracy impairments (and ignoring for now those whose impairment only become apparent when RT was also considered) the model comprising separate BIS measures (Model 2) outperformed the model comprising separate accuracy measures BIS (Model 1) for classification thus confirming the utility of BIS for classifying DP.

Table S3

Revised logistic regression models predicting group membership DP or Control

Models	Predictors	Overall model				AUC	% correctly classified	
		χ^2	p	Adj R^2 McF ^a	Adj R^2 N ^b		DP	Control
1	Accuracy: CFMT, CMT Difference, Old New Faces, FFT	47.4	<.001	0.762	0.84	0.90	83.3	96.2
2	BIS: Old New Faces, CFMT; FFT accuracy,	46.0	<.001	0.74	0.82	0.98	91.7	98.1
3	Global memory accuracy	41.7	<.001	0.67	0.77	0.97	75.0	96.2
4	Global memory BIS	35.8	<.001	0.58	0.69	0.96	66.7	94.3
5	CFMT accuracy (all)	22.5	<.001	0.36	0.48	0.88	41.7	90.6

^aAdj R^2 McF = McFaddens R^2 ; ^bAdj R^2 N = Nagkerkerke's R^2

Model 2 Coefficients - group

Predictor	Estimate	SE	Z	p	Odds ratio	95% Confidence Interval	
						Lower	Upper
Intercept	-8.65	3.62	-2.39	0.017	1.75 e-4	1.44 e-7	0.21
Old New Faces BIS	-1.49	0.75	-2.00	0.045	0.22	0.05	0.97
CFMT BIS	1.50	1.01	1.48	0.140	4.47	0.61	32.62
FFT	-3.07	1.47	-2.08	0.037	0.05	0.00	0.84
CMT Difference BIS	-1.73	0.82	-2.10	0.036	0.18	0.04	0.89

Note. Estimates represent the log odds of "group = DP" vs. "group = Control"

4. Correlational matrices

Table S4

Correlational matrices

Whole sample	Old New accuracy	Old New BIS	CFMT accuracy	CFMT BIS	FFT	CBMT accuracy	CBMT BIS
Old New accuracy	-	-					
Old New BIS	-	-					
CFMT accuracy	.04***	.45***	-	-			
CFMT BIS	.24**	.40***	-	-			
FFT	.30**	.42***	.43***	.30***	-		
CBMT accuracy	.13	.02	.11	.08	-.03	-	-
CBMT BIS	-.02	.11	.06	.32***	-.03	-	-
Self report	.33***	.51***	.46***	.42**	.44***	-0.08	0.09
Controls only							
	Old New accuracy	Old New BIS	CFMT accuracy	CFMT BIS	FFT	CBMT accuracy	CBMT BIS
Old New accuracy	-						
Old New BIS	-	-					
CFMT accuracy	.23*	.25*	-				
CFMT BIS	.20	.32**	-	-			
FFT	.11	.20	.33**	.19	-		
CBMT accuracy	.21	.16	.13	.12	.08	-	
CBMT BIS	.13	.24*	.09	.31	-.08	-	-
Self report	.20	.14	.32**	.26*	.01	.16	

Note: *** $p < .001$, ** $p < .01$, * $p < .05$. Correlation of standardised scores that were centred on age-group controls means thus controlling for age effects.

5. Mathematical formula for calculating BIS

$$BIS = Z_{i,j}^{PC} - Z_{i,j}^{\overline{RT}} = \frac{PC_{i,j} - \overline{PC}}{SD^{PC}} - \frac{\overline{RT}_{i,j} - \overline{\overline{RT}}}{SD^{RT}}$$

Where $Z_{i,j}$ is the standardised performance centred on age-group control means, (mean percentage correct and mean RT on correct trials respectively). $\overline{\overline{RT}}$ refers to the grand mean RT (again, correct trials only) and SD refers to the standard deviation.

Appendix 3: Chapter 5 supplementary materials

Mathematical formula for calculating BIS

$$BIS = Z_{i,j}^{PC} - Z_{i,j}^{\overline{RT}} = \frac{PC_{i,j} - \overline{PC}}{SD^{PC}} - \frac{\overline{RT}_{i,j} - \overline{\overline{RT}}}{SD^{\overline{RT}}}$$

Where $Z_{i,j}$ is the standardised performance centred on age-group control means, (mean percentage correct and mean RT on correct trials respectively). $\overline{\overline{RT}}$ refers to the grand mean RT (again, correct trials only) and SD refers to the standard deviation.

Table S1 Descriptive statistics Faces among non faces including outlier

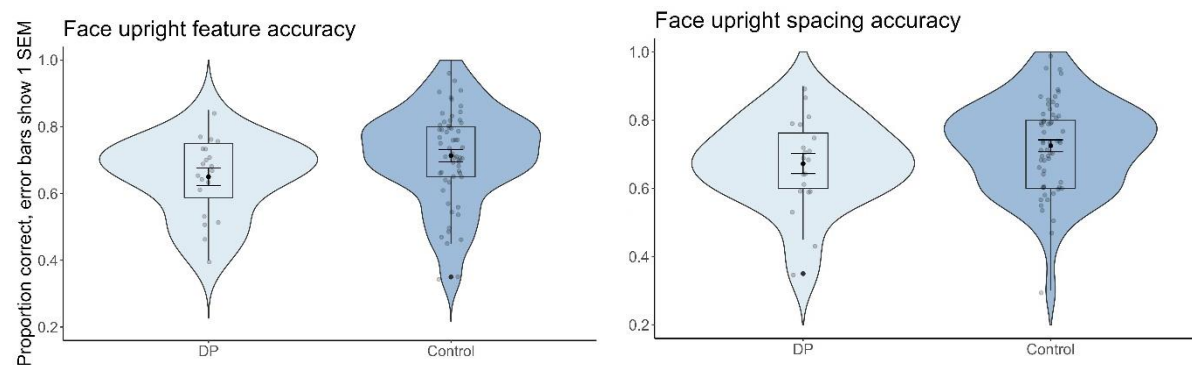
RT (ms)	DP			Control		
	<i>N</i>	<i>M</i>	<i>SD</i>	<i>N</i>	<i>M</i>	<i>SD</i>
14 - 35 years	7	1106	502	21	936	222
36 - 59 years	8	1544	786	21	1136	304
60 - 74 years	5	1795	351	18	1340	377
Total	20	1453	642	60	1127	340
BIS						
Total	20	-1.40	3.61	60	<.001	1.65

Note. Mean RT in ms calculated on correct, upright target-present trials. Because BIS is a standardised measure, BIS scores for controls are 0 or very close to 0 by definition.

