CORTEX 183 (2025) 131–145

Available online at www.sciencedirect.com

ScienceDirect

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

Research report

A taxometric analysis of developmental prosopagnosia: Evidence for a categorically distinct impairment

Sarah Bate [a,*], *Emma Portch* [a], *Rachel J. Bennetts* [b] *and Benjamin A. Parris* [a]

[a] *Department of Psychology, Bournemouth University, Faculty of Science and Technology, Poole House, Fern Barrow, Poole, UK*

[b] *College of Health and Life Sciences, Division of Psychology, Brunel University, Uxbridge, UK*

ARTICLE INFO

Article history:
Received 29 February 2024
Reviewed 30 April 2024
Revised 25 July 2024
Accepted 17 October 2024
Action editor Stefan Schweinberger
Published online 23 November 2024

Keywords:
Taxometrics
Pathology
Prosopagnosia
Face recognition
Individual differences

ABSTRACT

Poor performance on cognitive assessment tasks may indicate a selective 'impairment'. However, it is unclear whether such difficulties separate the individual from the general population qualitatively (i.e., they form a discrete group) or quantitatively (i.e., they represent the lower end of a continuous distribution). Taxometric methods address this question but have rarely been applied to cognitive disorders. This study examined the latent structure of developmental prosopagnosia (DP) – a relatively selective deficit in face recognition that occurs in the absence of neurological injury. Multiple taxometric procedures were applied to dominant diagnostic indices of face recognition ability across two independent datasets. All analyses supported a categorical outcome, even for mild cases of DP, suggesting that it is a qualitatively distinct condition. This finding has significant implications for our understanding of DP given it has traditionally been viewed as a continuous impairment. In particular, existing (arbitrary) diagnostic cut-offs may be too conservative, underestimating prevalence rates and prohibiting big-data approaches to theoretical study. More broadly, these conclusions support application of the taxometric method to many other cognitive processes where weaknesses are predominantly assumed to reside on a continuous distribution.

* *Corresponding author*. Department of Psychology, Bournemouth University, Faculty of Science and Technology, Poole House, Fern Barrow, Poole, UK.
E-mail address: sbate@bournemouth.ac.uk (S. Bate).

https://doi.org/10.1016/j.cortex.2024.10.021

1. Introduction

A fundamental question challenging clinicians and scientists is whether psychopathologies or 'disordered' constructs are more accurately conceptualized in a categorical (i.e., with a clear boundary between individuals with and without the condition) or dimensional (i.e., continuous) manner. Meehl (1992) refers to clinical conditions that qualify as categorical as "taxa", conceptually defined as categories that have a latent structure that is based on an enduring, objective, non-arbitrary and naturally-occurring distinction between members and non-members. The taxometric method (Meehl, 1992, 1995) is used to objectively determine whether particular constructs meet these criteria via the rigorous application of data-analytic procedures that search for abrupt changes in distributions of scores, which may indicate the presence of latent subgroups. To date, the technique has frequently been used to determine whether many personality and psychiatric conditions possess a single categorical boundary (Ruscio & Ruscio, 2004), critically informing the assessment and statistical techniques that are used in diagnostic practice as well as advancing research methodology and theory.

Debates regarding the categorical or dimensional nature of psychopathologies have also circulated for decades in the cognitive psychology literature, although the taxometric method has seldom been applied [but see taxometric analyses of specific language impairment (Dollaghan, 2004, 2011) and autism spectrum traits (Frazier et al., 2010; James et al., 2016)]. Here, it is generally accepted that broad cognitive abilities are dimensional due to multiple causal influences, although it is reasonable to suspect that more specific processes represent distinct taxa (Ruscio et al., 2006). Whether this is genuinely the case for many cognitive conditions, however, remains unknown. In fact, the current state of the art is often contradictory, acknowledging that some abilities vary widely in the typical population, while also regarding poor-performers as "clinically impaired" if they fail to reach some arbitrary threshold on an assessment task (Palermo et al., 2014; Peretz et al., 2008). Thus, the latent structure and conceptualisation of many cognitive processes remain unresolved, despite presenting as ideal candidates for taxometric analysis.

A key example is the domain of face recognition—a process that many believe is underpinned by specialised processing strategies (Yin, 1969; Young et al., 1987; Yovel & Kanwisher, 2004) and dedicated neural regions (Kanwisher et al., 1997; McCarthy et al., 1997). It has long been known that damage to these brain areas severely disrupts face recognition skills—a condition known as "acquired prosopagnosia" (Bodamer, 1947; de Renzi et al., 1991). Although this condition is rare, it has an observable pathological basis and the loss of the ability to recognise highly familiar faces severely contrasts with premorbid abilities. While the psychopathology of acquired prosopagnosia seems clear, in the last 30 years it has become apparent that much larger numbers of people experience difficulties with face recognition that occur in the absence of any concurrent neurological, psychiatric, intellectual or visual conditions (Bate et al., 2019a; Susilo & Duchaine, 2013). These apparently lifelong instances of face recognition deficits are often conceptualised as the developmental equivalent of acquired prosopagnosia (the condition has even been coined "developmental prosopagnosia": DP; McConachie, 1976), although there appears to be more variation in the severity of this form of the condition (Barton & Corrow, 2016; Bate & Tree, 2017). Further, the parallel study of face recognition skills in the typical population has also revealed a broad spectrum of ability (e.g., Bindemann et al., 2012; Wang et al., 2012; White & Burton, 2022; Yovel et al., 2014).

These observations raise the question of whether face recognition skills naturally fall on a vast continuum, where those at the "bottom end of normal" are regarded as experiencing DP (Barton & Corrow, 2016; Bate & Tree, 2017; Corrow et al., 2016). This viewpoint is bolstered by the fact that, in addition to the absence of organic damage, no firm biological or genetic basis for DP has yet been observed (though some evidence links structural and functional neurological differences to face processing deficits in those with the condition; e.g., Avidan & Behrmann, 2009, 2014; Furl et al., 2011; Rivolta et al., 2014; Zhang et al., 2015; and there is some suggestion of heritability of the condition; see Bate et al., 2024; Duchaine et al., 2007; Kennerknecht et al., 2008; Lee et al., 2010; Schmalzl et al., 2008). Thus, in the absence of a pathogenic marker, developmental face recognition difficulties are commonly regarded as continuous impairments, where diagnostic cut-offs are simply based on arbitrary criteria that are calculated using statistical rules of thumb (for discussion see Barton & Corrow, 2016; Bate & Tree, 2017).

Perhaps due to the lack of availability of large and relevant datasets, only one study to date has attempted to directly address the psychopathology of DP. DeGutis et al. (2023) performed cluster analyses on a large dataset (DeGutis & Evans, 2023; N = 3,116) containing scores on dominant diagnostic indicators of the condition (a subjective self-report measure, plus scores on an unfamiliar face recognition task and a famous face recognition task). The authors examined whether there was a natural grouping of poorer face recognizers, but failed to find consistent groupings beyond those with skills that were generally above-versus generally below-average. Further, when the authors reviewed the diagnostic criteria used across 43 existing DP studies, they did not find an association between greater diagnostic strictness and performance on a face perception task (although note that this measure is not typically used in DP diagnosis because perception may dissociate from memory in at least some individuals: Bate et al., 2019b; Bate et al., 2022; Biotti et al., 2019; Dalrymple et al., 2014; Murray et al., 2022; Stantic et al., 2022). While these findings appear to support dimensionality, it should be noted that cluster analyses cannot determine whether a categorical or dimensional model best captures the latent structure of data (see Beauchaine, 2003; Beauchaine & Beauchaine, 2002; Beauchaine & Marsh, 2006, for expanded discussions). That is, cluster analysis cannot identify discrete taxa because most algorithms always partition datasets into subgroups, yet methods for determining the correct number of clusters are only effective when symptom overlap is absent or minimal—a circumstance that rarely occurs in psychological research (see Grove, 1991; Milligan & Cooper, 1985; Tonidandel & Overall, 2004). Thus, the latent structure of DP has yet to be firmly addressed.

The current investigation addressed this issue, performing the first taxometric analysis of DP. In an initial study we analyse a dataset collected by our laboratory, containing the performance of confirmed DP participants and a complement (control) class on core diagnostic measures of face recognition ability. We specifically aimed to clarify whether people with DP possess face recognition skills that differ qualitatively and non-arbitrarily from the typical population, or whether they merely reside at the bottom end of a continuous distribution (i.e., their face recognition skills fall below an arbitrary cut-off but are not otherwise distinct). The outcome of this study supports a taxonic structure. To determine whether this finding replicates in a much larger, incidentally recruited population, a second study reanalyses the open access dataset collected by DeGutis and Evans (2023). In addition to replication of our earlier findings, we also investigate whether the taxonic outcome would extend to individuals with "Minor" DP, who are diagnosed under a more relaxed set of criteria.

2. Study one

To determine whether face recognition ability in adults distributes in a categorical or dimensional manner, we performed a series of taxometric analyses on a dataset that was collected by our laboratory, containing a cohort of pre-confirmed participants with DP and age- and ethnicity-matched complement participants. In brief (the full approach is expanded below), the taxometric method is a statistical technique that tests whether a latent variable, measured by two or more ordinal or continuous observed variables, is categorical or continuous. To achieve this, it takes scores on two or more diagnostic indicators from members of the hypothesized category (the taxon, here DP) and the "typical" population (the complement class), and orders them along any one of these indicators (referred to as the "input" variable). The resulting distribution is then divided into a number of "windows" or "cuts", where the relationship between the input and remaining ("output") variables is examined. At each of these points, the bootstrapped dataset is compared with idealised categorical and dimensional models, and a value is calculated that indicates which model best fits with the data. This value is supplemented by output plots, where a taxon, if present, is typically visible via a distinct peak.

2.1. Materials and methods

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.

2.1.1. Participants

The minimum total sample size required for taxometric analysis is 300, of which at least 10% of cases should be the proposed taxon (Meehl, 1995; Walters & Ruscio, 2009). Our online prosopagnosia screening programme (www.prosopagnosiaresearch.org) had yielded 31 individuals (24 female; aged 36–60 years, M = 49.8, SD = 7.3) within the year 2020 who met dominant diagnostic criteria (Bate & Tree, 2017) for a diagnosis of DP. This requires some element of self-report (in this case it was self-referral to our laboratory for screening purposes), and the application of strict exclusion criteria prior to task completion (i.e., no history of neurological injury or concurrent psychiatric, developmental or intellectual condition). In addition, each individual is required to score at least two SDs below the control mean (age-matched norms are applied from Bate et al., 2019c; raw scores can be found in the raw data files: https://osf.io/jx9d7/) on at least two objective tests of face recognition: the Cambridge Face Memory Test (CFMT; Duchaine & Nakayama, 2006), the Cambridge Face Perception Test (CFPT; Duchaine et al., 2007), and the Famous Faces Test (Bate et al., 2019c). To achieve a total cohort size of 300, we collected data from a further 269 participants (135 female) aged 35–59 years (M = 45.5 years, SD = 6.8),[1] who were recruited via the Prolific participant database (www.prolific.co) in the same year that the DP participants were tested. These participants were presumed to be members of the "typical" population, and all confirmed that they had no history of psychiatric, developmental or neurological conditions. All DP and typical participants were UK nationals who had lived in the country for the majority of their lives. Ethical approval for the study was received from the institutional Ethics Committee.

2.1.2. Materials and procedure

Taxometric analyses typically employ measures that are used for diagnosis of the proposed taxon. However, indicators need to meet a specific set of assumptions: they should (a) each substantially differ between the proposed taxon and complement group by a minimum Cohen's *d* effect size of 1.25 (Meehl, 1995), and (b) be relatively uncorrelated (mean $r < .3$) among the taxon and complement cases (Ruscio et al., 2006). While these input data requirements are required for the taxometric method to be able to detect a categorical structure if it is present, they do not guarantee a categorical outcome. In fact, Haslam et al. (2020) carried out a meta-analysis of 183 research articles, and reported that findings supporting dimensional models outnumbered those supporting taxonic models by five to one.

Although we already held data for our 31 DP participants across several diagnostic measures (see raw data: https://osf.io/jx9d7/), for budgetary reasons we could not run the entire battery across such a large number of typical participants on an exploratory basis. At the same time, we wished to collect the complement data in the exact same manner as the DP data. We therefore selected the two indicators that are completed at the very start of the DP screening session, which, based on existing smaller datasets, we anticipated would meet the input assumptions for taxometric analysis.[2] These

[1] One participant failed to engage with all the tasks and was replaced during recruitment. The demographical data presented here reflects the final sample.

[2] Our decision to use these two measures was also informed by the likelihood of their ability to detect a taxon should one exist. That is, neither is influenced by chance and scores within the full range are possible, and both tap familiar face recognition (one objectively and one subjectively) – the dominant aspect of face recognition that reflects everyday functioning and at which most typical perceivers excel (e.g., Bruce et al., 2001; Burton et al., 1999; Hancock et al., 2000; Johnston & Edmonds, 2009).

were the data-driven Prosopagnosia Symptom Checklist (PSC) that was developed and validated by our group (Murray et al., 2018), and a short version of the famous face recognition task that has been used in our published work to diagnose DP (Bate et al., 2019c):

Prosopagnosia Symptom Checklist (Murray et al., 2018): Participants were presented with a list of 16 hallmark symptoms of DP (https://osf.io/jx9d7/) that were developed from interviews with adults with DP, their significant others and the parents of children with DP. In response to each symptom, they were asked to provide a rating on a Likert scale ranging from 1 to 5 (1 = never; 5 = frequently), indicating how frequently they experienced each scenario in their everyday face recognition experiences. Responses were totalled to provide an overall score out of 80, where higher responses corresponded to greater everyday difficulties in face recognition. Analysis of data returned from DP and complement participants indicated that the PSC is highly reliable (Cronbach's alpha = .88 and .92 respectively).