Configural featural test instructions

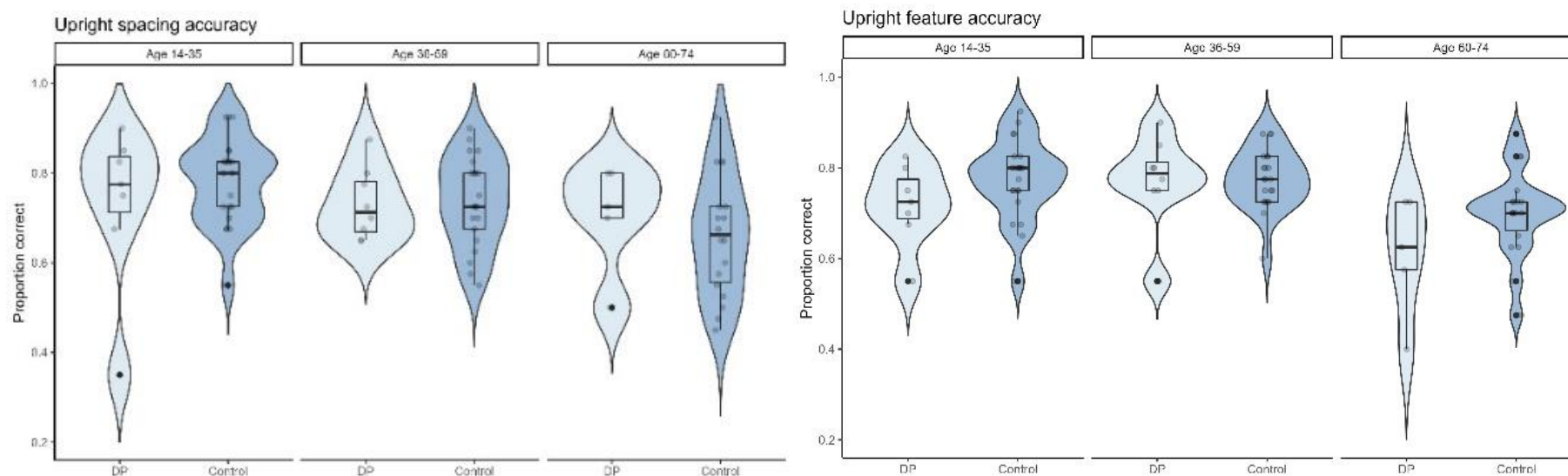
The test began with two practice trials using cartoon faces to familiarise participants with the task. Next participants were told “*The faces you will see in the test are trickier because they are real faces and the differences between them are very small. The different faces are a bit like twins, they look almost the same but there is a small difference that you will have to look hard to spot. You will now see some examples of faces that only have very small differences*”. The test instructions were as follows: “*Your task is to decide if the pairs of faces that will appear very quickly one after the other are the same or different. If the two faces are exactly the same press “S” and if they are NOT exactly the same press “N”. Be as fast and accurate as you can. If you are not sure, just try your best. Get your fingers ready next to S and N and when you are ready press START.*”

Figure S1 Comparing group performance discriminating spacing and feature changes: Configural featural task



Note: Unstandardised proportion correct by group (DP versus control). Spacing accuracy = overall accuracy for detecting configural changes to faces and houses combined, feature accuracy = overall accuracy for detecting features changes to face and houses combined. Each dot represents a single data point, the box shows the interquartile range (IQR) and the midline indicates the group median score. The end of each whisker line represents 1.5 x the IQR.

Figure S2 Comparing group performance discriminating spacing and feature changes by age group



Note: Unstandardised proportion correct by group (DP versus control) by age group. Spacing accuracy = overall accuracy for detecting configural changes to faces and houses combined, feature accuracy = overall accuracy for detecting features changes to face and houses combined. Each dot represents a single data point, the box shows the interquartile range (IQR) and the midline indicates the group median score. The end of each whisker line represents 1.5 x the IQR.

Table S2 Benjamini Hochberg adjusted p values

Measure	Original <i>p</i> value	Rank	B-H adjusted alpha threshold	BH sig?	BF ₁₀
CFPT	.001	1	.004	Yes	31.2
CFPT normalised inversion index	.005	2	.008	Yes	1.71
Configural Featural: Face Advantage	.005	3	.012	Yes	2.49
Faces among non faces RT	.008	4	.015	Yes	1.88
Gender Categorisation BIS	.015	5	.019	Yes	1.49
Faces among non faces RT inversion index normalised	.020	6	.023	Yes	1.08
Configural Featural: Face accuracy	.021	7	.027	Yes	2.21
Mooney Faces	.062	8	.031	No	0.60
Age categorisation BIS	.129	9	.035	No	0.60
Configural featural: Face inversion index	.245	10	.038	No	0.47
Configural featural: Houses accuracy	.331	11	.042	No	0.25
Configural featural: Spacing advantage	.430	12	.046	No	0.38
Configural featural: House inversion index	.847	13	.050	No	0.27

Note: *p* values in bold remain significant after BH adjustment. FDR = 0.05, Number of measures = 13

Table S3 Intraclass correlation estimates for original preregistered measures

	Intraclass Correlation (model type)	95% confidence interval		F test with true value 0			
		Lower Bound	Upper Bound	Value	df1	df2	p
<u>Pre-registered</u>							
DP group <i>all</i> individual measures	.039 (C,1)	-.025	.182	1.40	17	153	.143
Control group <i>all</i> individual measures	.109 (C,1)	.049	.199	2.22	50	450	<.001
<u>Exploratory</u>							
DP group individual face <i>perception</i> measures only	.038 (C,1)	-.060	.225	1.23	19	95	.248
Control group individual face <i>perception</i> measures only	.040 (C,1)	-.025	.135	1.25	56	280	.123
DP group <i>average</i> face perception & face memory	.510 (C,2)	-.238	.806	2.04	19	19	.065
Control group <i>average</i> face perception & face memory	.679 (C,2)	.462	.808	3.11	59	59	<.001

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

Single case results

Table S4a Single case results showing z scores for accuracy and RT on nonidentity tasks

ID	Age group	Faces among non faces RT		Faces among non faces inversion effect		Faces among non faces normalised inversion effect		Mooney faces		Age categorisation		Gender categorisation	
		<i>Effect size</i> <i>Z CC</i>	Estimated % of controls faster than case's RT	<i>Effect size</i> <i>Z CC</i>	Estimated % of controls with a larger FIE than case's	<i>Effect size</i> <i>Z CC</i>	Estimated % of controls with a larger FIE than case's	<i>Effect size</i> <i>Z CC</i>	Estimated % of controls falling below case's score	<i>Effect size</i> <i>Z CC</i>	Estimated % of controls falling below case's score	<i>Effect size</i> <i>Z CC</i>	Estimated % of controls falling below case's score
CF059	Age 14-35	5.70	99.99%	0.98	82.4%	2.05	97%	0.21	57.9%	-13.2	0.00%	-9.72	0.00%
CF005	Age 14-35	0.04	51.6%	0.61	68.9%	0.58	71%	-0.10	46.1%	0.55	70.2%	-0.42	34.4%
AF016	Age 14-35	0.67	0.26%	-0.17	43.3%	-0.56	30%	-0.41	34.7%	0.55	70.2%	1.12	85.6%
AF002	Age 14-35	-0.08	46.8%	-0.21	42.1%	-0.23	41%	0.51	68.9%	0.55	70.2%	1.16	85.6%
AF017	Age 14-35	-0.81	22.0%	0.51	68.9%	0.03	51%	0.21	57.9%	0.55	70.2%	-3.54	0.13%
AF004	Age 14-35	1.13	0.86%	2.57	99.0%	3.23	100%	0.51	68.9%	0.55	70.2%	-0.42	34.4%
AF007	Age 14-35	0.06	52.4%	-2.90	0.5%	-1.85	4%	0.21	57.9%	0.55	70.2%	-0.42	34.4%
AF009	Age 36-59	2.07	95.4%	0.79	77.5%	1.39	91%	0.38	64.4%	0.55	70.2%	-1.49	8.20%
AF075	Age 36-59	1.12	83.6%	0.49	68.1%	0.91	81%	0.38	64.4%	0.55	70.2%	0.26	60.0%
AF018	Age 36-59	0.12	54.3%	1.29	88.8%	1.20	87%	-3.65	0.97%	0.55	70.2%	1.13	85.8%
AF021	Age 36-59	-0.07	47.3%	-0.46	32.8%	-0.33	38%	-0.02	49.2%	0.55	70.2%	1.13	85.8%
AF001	Age 36-59	-0.28	40.1%	0.12	54.7%	-0.02	49%	0.38	64.4%	0.55	70.2%	1.13	85.8%
AF010	Age 36-59	7.34	99.9%	3.23	99.8%	3.24	100%	-2.44	1.35%	0.55	70.2%	0.26	60.0%
AF006	Age 36-59	-0.06	30.4%	1.49	91.9%	1.12	86%	0.38	64.4%	0.55	70.2%	0.26	60.0%
AF003	Age 36-59	0.97	80.5%	-0.67	26.2%	0.08	53%	0.38	64.4%	0.55	70.2%	1.13	85.8%
AF060	Age 60-74	1.36	89.6%	0.16	56.1%	0.64	73%	0.66	73.3%	0.78	77.1%	0.20	57.5%
AF098	Age 60-74	2.30	98.0%	1.42	90.6%	1.89	96%	-10.5	0.00%	0.78	77.1%	-1.46	8.72%
AF019	Age 60-74	2.02	96.6%	2.26	96.7%	2.46	99%	-0.74	24.2%	0.78	77.1%	-0.64	27.0%
AF099	Age 60-74	-0.19	42.8%	-1.67	6.2%	-1.34	10%	0.66	73.3%	0.78	77.1%	-0.64	27.0%
AF008	Age 60-74	1.22	87.4%	0.52	68.9%	0.90	80%	-0.74	24.2%	0.78	77.1%	-2.27	2.05%

Table S4b Single case results showing configural featural face advantage and house advantage z scores