Famous Faces Test (Bate et al., 2019c): We used a famous face recognition task that had been specifically developed for individuals in our target age group, containing the faces of celebrities that were identified to be highly familiar to this age-range in initial pilot-testing. Here, we presented the faces of the most well-known 30 celebrities from a pool of 60 (as identified in Bate et al., 2019c). Each face was displayed sequentially, in a random order, for an unlimited time period. Participants were asked to type the person's name or some uniquely identifying biographical fact about that individual. As the measure of interest for prosopagnosia diagnosis is recognition rather than naming (Barton & Corrow, 2016; Biotti et al., 2019; Dalrymple & Palermo, 2016; Tsantani & Cook, 2020), either a name or uniquely-identifying biographical fact was accepted as a correct response. At the end of the study, participants were provided with a list of the names of the celebrities they had just viewed and were asked to rate their familiarity on a Likert scale that ranged from 1 (not at all familiar) to 5 (very familiar). Any celebrity that was unknown to each participant by name (i.e., those that were scored as a 1 or 2 on the Likert scale) were removed from the overall score and the percentage correct was adjusted accordingly. This design is adhered to by the majority of prosopagnosia laboratories worldwide (e.g., Arizpe et al., 2019; Bate et al., 2019b; Dalrymple & Palermo, 2016; Eimer et al., 2012; Mishra et al., 2021; Tsantani & Cook, 2020), and these tasks have very good reliability (e.g., .75-.80, Pozo et al., 2023). Here, we calculated split-half reliability using REL_{EX} (Steinke & Kopp, 2020), a software programme that repeatedly samples reliability coefficients from random splits while accounting for missing data (i.e., trials in which participants indicated they were not familiar with the person's name). A congeneric measurement model was assumed, so we report ρ_{SC} (the Angoff–Feldt coefficient; see Steinke & Kopp, 2020). Using this method, split-half reliability sampling with 10,000 iterations revealed a median reliability estimate of $\rho_{SC} = .914$; 95% of the sampled reliability coefficients lay between $\rho_{SC} = .885$ and $\rho_{SC} = .932$.

We do not have legal permission to publicly archive the stimuli used in this task; however, they can be accessed by contacting Prof. Sarah Bate (sbate@bournemouth.ac.uk).

Both DP and complement participants completed both tasks online via the Testable (www.testable.org) platform, in an identical manner. The PSC was always completed first so that participants did not reflect upon their performance in the more objective famous face recognition test prior to completing ratings of their everyday experiences with faces. Data for this study is publicly available at https://osf.io/jx9d7/.

2.1.3. Statistical analyses

There are two taxometric procedures that can be applied to datasets containing two diagnostic indicators:

MAMBAC (Mean Above Minus Below A Cut; Meehl & Yonce, 1994): MAMBAC is the simplest taxometric procedure. It assumes that data has a taxonic structure when an optimal cutting score is found that accurately classifies the cases into two groups, with only a small number of false positives and false negatives. Conversely, when no optimal cutting score is found, it assumes there are no groups to be distinguished and structure is therefore dimensional. MAMBAC searches for this optimal cutting score by sorting cases along the input indicator; a moving cutting score is then applied to this distribution, and cases falling above or below the cutting score are used to calculate mean differences from the output indicator(s). These differences are plotted along the *y* axis of a MAMBAC graph, with the number of cases plotted on the *x* axis. A peaked MAMBAC curve suggests a taxonic structure, where the optimal cutting score lies in the region of the peak. In contrast, concave curves without a clear peak reflect dimensional structure, often curving upwards at one or both endpoints.

MAMBAC also calculates an objective index that quantifies the relative fit of the results against categorical and dimensional comparison data: the Comparison Curve Fit Index (CCFI; for a review see Ruscio et al., 2011). CCFI values range from 0 to 1, where lower scores provide the strongest support for dimensional structure, and higher scores provide the strongest support for taxonic structure. Ruscio et al. (2010) recommended that CCFI values between .45 and .55 are treated as ambiguous, and this can be extended to those between .40 and .60 if one wishes to be particularly cautious.

MAXSLOPE (Maximum Slope; Grove & Meehl, 1993): MAXSLOPE is a simplified version of the MAXCOV (Maximum Covariance; Meehl & Yonce, 1996) procedure, offering an alternative procedure to MAMBAC when only two variables are available. MAXSLOPE creates a scatterplot between the two variables, and Cleveland's (1979) LOWESS (locally weighted scatterplot smoother) procedure is used to perform a nonlinear regression. When the resulting LOWESS curve follows an S-shaped trajectory, taxonic structure is inferred (slopes are fairly flat for low- and high-scoring cases, but steeper for those with intermediate scores). In contrast, a dimensional structure results in a fairly straight LOWESS curve with a positive slope. Ruscio and Walters (2011) recommended that the MAXSLOPE graph should be changed from a scatterplot to a plot of slopes by indicator scores, allowing for easier interpretation and the calculation of a CCFI value.

A key recommendation in the taxometric method is to carry out multiple analyses to assess the consistency of findings (Meehl, 1992). Here, we applied this recommendation by

Table 1 – Mean (SD) scores for the DP and complement groups on the PSC and Famous Faces Test.

	PSC (raw score/80)	Famous Faces Test (%)
DPs	67.29 (8.46)	47.04 (16.47)
Complement	39.71 (10.72)	96.29 (5.26)

implementing both the MAMBAC and MAXSLOPE procedures. The analyses were conducted using the *R* package and script developed by Ruscio and Wang (2017). Both the Famous Face Test and PSC indicators acted as inputs across 50 evenly spaced cuts, with 10 replications to ensure stability at each partition point (for evidence that this approach ensures stability of the outcome, see Ruscio & Walters, 2009; Walters & Ruscio, 2010). Two hundred data sets (100 categorical and 100 dimensional) were simulated as comparison data. Note that no part of the study procedures or analyses was preregistered prior to the research being conducted.

2.2. Results

Mean and SD scores for each indicator are presented as a function of group in Table 1, and the overall distributions are presented in Fig. 1. Inspection of individual scores in the complement group did not highlight any cases that could be suspicious for the taxon: where PSC scores were high (indicating high agreement with DP symptoms) they were not accompanied by poor famous face recognition scores and vice versa. Thus, all complement data were retained for the main analyses. While all DPs performed poorly on the FFT (i.e., at performing below the 2SD cutoff), 10 individuals returned PSC scores that were between 1 and 2 SDs from the complement mean.

Next, we inspected our dataset to ensure the taxometric method could be applied. Both the inter-group effect sizes for the Famous Faces Test and the PSC met the required criterion (Cohen's $d > 1.25$; Meehl, 1995): t (30.709) = 16.551, $p = .001$, $d = 6.82$, and t (298) = 13.835, $p = .001$, $d = 2.62$, respectively. Further, correlations between the two tasks were at an acceptable level ($r < .30$; Ruscio et al., 2006) in both the DP ($r = -.29$, $p = .120$) and complement ($r = -.09$, $p = .150$) groups. We were therefore able to proceed to the main taxometric analyses.

Examination of the comparison curves from two taxometric analyses indicated a better fit with a categorical structure. Fig. 2A displays the averaged MAMBAC curve against simulated categorical and dimensional data: while the categorical comparison and observed data are a close fit, the dimensional data are not. This is supported by the MAMBAC CCFI of .85. A similar finding was observed for the MAXSLOPE analysis (CCFI = .69; see Fig. 2B). Given the averaged CCFI value of .77 is well above the most conservative threshold for a categorical outcome (.6), these findings are interpreted as support for DP being a taxon.

2.3. Discussion

Two taxometric analyses provided evidence for DP being a categorically distinct condition, using objective and subjective indices of face recognition ability. This finding differs from existing views that DP is best conceptualised as a dimensional condition (e.g., Corrow et al., 2016; Tardif et al., 2019, but see Tian et al., 2020). While these assumptions have not previously been tested, they are largely based on observations that face recognition ability varies widely in the typical population (e.g., Bate et al., 2010; Bindemann et al., 2012; Wang et al., 2012; White & Burton, 2022; Yovel et al., 2014) and in the severity of DP itself (Barton & Corrow, 2016; Bate & Tree, 2017).