ID	Configural featural task									
	Face advantage					Spacing advantage				
	<i>Test for deficit in house accuracy Effect size</i>	<i>Test for deficit in face accuracy Effect size</i>	Z score for the difference between case and controls (Z-DCC)	Estimated % of population with a more extreme discrepancy	Presence of strong (S), classical (C), or no dissociation (N)	<i>Test for deficit in spacing condition Effect size</i>	<i>Test for a deficit in feature condition Effect size</i>	Z score for the difference between case and controls (Z-DCC)	Estimated % of population with a more extreme discrepancy	Presence of strong (S), classical (C), or no dissociation (N)
CF059	-5.28	-2.29	-2.27	2.2%	S	-4.47	-2.58	-1.67	6.6%	N
CF005	-0.64	-1.09	0.34	37.6%	N	-1.1	-0.85	-0.22	41.9%	N
AF016	0.79	-0.69	1.12	0.2%	N	-0.06	-0.28	0.19	43.0%	N
AF002	1.14	0.11	0.78	0.2%	N	0.46	0.56	-0.09	46.6%	N
AF017	0.79	-1.09	1.42	0.1%	N	-0.32	-0.56	0.21	42.3%	N
AF004	4.28	-0.49	0.70	0.3%	N	0.72	-1.13	1.59	7.5%	N
AF007	1.85	0.11	1.32	0.1%	N	1.23	0.28	0.82	22.4%	N
AF009	0.58	-0.48	0.74	24.4%	N	0.37	-0.38	0.60	28.9%	N
AF075	0.84	-0.21	0.74	24.6%	N	-0.62	1.89	-2.02	3.6%	N
AF018	-1.50	-1.83	0.23	41.3%	N	-0.87	-3.4	2.03	3.5%	N
AF021	1.62	0.61	0.71	25.3%	N	1.36	1.14	0.18	43.4%	N
AF001	0.58	0.07	0.36	36.7%	N	0.61	0	0.49	32.3%	N
AF010	-0.20	0.07	-0.18	43.1%	N	-0.38	0.38	-0.60	28.7%	N
AF006	0.84	-1.56	1.69	6.2%	N	-0.87	0.38	-1.00	17.7%	N
AF003	-0.20	-0.21	0.01	49.8%	N	-0.13	-0.38	0.20	42.6%	N
AF060	1.22	0.28	1.02	17.9%	N	1.03	0.26	0.76	24.4%	N
AF098	-0.88	-1.63	0.81	23.2%	N	-1.19	-1.31	0.12	45.6%	N
AF019	-0.65	-0.78	0.15	44.7%	N	1.03	-3.16	4.16	0.1%	C
AF099	0.76	-0.14	0.97	19.0%	N	0.29	0.26	0.03	49.1%	N
AF008	-0.65	0.49	-1.23	13.5%	N	0.47	-0.79	1.25	13.0%	N

Table S4c Single case results showing configural featural inversion effects and dissociation analysis

Configural featural task: Inversion effects					
ID	<i>Test for deficit in face inversion Effect Size</i>	<i>Test for a deficit in house inversion Effect Size</i>	Z score for the difference between case and controls (Z-DCC)	Estimated % of population with a more extreme discrepancy	Presence of strong (S), classical (C), or no dissociation (N)
CF059	-2.77	-1.71	-0.76	23.8%	N
CF005	0.22	1.20	-0.70	25.6%	N
AF016	-0.81	-0.11	-0.50	31.9%	N
AF002	0.81	0.27	0.38	35.9%	N
AF017	-0.76	2.42	-2.29	2.09%	N
AF004	-0.40	-1.63	0.89	20.4%	N
AF007	-0.67	0.27	-0.68	26.3%	N
AF009	-0.90	0.66	-1.00	17.5%	N
AF075	-0.08	-0.36	0.18	43.3%	N
AF018	-0.40	-1.24	0.53	30.7%	N
AF021	-0.16	0.16	-0.20	42.4%	N
AF001	0.81	0.95	-0.09	46.7%	N
AF010	0.33	-0.47	0.51	31.5%	N
AF006	-0.90	0.66	-1.00	17.5%	N
AF003	1.15	-0.99	1.36	10.4%	N
AF060	-0.31	-0.42	0.08	47.2%	N
AF098	-1.23	-0.08	-0.81	22.7%	N
AF019	-0.79	-0.86	0.05	48.1%	N
AF099	-0.61	0.99	-1.13	15.1%	N
AF008	1.57	-0.62	1.54	8.2%	N

Note: Inversion effects < 0 indicate relatively better accuracy on inverted vs upright trials and thus atypical performance

Table 4d Single case results CFPT upright accuracy

CFPT		
ID	Effect size Z CC	Estimated % of controls falling below case's score
CF059	-2.38	1.57%
CF005	-2.38	1.57%
AF016	-1.69	5.82%
AF002	-0.63	27.2%
AF017	-1.86	4.26%
AF004	-0.10	46.1%
AF007	-0.63	27.2%
AF009	-0.68	25.6%
AF075	-1.14	13.9%
AF018	-0.99	17.4%
AF021	-1.30	11.0%
AF001	-2.70	0.79%
AF010	-1.30	11.0%
AF006	-4.10	0.03%
AF003	-0.99	17.4%
AF060	0.58	70.9%
AF098	-0.80	22.4%
AF019	-1.05	16.3%
AF099	-0.30	38.8%
AF008	0.32	62.2%

Table S4e Single case dissociation of global face perception and face memory

Dissociation analysis global face perception (BIS) and global face memory (BIS)					
ID	Test for deficit in global face perception Effect size	Test for deficit in global face memory Effect size	Z score for the Difference between Case and Controls (Z-DCC)	Estimated % of Population with a more extreme discrepancy	Presence of strong (S), classical (C), or no dissociation (N)
CF059	-14.8	-2.81	-12.10	0.0%	S
CF005	-1.13	-2.93	1.82	5.2%	N
AF016	-0.22	-2.00	1.80	5.4%	N
AF002	-0.27	-2.21	1.96	4.1%	N
AF017	-0.74	-1.48	0.75	24.4%	N
AF004	-0.77	-1.91	1.16	14.6%	N
AF007	-2.50	-2.72	0.22	41.8%	N
AF009	-1.79	-2.48	0.78	23.6%	N
AF075	-0.90	-2.23	1.50	8.8%	N
AF018	-1.34	-2.19	0.95	19.2%	N
AF021	0.20	-1.95	2.42	1.8%	C
AF001	0.08	-1.19	1.43	9.8%	N
AF010	-2.23	-5.20	3.34	0.3%	S
AF006	-0.71	-1.54	-0.94	19.5%	N
AF003	-0.29	-1.51	1.37	10.8%	N
AF060	-0.01	-1.28	1.95	4.6%	N
AF098	-3.76	-1.59	-3.34	0.4%	C
AF019	-2.13	-3.34	1.86	5.3%	N
AF099	0.31	-2.08	3.68	0.2%	C
AF008	-1.56	-1.61	0.08	47.1%	N

Appendix 4: Chapter 6 supplementary materials

S1 Survey questions

Please answer as accurately and honestly as you can. If you prefer not to answer any of the questions please tick this option or write “prefer not to say”.

Section 1: Your experience of living with face recognition difficulties.

- 1) Please describe how you try to recognise people you know (colleagues, family, friends, acquaintances). [open text response]
- 2) Please tell us what you find hardest, if anything, about having trouble recognising faces? Are certain settings more challenging, has it affected your relationships, social or work life? [open text response]
- 3) Please tell us about any positives, if any, about having trouble recognising faces? [open text response]

Thank you. The following questions ask you to provide some numbers, please answer as accurately as you can, even if it is an estimate.

- 4) Approximately how many people that you know can you reliably recognise by their face alone when meeting them in context e.g., at a large family gathering, at a regular social or sports club you attend, at college, at work? You may want to list the people on a piece of paper, count them up and enter the total here.
- 5) Approximately how many people can you reliably identify by their face alone if you bumped into them unexpectedly and out of context e.g. in another town? If none write 0
- 6) Can you always recognise your immediate family (e.g. partner/children/parents/siblings) if you meet them unexpectedly out of context? *Yes/No/Don't know/Prefer not to say*
- 7) Can you always recognise your three closest friends if you meet them unexpectedly out of context? *Yes/No/Don't know/Prefer not to say*
- 8) Approximately how many times do you typically need to meet a person before you might recognise them (if you never recognise anyone, please state never)?
- 9) Thinking about different situations where you need to recognise someone, do you find recognising a familiar face more difficult: *In a crowd/One to one/No difference/Don't know/Prefer not to say*

10) Do you find recognising a familiar face more difficult: Indoors/Outside/No difference/Don't know /Prefer not to say

11) Is there a setting or situation in which you find it particularly challenging to recognise faces? [open text response]

Next, please rate how the following statements apply to you:

12) I tend to forget familiar faces if I don't see them regularly: Very often, quite often, occasionally, never/ Don't know /prefer not to say

13) I can easily recall facts about people I know e.g. their pet, where they went on holiday: *Very often, quite often, occasionally, never/ Don't know /prefer not to say*

14) I have mistaken a stranger for someone I know: Very often, quite often, occasionally, never/ Don't know /prefer not to say

Section 2: Your experiences of seeking help and diagnosis.

Thank you. Next we are interested to hear about your experiences of seeking help or support for face recognition difficulties

15) As far as you know, have you had trouble with faces your whole life? Yes/No

a) If **yes**, have you been formally diagnosed with the lifelong form of prosopagnosia (developmental prosopagnosia)? *Yes/no/Don't know/prefer not to say*

i) If yes who provided this diagnosis e.g. NHS neuropsychologist, private neuropsychologist, neurologist, university researcher, GP, Optometrist, other (please state occupation of the person who diagnosed you)

b) If **No**: Did the difficulty recognising faces start suddenly after previously normal face recognition ability? *Yes/no/Don't know/prefer not to say*

i) if yes, please describe what happened immediately before first noticed you suddenly had trouble recognising faces e.g. head injury, illness, stroke OPEN TEXT

16) Has anyone ever suggested that you get tested for prosopagnosia/face blindness *Yes/No Don't know/prefer not to say*

a) If yes, who? tick as many as apply: Family member, friend, GP, Neurologist, neuropsychologist, psychologist, eye health professional, teacher, educational psychologist, other (state)

17) And have you ever sought a diagnosis or requested to be tested for prosopagnosia?
Yes/No Don't know/prefer not to say

- a) If yes, who did you approach: GP/Teacher/Educational psychologist/Neuropsychologist/Eye health professional e.g. optometrist, optician, orthoptist/university researcher/Other (please state)
- b) Please describe what happened/the response (e.g. did the professional you approached know about prosopagnosia, were you offered support and advice, or tested for prosopagnosia?) [open text response]

18) Have you every sought information/advice on prosopagnosia? *Yes/No*

- a) If yes... where from: GP, School, education psychologist, NHS Website, NHS (please state which part/service), Google, University (please name), charity (please name), Online support group (please name), social media, OTHER (please state)

19) Have you told any of the following that you have or think you have prosopagnosia?
Tick all that apply:

- Spouse/partner
- Close family member
- Extended family
- Teacher/lecturer
- Colleague
- Manager
- GP or Doctor
- Close friend
- Acquaintance
- Counsellor/psychologist
- Other (please state)
- I have not told anyone

If yes, what was their response? Please state who you are referring to

20) What sources of information and advice about prosopagnosia you have you personally found particularly useful and would recommend to others, if any? [open text response]

21) What would you recommend as a first source of advice and information to a **parent** of a child or **family member** of someone with prosopagnosia, if any? [open text response]

22) What do you think would most make life easier for people with prosopagnosia e.g. information for schools, practical changes, new rules? [open text response]

Section 3 The next questions ask about strategies you use for managing face recognition difficulties

23) Please describe any strategies you use to help you recognise people in everyday life. Please state None if you don't use any [open text response]

24) Some people have reported that they use some of the following strategies to help them recognise faces. Please tell us about your experience of using the following strategies: *I tried and it helped, I tried and it didn't really help, I have not tried, prefer not to say*

- I try to memorise details about individual faces like freckles or scars
- I look up photos of people before I meet them e.g. staff directory, social media
- I ask a companion to identify the person for me
- I arrive first so that person I am meeting will find me rather than the other way around
- I ask people to introduce themselves
- I ask people to tell me what they will be wearing when we meet so I can find them by their clothing
- I take note of how people walk
- I try to wait until someone speaks so I can recognise them by voice
- I keep notes with descriptors of people that will help me recognise them again
- I keep photos of people I meet often and look over these on a regular basis
- I tell people I have trouble with faces
- I try to remember details like clothing, hairstyle, accessories, glasses
- When I meet someone, I remind myself to look closely at their face, even if it doesn't feel comfortable

25) Thinking about your place of work or study, please tell us about any adaptations you use(d) in workplace/school/college to help you recognise people (e.g., ask your colleagues to always introduce themselves, name badges). State NA if you don't use any [open text response]

26) Have you ever formally requested your place of work or study to provide adaptations for your face recognition difficulties E.g. using name badges? *Yes/no/Don't know/Prefer not to say*

If yes,

- a) Have you ever formally requested your place of work or study to provide adaptations for your face recognition difficulties E.g. using name badges?