There is good reason to challenge this notion in light of the current set of findings. First, the variation in face recognition performance in typical perceivers is observed on measures of *unfamiliar* face recognition. That is, many people struggle to recognise faces they have only briefly seen before (e.g., Bruce et al., 2001; Hancock et al., 2000). In contrast, the vast majority of the population find it remarkably easy to recognise the faces of highly familiar people, and can even recognise these faces under the most challenging of viewing conditions (e.g., Burton et al., 1999; Jenkins et al., 2011; Johnston & Edmonds, 2009; Kramer et al., 2018).

Although prosopagnosia is often assessed using more convenient unfamiliar face recognition tasks, the traditional and most striking symptom (and the one that drives people to seek a diagnosis) is the failure to recognise highly familiar faces in everyday life (Adams et al., 2020; Murray et al., 2018; Portch et al., 2023). The two indicators used in the current study offer a closer reflection of this symptom than unfamiliar face recognition indices, and the categorical outcome supports our rationale of using the most appropriate measures to seek initial evidence that DP is a taxon. However, given unfamiliar face recognition skills are also disrupted in DP, together with the practical advantages of administering unfamiliar face recognition tasks, the next stage of investigation requires exploration of whether a taxonic outcome would also be observed when this indicator is combined with subjective and/or objective measures of familiar face recognition.

Second, the notion that DP itself may vary in severity also requires further exploration. Again, this concept has particularly been observed in the distribution of performance on unfamiliar face recognition tasks, where wide variation in performance may also reflect the influences of chance or measurement error. However, the distribution of DP performance in the current study also indicates some variation (see Fig. 1), despite the fact that these individuals were preidentified to have DP via the application of strict diagnostic protocols. What we cannot see from the current dataset is the performance of individuals who failed to meet these strict diagnostic criteria but nevertheless claim that they experience face recognition difficulties in everyday life. Such individuals may genuinely be those that experience the condition but have not been detected by the diagnostic approach used here, perhaps because they have found ways to effectively cope with or circumvent their difficulties in daily life (reducing their ratings of poor functioning on self-report questionnaires) and/or in objective assessment tasks (i.e., via the application of suboptimal cognitive strategies that result in a higher proportion of correct responses; Portch et al., 2023). A further alternative is that self-report of face recognition difficulties may tap different constructs to objective tasks, prohibiting

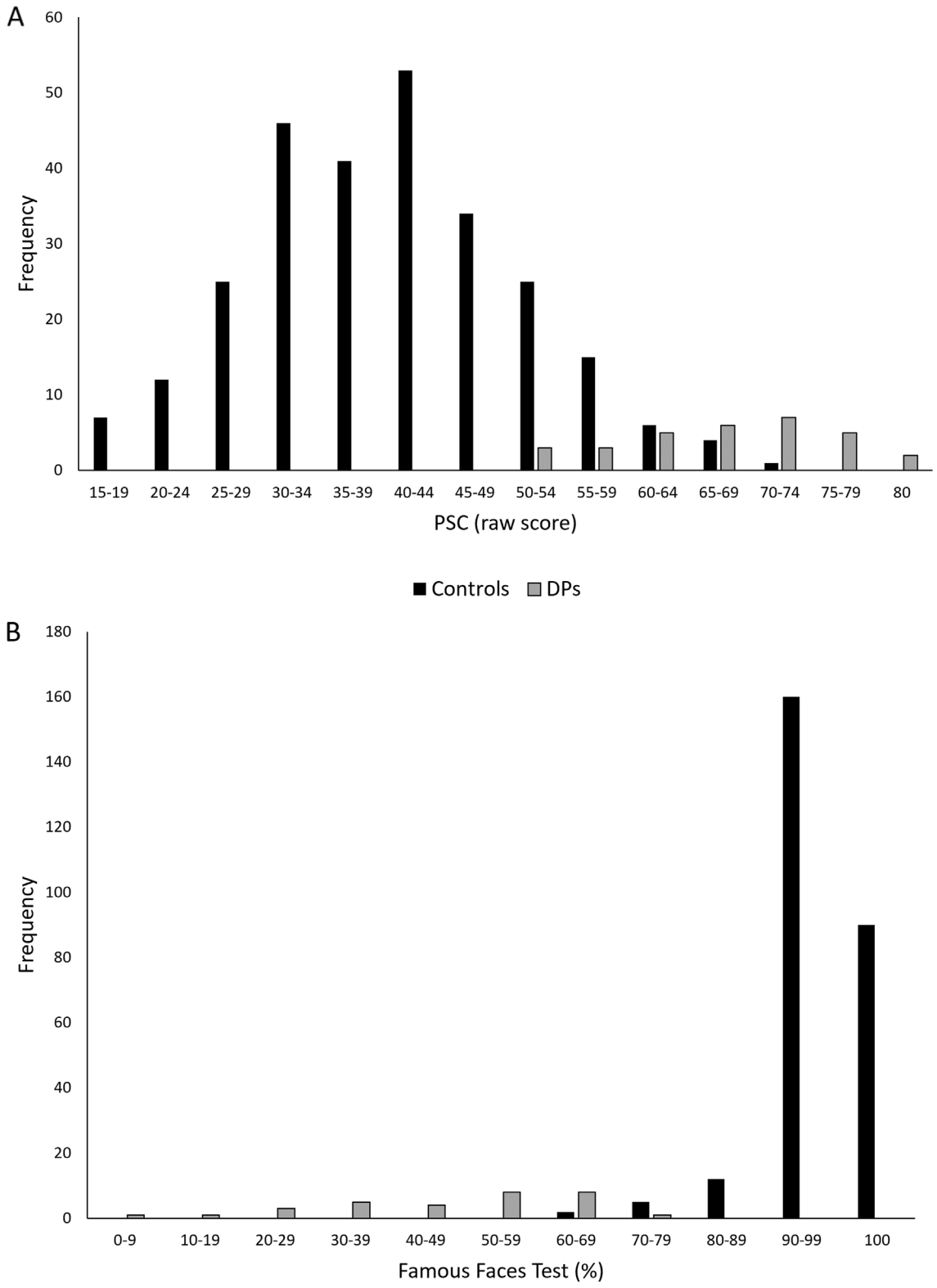

Fig. 1 – The distribution of scores for the DP and complement groups on (A) the PSC and (B) the Famous Faces Test.

some individuals from meeting the required diagnostic criteria across multiple measures (see Gerlach et al., 2024). While several authors have very recently debated the need for more inclusive diagnostic criteria that take account of both subjective and objective measures, while also relaxing the thresholds for cut-off performance (e.g., Burns et al., 2023; DeGutis & Campbell, 2024; Epihover & Astle, 2024; Gerlach et al., 2024), further work is required to map the underpinning constructs of these measures to examine their interrelationships.