What response did you receive? E.g. Did the organisation agree? Were adaptations put in place? How was your request received? [open text response]

- b) To what extent did you find the adaptations helped you if at all? [open text response]

Section 4: Research priorities

27) What is important to you in terms of what researchers dedicate their time to examining prosopagnosia? What would you like to see more research on? [open text response]

28) How would you rate the following priorities for future prosopagnosia research (order them 1-10 where 1 is top priority and 10 is lowest priority)

- Developing training to help improve face recognition
- Developing drug treatments to help improve face recognition
- Building a support community of people and families with DP
- Awareness raising for the general public
- Awareness raising for health professionals
- Awareness raising for workplaces
- Awareness raising for schools
- Improving access to diagnosis for prosopagnosia
- Strategies to help people manage/work around their face recognition difficulties
- Coping with anxiety resulting from prosopagnosia
- Researching the causes of developmental prosopagnosia
- Other

29) Finally, is there anything else you would like to share with us? [open text response]
Demographics: Please tell us about yourself

What year were you born?

Gender (open text),

Where do you live? *England, Scotland, Wales, Northern Ireland, Republic of Ireland, Channel Islands, Isle of Man , Other* (please state)

Employment status (*working, unemployed but available to work, not working due to illness/ disability or family reasons, student, retired, prefer not to say*)

As far as you know, does anyone else in your family have trouble recognising faces ?
(*yes/no/don't know/prefer not to say*)

Table S2 Correlational analysis examining relationship between real life measures and test score

		PI20	FFT	CFMT_zscore	CFMT_bis	cfm_diff_bis	Old New BIS	Recog_out_context	Recog_context
PI20	Spearman's rho	—							
	df	—							
	p-value	—							
FFT	Spearman's rho	-0.20	—						
	df	26	—						
	p-value	0.307	—						
CFMT_zscore	Spearman's rho	-0.59***	0.34	—					
	df	27	26	—					
	p-value	< .001	0.075	—					
CFMT_bis	Spearman's rho	-0.43*	0.25	0.50**	—				
	df	27	26	27	—				
	p-value	0.019	0.196	0.006	—				
cfm_diff_bis	Spearman's rho	-0.47**	0.12	0.51**	0.52**	—			
	df	27	26	27	27	—			
	p-value	0.009	0.548	0.005	0.004	—			
Old New BIS	Spearman's rho	0.07	0.44*	0.24	-0.02	-0.25	—		
	df	27	26	27	27	27	—		
	p-value	0.708	0.020	0.211	0.938	0.199	—		
Recog_out_context	Spearman's rho	-0.35	0.29	0.23	-0.02	-0.13	0.23	—	
	df	27	26	27	27	27	27	—	
	p-value	0.063	0.128	0.235	0.904	0.511	0.232	—	
Recog_context	Spearman's rho	-0.31	0.24	0.26	-0.14	-0.18	0.24	0.83***	—
	df	27	26	27	27	27	27	27	—
	p-value	0.098	0.219	0.168	0.470	0.360	0.217	< .001	—

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

Note: PI-20 = Prosopagnosia Index 20, CFMT = Cambridge Face Memory Test, z score _ standardised proportion correct, BIS= Balanced Integration Score, cfm_diff_bis = difference between face memory BIS and Bicycle memory BIS, FFT = Famous Faces Test, Old New = Old new Faces test, Recog_out_context = number of faces participants calculated they could recognise if meeting out of context, Recog_context = number of faces participants calculated they could recognise if meeting in context.

Table S3 Independent Samples *t* test comparing scores of participants who do (Group = Yes) and do not (Group = No) recognise immediate family

Independent Samples T-Test

		Statistic	±%	df	p		Effect Size
pi20	Bayes factor ₁₀	0.41	0.00				
	Welch's <i>t</i>	0.52		15.86	0.607	Cohen's <i>d</i>	0.21
on_zscore	Bayes factor ₁₀	0.77	0.00				
	Welch's <i>t</i>	-1.42		17.99	0.173	Cohen's <i>d</i>	-0.56
on_bis	Bayes factor ₁₀	0.80	0.00				
	Welch's <i>t</i>	-1.42		16.94	0.174	Cohen's <i>d</i>	-0.57
CFMT_zscore	Bayes factor ₁₀	0.41	0.00				
	Welch's <i>t</i>	-0.57		18.89	0.573	Cohen's <i>d</i>	-0.23
CFMT_bis	Bayes factor ₁₀	0.44	0.00				
	Welch's <i>t</i>	0.68		17.88	0.503	Cohen's <i>d</i>	0.27
cfmt_diff_zscore	Bayes factor ₁₀	0.41	0.00				
	Welch's <i>t</i>	-0.56		19.68	0.584	Cohen's <i>d</i>	-0.22
cfmt_diff_bis	Bayes factor ₁₀	0.39	0.00				
	Welch's <i>t</i>	0.43		17.37	0.670	Cohen's <i>d</i>	0.17
FFT_zscore	Bayes factor ₁₀	0.54	0.00				
	Welch's <i>t</i>	-1.04		20.07	0.308	Cohen's <i>d</i>	-0.41

Note: PI-20 = Prosopagnosia Index 20 CFMT = Cambridge Face Memory Test, *z* score = standardised proportion correct, BIS= Balanced Integration Score, cfmt_diff = difference between face memory and bicycle memory BIS, FFT = Famous Faces Test, on = Old new faces test