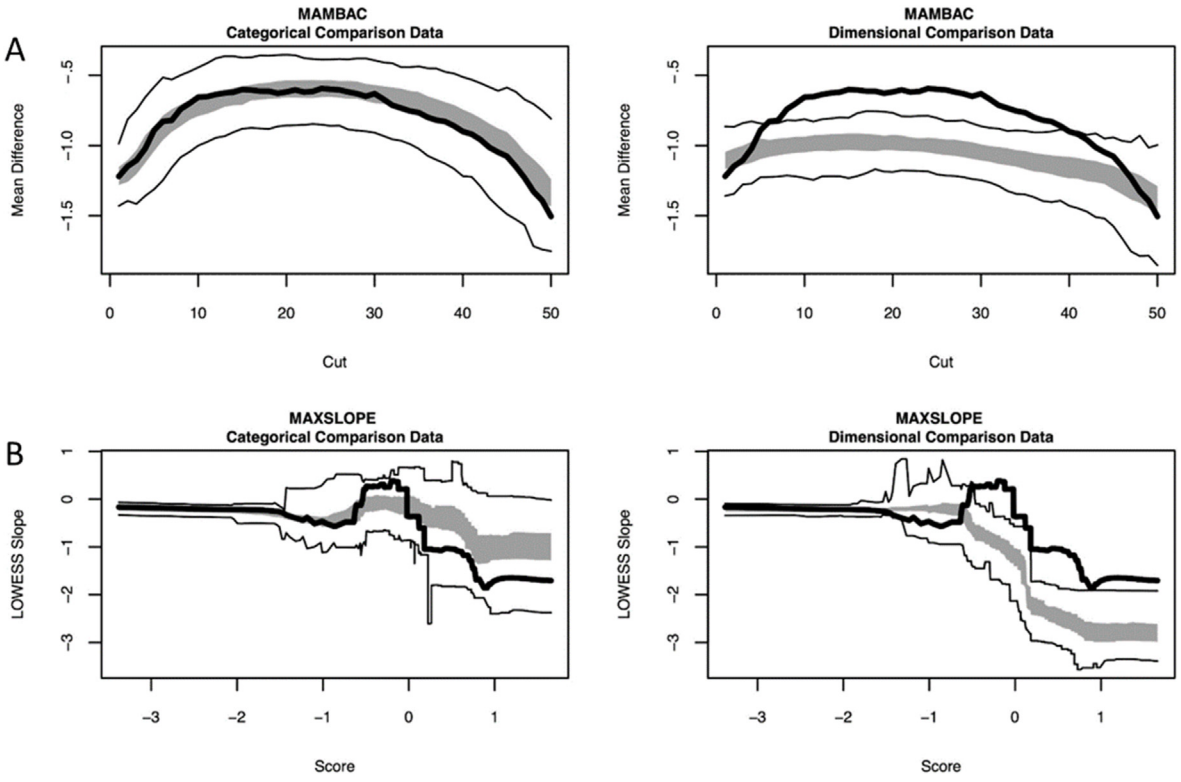

Fig. 2 – Categorical and dimensional comparison data from the (A) MAMBAC (CCFI = .85) and (B) MAXSLOPE (CCFI = .69) analyses. ***The average curve from each taxometric procedure is portrayed by the thick black line, against curves that are produced from simulated comparison datasets. The grey band represents the middle 50% of values from the comparison datasets, framed by the two thin dark lines that mark the largest and smallest values of the comparison datasets. For both taxometric procedures, the graphs show that the categorical comparison and observed data offer a closer fit than the dimensional data, which falls outside of the boundaries of the comparison values.***

To address the issues outlined above, and to replicate the basic finding that DP is a taxon, we performed a second taxometric analysis on a different dataset. This was also important because there are other potential limitations in our first study: we used the minimum sample size for taxometric analysis, we selected indicators that were most likely to result in a taxonic outcome, and we used a pre-existing group of DP participants and a separate complement cohort rather than random sampling from a single population. In Study 2 we were able to overcome these limitations by reanalysing the open access dataset offered by DeGutis and Evans (2023), incorporating a measure of unfamiliar face recognition into the taxometric analysis and including a group of "Minor" DPs as an additional proposed taxon (i.e., those who report everyday face recognition difficulties but only meet a lower-bound cut-off of 1–2 SDs on diagnostic tests).

3. Study 2

To further determine whether face recognition ability in adults distributes in a categorical or dimensional fashion, we performed a series of taxometric analyses on the large existing dataset (N = 3,116) offered by DeGutis and Evans (2023). This dataset presents scores on three diagnostic measures that are variations of those dominantly used in DP screening: An alternative version of the CFMT referred to as the "CFMT3", various versions of a famous face recognition task, and a self-report questionnaire that enquires about everyday experiences when recognising faces (the Cambridge Face Memory Questionnaire, CFMQ: Arizpe et al., 2019). The authors previously applied cluster analyses to this dataset, resulting in groupings that only reflected generally above-versus generally below-average face recognition skills. However, taxometric analysis—the technique that can more definitively inform whether DP is a taxon—has not yet been applied to this dataset.

Pertinently, DeGutis et al. (2023) also challenged the conservativeness of existing DP diagnostic criteria in a review of 43 studies, finding only a weak association between diagnostic strictness and face perception ability (as measured by the CFPT). The authors suggest this finding supports the use of more relaxed diagnostic criteria, as the presentation of the condition does not fundamentally change when this occurs. Burns et al. (2023) reported a similar finding: participants missing the typical 2SD CFMT cut-off still exhibited significant (but milder, i.e., ≤ 1SD) impairments on the CFPT (Duchaine et al., 2007) and a famous face test. Further, these individuals performed comparably to a severely impaired group on a measure of holistic processing and in their level of self-reported face recognition difficulties. As such, DeGutis et al. not only recommend that face recognition ability should be

viewed on a continuum, but they also suggested that the diagnosis of DP should follow the rule-of-thumb recommended by DSM-V (Sachdev et al., 2014), partitioning the condition into those with major (more than 2 SDs from the mean) versus mild (1–2 SDs from the mean) symptoms.

In our reanalysis of this dataset we were able to address this issue as well as the other limitations from Study 1, while examining whether the categorical outcome replicated in an independent cohort. Thus, we applied the taxometric method to this much larger sample where (a) we could conduct multiple replication analyses, (b) indicators of both unfamiliar and familiar face recognition ability were available, and (c) the DP group was extracted from the same sample as the complement and could be separated into those with "Major" versus "Minor" forms of the condition.

3.1. Dataset

3116 adult volunteers (1904 females) from the USA contributed to DeGutis and Evans's (2023) dataset. Their age ranged from 18 to 55 years (M = 31.0 years, SD = 10.5), and 61% were female. The study description did not call for participants of a certain face recognition ability, although the authors note that it may have inevitably attracted more individuals who suspect they struggle with face recognition than would be found in a fully representative sample (i.e., these participants may have participated because they wished to gain insight into their face recognition ability). All participants completed three assessments of face recognition online, via the testmybrain.org website. Tasks were completed in the following order:

CFMQ: This previously validated (see Arizpe et al., 2019) 18-item questionnaire requires participants to self-assess their face recognition skills in daily life. Items assess the frequency of positive and negative instances of face recognition performance, with one question requiring participants' assessment of their own skills relative to that of others. DeGutis et al. (2023) reported the CFMQ to be highly reliable (Cronbach's alpha = .91).

CFMT3: This task is an alternate version of the dominant CFMT (Duchaine & Nakayama, 2006) that is widely used in prosopagnosia screening (Bate et al., 2019c; Burns et al., 2023; Dalrymple & Palermo, 2016). The paradigm is described in depth elsewhere (see Duchaine & Nakayama, 2006); in brief, participants are required to learn six target faces and then recognize them across 72 triads of faces with varying levels of difficulty. The identical paradigm is applied to the CFMT3, but the facial images have been replaced with artificially-generated faces that depict different identities to those used in the original version. DeGutis et al. (2023) report that the CFMT3 has high internal consistency (Cronbach's alpha = .76) and found a robust correlation with performance on the original CFMT in 67 individuals ($r = .61$, $p < .001$).

Famous face recognition test: Each participant took part in one of three versions of a famous face recognition task that has been used in previous work (see Mishra et al., 2019), where they were required to identify images of 26, 27 or 40 celebrities from images that had been cropped around the jawline (24 images were repeated in at least one of the other versions). Participants were asked to type the name or some unique biographical information about the person in response to each face; they then viewed the correct answer and were asked to score their response (this facilitated a score that was based on face recognition itself, rather than naming and/or spelling accuracy). As in our procedure for Study 1, participants were then asked if they were actually familiar with the person, and trials were removed where the participant had no prior exposure to the target (note, as in Study 1, this prohibited a measure of reliability to be calculated for this particular task). To normalize the scores across the three different versions of the task, DeGutis et al. (2023) calculated the version-specific z-score for each participant.

3.2. Data overview

We initially inspected DeGutis and Evans's (2023) dataset to ensure the taxometric method could be applied. The first step required separation of the proposed taxon from the complement class. We initially applied the dominant and most conservative practice (in line with the criteria used in Study 1) of using some element of self-report together with scores that fall at least two SDs from the mean on two objective tests of face recognition ability. In the current dataset, DeGutis et al. (2023) determined self-awareness of face recognition difficulties by using responses on a single question from the CFMQ that has previously been used to determine subjective eligibility for DP (Arizpe et al., 2019): "Compared to my peers, I think my face recognition skills are …" (Far below average/Below average/Average/Above average/Far above average). Thus, we replicated this approach (accepting the two lowest response options as eligibility for DP), and used objective scores on the CFMT3 and famous face tasks (norms for the CFMT3 were calculated from the overall dataset, whereas the existing z-scores for the famous face tasks were simply maintained). The application of these protocols resulted in a pool of 23 Major DPs. We then followed DeGutis and colleagues' recommendation for identifying individuals with Minor DP by retaining the subjective requirement, but reducing the diagnostic cut-off for the two objective scores to 1–2 SDs from the mean. This approach identified 55 individuals.

However, while the sample size of those meeting the criteria for Minor DP falls within the minimum sample size for taxometric analysis (a minimum sample of 300 is required, of which at least 10% should represent the taxon: Meehl, 1995; Walters & Ruscio, 2009), unfortunately this was not the case for those who met the full 2SD criteria for Major DP (N = 23). We therefore combined this sample with other individuals who reported a subjective difficulty with face recognition on the single question, and achieved a "major" (at least 2SD) score on one objective indicator and a "minor" (at least 1SD) score on the other (note these individuals were not also included in the Minor DP group as we wished to firmly adhere to the criteria of two "minor" diagnostic scores for the taxometric analyses). This resulted in a final sample of 83 individuals (24 individuals presented with their major impairment on the CFMT3, and 36 with the major impairment on the Famous Faces Test), hereon referred to as the "Major DP" taxon (for demographical information see Table 2).

All remaining participants were allocated to the complement dataset. Thus, in addition to a large number of "typical" perceivers, this sample included some individuals who met

Table 2 – Demographical information and indicator validity for the Major and Minor DP groups, and the Test (T) and Replication (R) complement cohorts that were randomly selected for comparison to each DP group.

	Sample			Inter-group difference Cohen's *d*			Correlations (*r*)		
	N	Mean age (SD)	Sex (No. F)	CFMT	CFMQ	FFT	CFMT v CFMQ	CFMT v FFT	CFMQ v FFT
Major DPs	83	30.4 (10.9)	51				.19	−.31*	.10
Complement (T)	500	30.5 (10.4)	288	2.50**	1.92**	2.56**	.28**	.38**	.39**
Complement (R)	500	31.0 (10.7)	327	2.46**	2.23**	2.58**	.29**	.40**	.39**
Minor DPs	55	34.5 (11.5)	31				.05	.04	−.11
Complement (T)	350	31.1 (10.8)	217	1.74**	1.94**	1.75**	.27**	.28**	.49**
Complement (R)	350	31.9 (11.1)	214	1.83**	2.08**	1.81**	.28**	.32**	.36**

**$p < .001$, *$p < .05$ (*r* or *t*).

objective criteria for DP but did not self-report difficulties with face recognition in everyday life, as well as those who self-reported real-world difficulties but performed well on the objective tasks. We made the decision to include these participants in the complement sample to ensure that we were not skewing our data in favour of a categorical outcome; indeed, any of these individuals could arguably fall towards the bottom end of a continuous distribution. For the same reason, we did not perform any outlier analyses on our complement group and opted to retain all cases for analysis. However, due to the required sample size for taxometric analysis (a minimum sample of 300 of which at least 10% is the taxon), the size of the Major and Minor DP taxa prohibited us from including all of the complement group in each analysis. We therefore opted for a taxon base rate of 14%, allowing the size of the complement group to be 500 for the Major DP analyses, and 350 for the Minor DP analyses. Given a core principle of the taxometric method is replication using different samples (Meehl, 1992), we opted to repeat the analysis of each taxon twice, comparing them to different samples that were randomly extracted from the overall complement group. Thus, we randomly selected two groups of 500 members of the complement for "test" and "replication" analyses of the Major DP taxon, and two groups of 350 for "test" and "replication" analyses of the Minor DP taxon (for demographical information see Table 2).

Finally, it was necessary to ensure that the three diagnostic indicators for the newly grouped data met the assumptions for taxometric analysis. We therefore checked that all the inter-group effect sizes and correlations between the three diagnostic indicators (CFMQ, CFMT3, Famous Faces Test) were acceptable. While all effect sizes met the required criterion (all *d*s > 1.25), an unacceptable level of correlation was found in all complement groups between the famous face task and the other two measures (see Table 2). This restricted our taxometric analysis to two indicators: the CFMT3 and the CFMQ. Data that were used in each analysis are available at: https://osf.io/jx9d7/.

3.3. *Taxometric analyses*

As stated above, a key recommendation in the taxometric method is to carry out multiple analyses to assess the consistency of findings (Meehl, 1992). Here, we further applied this.

recommendation by not only analysing multiple complement groups, but also by implementing both the MAMBAC and MAXSLOPE procedures for each of the Major and Minor DP cohorts. Again, the analyses were conducted using the same *R* resources developed by Ruscio and Wang (2017). No part of the study procedures or analyses was pre-registered prior to the research being conducted.

Major DPs: All analyses supported a categorical model, with an examination of the comparison curves revealing that a categorical structure was a much better fit of the data. Fig. 3 shows the averaged MAMBAC curve against comparison categorical and dimensional data, for both the Test (panel A) and Replication (panel B) Complement groups. The graphs show that the observed data is a closer fit to the categorical than the dimensional comparison data. A similar finding was observed for the MAXSLOPE analysis (see Fig. 3), and again in comparison to both Complement cohorts. In all analyses, the CCFIs were greater than .80 (Test group: MAMBAC: .84, MAMSLOPE: .83, average: .83; Replication group: MAMBAC: .82, MAMSLOPE: .81, average: .81), with the overall average (.83) supporting a categorical model.

Minor DPs: All analyses and comparison curves (see Fig. 4) continued to support a categorical model. Again, the graphs show that the observed data are a closer fit to the categorical rather than the dimensional comparison data, and this was replicated for both complement cohorts. The CCFIs were all greater than .60 (Test group: MAMBAC: .86, MAMSLOPE: .79, average: .83; Replication group: MAMBAC: .69, MAMSLOPE: .60, average: .65), with the overall average (.74) again supporting a categorical model.