Group Descriptives

	Group	N	Mean	Median	SD	SE
pi20	No	10	84.20	84.50	9.60	3.04
	Yes	18	82.33	83.00	7.90	1.86
on_zscore	No	10	-3.10	-3.08	2.92	0.92
	Yes	18	-1.50	-0.54	2.79	0.66
on_bis	No	10	-4.96	-4.65	3.20	1.01
	Yes	18	-3.23	-2.70	2.85	0.67
CFMT_zscore	No	10	-1.43	-1.54	0.94	0.30
	Yes	18	-1.22	-1.22	0.95	0.22
CFMT_bis	No	10	-1.71	-1.34	1.60	0.51
	Yes	18	-2.13	-2.07	1.52	0.36
cfmt_diff_zscore	No	10	-1.75	-1.67	0.92	0.29
	Yes	18	-1.54	-1.76	0.97	0.23
cfmt_diff_bis	No	10	-1.70	-1.65	1.36	0.43
	Yes	18	-1.93	-1.80	1.24	0.29
FFT_zscore	No	10	-2.17	-1.55	1.90	0.60
	Yes	17	-1.36	-0.63	2.03	0.49

Table S4 Participants' spontaneous suggestions for future DP research

Theme	Suggestions
Rehabilitation of DP	▪ Effective techniques for recognising people ▪ Strategies for managing ▪ Help for the severe cases that most likely require rehabilitation ▪ Top tips ▪ Can it be improved? ▪ How to help children understand and adapt to DP ▪ Is there a cure? ▪ More therapeutic and/or technical strategies ▪ What may be successful non-clinical interventions and training or therapy. ▪ Whether you can train yourself to improve your facial recognition (similarly to how people with dyslexia can improve their reading). ▪ Adaptations which might help ▪ Strategies that help those with it
Understanding	▪ How and when does it develop? ▪ Are there different subtypes ▪ Gain a better understanding of the condition ▪ Is the part of the brain that's doing face recognition badly, doing something else ? ▪ Why this happens and in particular if there are circumstances where it is more likely to occur. ▪ An explanation for why I struggle would be interesting ▪ Can DP happen randomly with no trauma involved? ▪ A better understanding of what causes it and why would be very interesting. ▪ Why it happens ▪ Whether there are any benefits to having prosopagnosia (eg better voice recognition). ▪ What it is, how it develops, how it differs between people, genetic basis, ▪ Is there any hormonal link. e.g. I find that I'm slightly better at recognising people during certain times of the month. Also whilst I was pregnant.
Co-occurrence with other conditions and identities	▪ Association with other conditions e.g. Autism, navigation, object recognition, synaesthesia, anxiety, dyslexia, poor direction finding, mixed handedness, ADHD, aphantasia were specifically mentioned ▪ Links with stress and general mental load, e.g. I've noticed I find it harder to recognise faces during times of greater workload, or if I've been to an event this week where I've met lots of people.
Awareness of DP	▪ What are the current levels of awareness of DP in general population

Theme	Suggestions
	▪ Other peoples' understanding of face blindness / prosopagnosia (i.e. people who don't have it). What do they make of it? What effect would raising public awareness have on other peoples' understanding. Including in the work place. ▪ Perhaps raising awareness would be useful ▪ Wider discussion in the media.
Early intervention	▪ Mass testing of children ▪ How and when does it develop? Are there interventions for children that would make it less likely?
Other	▪ I think we need as much data as possible before we can decide what the most beneficial research direction will be. So more research on the lived experience is probably a good thing in the first instance. ▪ Sexuality, (are faceblind people more likely to be asexual?) ▪ Are faceblind people more interested in reading and is this associated with increased empathy, focus and concentrate, better sleep and brain ageing, ▪ Whether the condition is associated with different personality types and whether different personality types deal with the condition differently

Table S5 Overview of themes, sub themes and associated codes

Theme and description	Sub themes	Associated codes
1. Seeking explanation, validation and diagnosis This theme describes the process by which participants came to realise their face recognition difficulties were atypical, and the existence of DP as a condition. It captures the sources of information participants sought out in order to understand or explain their difficulties, and their experiences of seeking a diagnosis (or classification) whether formal, informal or self-diagnosis and how this made them feel.	1.1 “I thought it was just me“	Realising has DP
	1.2 The struggle to find reliable information	Internet Media Medical sources Other DPs/social media Support organisations University researchers Other sources
	1.3 Perceived value in diagnosis	Who was approached for diagnosis Misdiagnosis
2. “You don't know how difficult it is navigating life as a prosopagnosic”. This theme captures depictions of what it's like to live with poor face recognition, including whether this has changed over time – for example whether certain life stages were more difficult and if the condition gets more, or less challenging, with age. It describes the settings and circumstances that participants said they find the hardest as well as any perceived advantages of DP. It captures the techniques participants use in their everyday lives to try to recognise people and the extent to which they find these helpful. It also captures beliefs participants had about the condition and these, and the condition itself, affects their self-identity.	2.1 The most challenging aspects of DP	Main challenges Challenging settings and situations Factors that exacerbate difficulties Factors that exacerbate difficulties: Cognitive load Work experiences

	2.2 The use of effortful strategies for recognition	Body Clothing Gait Hair Mannerisms Memorise features Other strategies Paraphernalia Prompts from others Situational context Staged approach Stringing conversation along Use of notes Voice Wait for others to speak first Effortful Strategy failures
	2.3 Differing beliefs about DP and its impact	Age and stage Surprise oneself DP as a disability Severity Self-identity Lack of agency Acceptance Coping strategies
	2.4 Silver linings	Advantages of DP Provides perspective Positive affect
3. Widespread negative psychosocial impact This theme captures the varied aspects of life that are affected by DP and how the condition negatively impacts the lives of many participants; emotionally and psychologically, in relation to restricted social relationships and social development, family life, school and career. This theme also highlights that DPs often struggle to recognise those they are closest to, family and friends as well as colleagues they see very regularly and know well e.g. at work.	3.1 This is emotionally hard.	It's hard Negative emotions Feelings of embarrassment