3.4. *Discussion*

A series of taxometric analyses consistently provided evidence for DP being a categorically distinct condition (a taxon) in both "Major" and "Mild" cases, and this replicated when the taxa were compared to different complement samples. It should be noted that one potential limitation is that the participants in this dataset were not screened for visual difficulties or concurrent socio-emotional disorder, allowing for the possibility that the taxon may extend beyond DP itself. However, the findings build upon those from Study 1 where these exclusion criteria were strictly applied, replicating the taxonic outcome in an independent dataset across multiple analyses. This finding differs to previous conclusions that were drawn from the same dataset, where cluster analyses failed to identify a distinct "DP" cluster that separated from the typical population

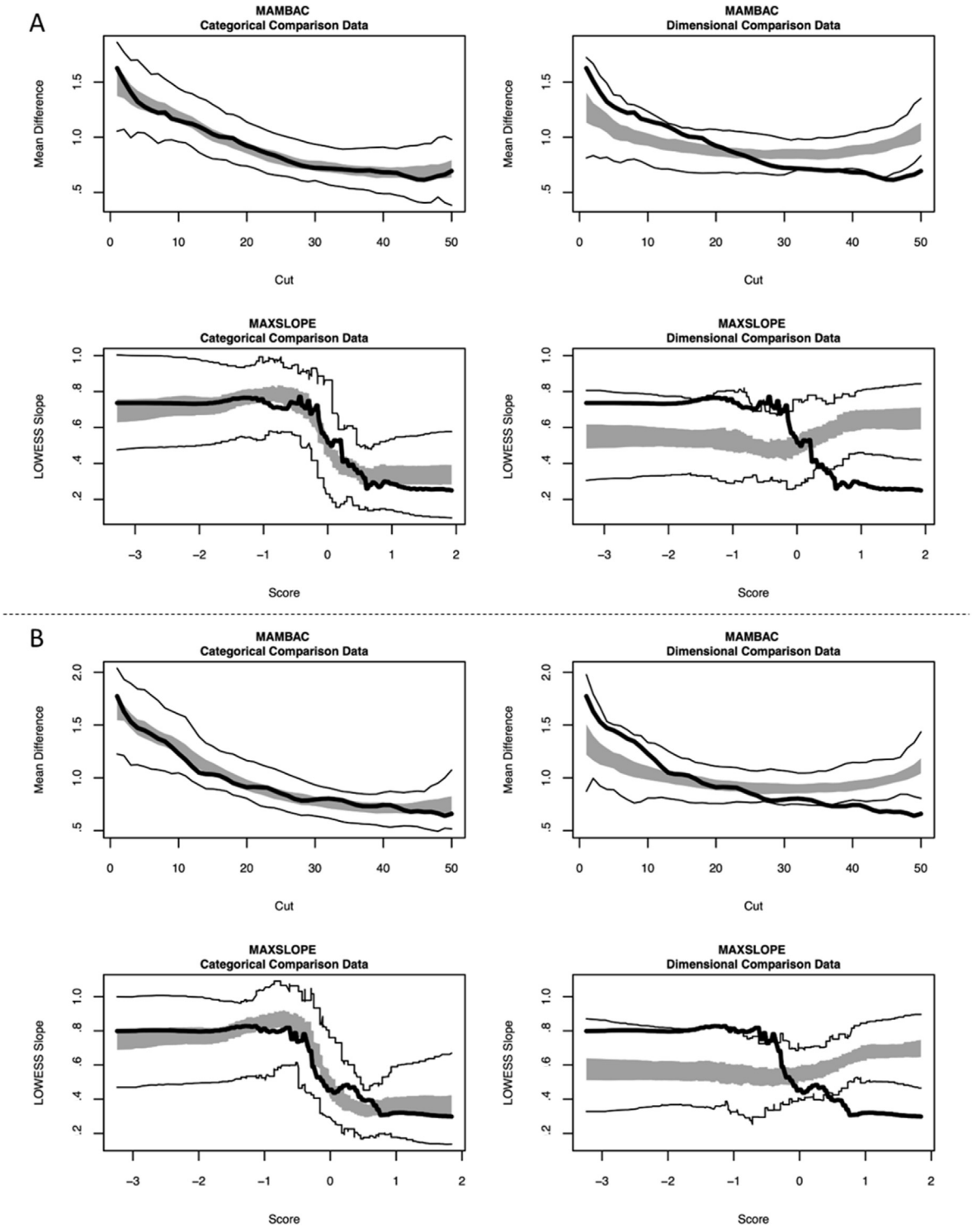

Fig. 3 – MAMBAC and MAXSLOPE curves for the proposed Major DP taxon against the (A) Test and (B) Replication complement cohorts. ***The average curve from each taxometric procedure is portrayed by the thick black line, against curves that are produced from simulated comparison datasets. The grey band represents the middle 50% of values from the comparison datasets, framed by the two thin dark lines that mark the largest and smallest values of the comparison datasets. For both taxometric procedures, the graphs show that the categorical comparison and observed data offer a closer fit than the dimensional data, which falls outside of the boundaries of the comparison values.***