	3.2 Isolation and social withdrawal	Effects on social relationships School experiences Social circles are kept deliberately small Perception that social skills are underdeveloped Social withdrawal and isolation Affects ability to believe others
	3.3 Negative impacts on family relationships	Impact on family relationships Family history of poor face recognition
4. Worrying what others think This theme captures how extremely important the positive regard of others is to DPs. It contains examples of the widespread and prominent concerns that almost all DPs expressed about potential and actual negative evaluation from others, a deep-seated worry that others would misinterpret their difficulty recognizing others for rudeness, being stuck up, cold or uncaring. The sub themes capture both <i>anticipated</i> negative reaction and judgments, i.e. guessing how others might interpret and judge their difficulties and also <i>direct experiences of</i> other's negative reactions both in relation to observed instances of failures to recognise familiar people as well as experience of disclosure.	4.1 Anticipating and receiving negative evaluation causes concern	Worrying what others think Fear of being considered uncaring Fear of being thought rude Fear of being thought standoffish or stuck up Fear of being considered stupid or incapable Negative reactions from others Hostile or impatient Teasing or joking Others misunderstand or misinterpret
	4.2 Mixed social reactions to disclosure	Feeling believed Not feeling believed Supportive response Lack of response Surprised Unsurprised Tried but failed to understand
5. Appropriate support helps but is often unavailable. This theme captures participants' recommendations for things that others can do (rather than that individual DPs themselves can do) to make navigating life as a DP easier, including informal support. This theme also captures the other side of the coin from theme 4, instances of others being supportive and understanding of DPs' difficulties and how valued this is.	5.1 Support helps to navigate life with DP	Supportive Pro-actively supportive Family and friends Sources of support: employer Sources of support: colleagues Employer Colleagues Other sources of support
	5.2 Reluctance to request or provide adaptations and accommodations	Adaptations and accommodations Disclosure as a strategy
	5.3 Increased awareness to reduce anxiety.	Need for awareness

	Disclosure as a strategy
5.4 Seek advice from experts	Advice for others

S6: Recommendations from DPs for sources of advice and information¹⁵

What sources of information and advice about prosopagnosia have you personally found particularly useful and would recommend to others with face blindness, if any?

- Faceblind UK have a marvellous book about what it's like living with prosopagnosia, with quotes from prosopagnosics. I lent it to several people at work to help them understand.
- I realised I had face blindness after I read the young adult novel "Holding Up the Universe" by Jennifer Niven. I recommend it as it opened my eyes to face blindness and on reflection, I find it valuable to see someone like me in a book. I recommend faceblind.org and their newsletter.
- I like the little badge that I got from Faceblind UK- it's something subtle that I wear on my lanyard and sometimes people ask.
- I would advise others to visit faceblind.org.uk, and to seek out and engage with university researchers.
- Can't think of anything specific. I think by the time you know you have prosopagnosia you tend to be an expert! The Facebook groups can be helpful but mostly for support, not information.
- From my research there isn't a huge amount out there, but there's a few self diagnosis questionnaires and I would recommend faceblind.org
- The Facebook groups. Reading other peoples' situations that I recognise so well. It has been quite liberating. Until a few years ago, I didn't know there was such a thing as Prosopagnosia! I always thought that it was "my own fault" somehow. I would definitely recommend these kind of groups.
- Just that I'm not alone
- I enjoy listening to the podcasts on face blindness because they provide real life experiences from regular people who are experiencing the same thing as me. I read academic articles and controlled news or university websites on the topic. I studied Psychology at university so I also have class notes on the disorder.

¹⁵ Duplicate answers are combined

- I think mainly that some of the anxiety I feel in social situations is a result of face blindness, and that doesn't make me stupid or shallow. Otherwise, I think, just get on with it
- Other than the information offered to me by Judith I don't think I have any other advice or information offered to me.
- Don't sweat about it! Just let people know you're not rude. You have a neurological condition
- Faceblind UK has lots of interesting information, but I also found lots of information on random websites that I cannot remember details of. Joining the research project was also very helpful and interesting.
- Faceblind.org. The NHS page was also useful as an overview, but the info is v basic.
- Radio 4 programme which initially gave a name to this (about 12-15 years ago). Before that I had thought it was just me. Following this I found the faceblind.org website and signed up for their newsletter which I read. I have read the young adult book "holding up the universe" by Jennifer Niven which included a main character who had severe face-blindness. I would recommend this book.
- Facebook page for people with the condition

What would you recommend as a first source of advice and information to a parent of a child or family member of someone with prosopagnosia, if any?

- Google
- All I could suggest is they do their own research and offer to discuss their [child's or family member's] experiences with them
- The majority of current resources I've found are inadequate. Many do not chime with my personal experience.

- Face Blind UK is the only source I am aware of, never really thought about googling it
- Nancy Mindick's book Understanding Facial Recognition Difficulties in Children.
- BBC news article on face blindness and podcasts by academics based on scientific evidence.
- Google I think. I think don't regard it a source of embarrassment, just treat it as something that makes you different, but not worse. And be honest with people if you don't recognise them, say you are bad at recognising, which is what I say to people.
- I think I would now point them toward Judith! But other than that I suppose the internet would be my obvious place to turn.
- Look for a self help group or society. If it's come on suddenly or a child involved, seek investigations from your GP to rule out anything serious.
- Face blind UK, and I possibly would recommend an official diagnosis for someone young, as this may help with coping strategies.
- With caveats, online
- I don't really know but I would say research a reliable specialist charity and contact them (if they exist).
- NHS website, Wikipedia
- Faceblind.org website (or if a child the book I mentioned "Holding up the universe" by Jennifer Niven).