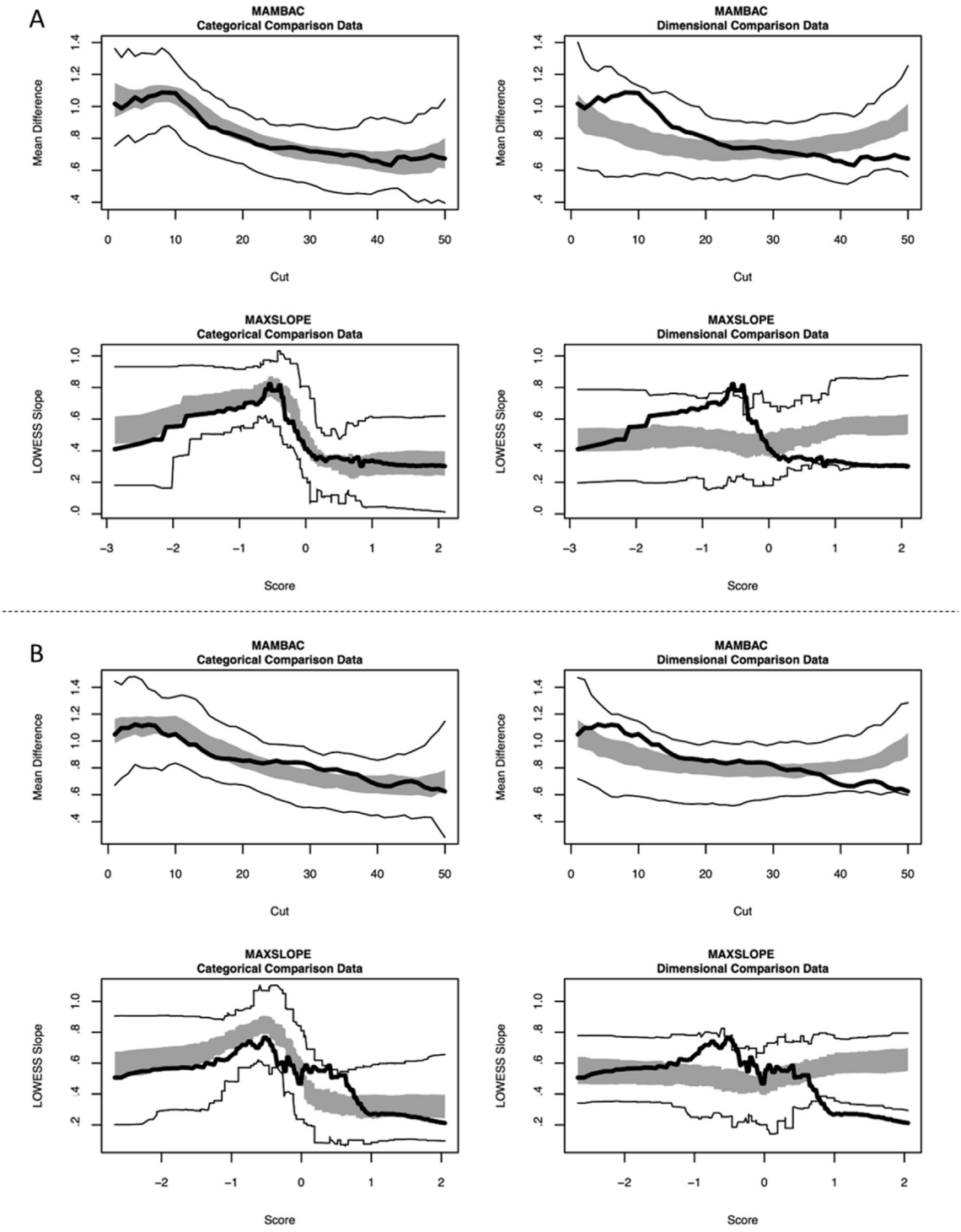

Fig. 4 – MAMBAC and MAXSLOPE curves for the proposed Minor DP taxon against the (A) Test and (B) Replication complement cohorts. ***The average curve from each taxometric procedure is portrayed by the thick black line, against curves that are produced from simulated comparison datasets. The grey band represents the middle 50% of values from the comparison datasets, framed by the two thin dark lines that mark the largest and smallest values of the comparison datasets. For both taxometric procedures, the graphs show that the categorical comparison and observed data offer a closer fit than the dimensional data, which falls outside of the boundaries of the comparison values.***

(DeGutis et al., 2023). This discrepancy in findings can be attributed to the two very different analytical techniques, with much existing discussion supporting the use of the taxometric method over cluster analyses for the purpose of determining psychopathology (Cleland et al., 2000; Ruscio & Ruscio, 2004). Alternatively, it is possible that exclusion of the FFT in the taxometric analysis prompted the different outcome.

There were two key findings in this study that developed those from Study 1. First, a measure of unfamiliar face recognition was used in this study but not in Study 1. The CFMT paradigm has been well-used to document individual differences in face recognition ability in the typical population (e.g., McCaffery et al., 2018; Verhallen et al., 2017), as well as showing varied performance in individuals with DP (Bowles et al., 2009; Burns et al., 2023; Murray & Bate, 2020). This data has been used to generate hypotheses of a dimensional distribution of face recognition ability. While the categorical outcome was supported by the additional inclusion of the subjective CFMQ measure, the findings reported here nevertheless lend support to the continued use of unfamiliar face recognition tests in diagnostic practice. Further, the additional support for DP emerging as a taxon is bolstered by the successful application of a measure that was not expected to provide such a strong categorical outcome.

Second, the taxonic outcome emerged even for "Minor" DPs, who only differed from the complement class by 1-2SDs. Traditionally, a strict 2SD cut-off has been recommended in DP diagnosis (e.g., Barton & Corrow, 2016; Dalrymple & Palermo, 2016), and this approach has resulted in prevalence estimates of 2–2.5% of the population (Bowles et al., 2009; Kennerknecht et al., 2006). Pertinently, when DeGutis et al. (2023) examined the difference in prevalence rates when diagnostic criteria are relaxed, an estimate of 3.08% emerged. This figure comprised both major (.93%) and mild cases (2.15%), as assessed via subjective self-report and objective performance on at least two valid screening tests. However, DeGutis et al. returned estimates as high 5.42%, dependent on the number and type of screening tests used, and the statistical method used to infer impairment.

While the CCFI indices were lower for the "Minor" compared to the "Major" DPs, most likely resulting from the better performance of the former group on the input measures of the analyses, the consistent categorical outcome nevertheless suggests that a 1SD cut-off on at least two tasks is a sufficient and non-arbitrary means of diagnosing DP, at least using the measures adopted within this study. Whether alternative indicators would result in the same outcome remains to be uncovered (see Gerlach et al., 2024), but there would be several advantages of implementing this approach. First, it would open the research field to much larger samples of research participants with DP, supporting a sorely needed big-data approach to key theoretical issues. Second, the application of a non-arbitrary cut-off gives a stronger framework for dealing with borderline cases and the influence of chance and measurement error in diagnosis. Finally, the relaxing of diagnostic protocols would have important implications for the many people who believe they experience everyday face recognition difficulties but are able to circumvent these deficits both in daily life and during formal assessment (Portch et al., 2023).

4. General discussion

In sum, the two studies reported here provide strong evidence to suggest that DP can be regarded as a taxon that can be categorically partitioned from the typical population. The indicators used in our analyses derive this conclusion from measures that reflect the core everyday symptoms that perhaps more convincingly set DPs apart from others (e.g., the striking failure to recognise familiar others in everyday life: e.g., McConachie, 1976; Murray et al., 2018), as opposed to performance on unfamiliar face recognition tasks where there is considerable variation both within the typical and DP population (Bindemann et al.,. 2012; Burns et al., 2023; McCaffery et al., 2018; Murray & Bate, 2020).

The finding that DP is a taxon has important implications for theory as well as diagnostic practice, particularly when combined with our observation that the taxon is maintained when diagnostic criteria are relaxed to 1SD on at least two objective tasks, plus some element of self-report. As discussed above, there are a range of benefits of moving to this more lenient approach to identifying DP, rather than maintaining arbitrary cut-offs for diagnosis: these include relief to the individuals concerned, as well as a stronger diagnostic framework that will aid decisions on borderline performance and inclusion criteria for theoretical study. Indeed, a major implication of this shift would be to make much larger samples of DPs available for research participation, allowing the field to carry out more rigorous work. Further, it would allow more valid calculations of prevalence, that are likely to shift from current estimates (2–2.5%) to a larger figure of ~3.08% (DeGutis et al., 2023).

In addition, the categorical outcome has further implications for research examining DP. It supports the continued use of a dichotomous rather than continuous approach to research design (i.e., when comparing DP groups to the typical population). Further, in terms of progressing diagnostic techniques, the findings support the use of a relatively small number of items that have discriminatory power close to the categorical boundary. While we do not wish to discount views that developmental conditions most likely result from different combinations and gradations of multiple and sometimes minor contributing factors (e.g., Susilo & Duchaine, 2013), and that these can result in different phenotypes of a condition (Bennetts et al., 2022), we should acknowledge that the categorical outcome supports continued exploration of the potential genetic underpinnings of DP. This does not necessarily mean that DP is explained by autosomal dominant inheritance (c.f. Kennerknecht et al., 2006), but may arise from an interaction effect between a genetic predisposition and a specific environmental issue, or a threshold effect involving a particular trigger (Ruscio et al., 2011). Further research is therefore required into the potential genetic underpinnings of DP, and the developmental trajectory of face recognition difficulties as they unfold during maturation (see Epihova & Astle, 2024). These two applications of DP research have received very little attention to date, and may be ripe grounds for theoretical progress.

In sum, this paper presents the first taxometric analysis of DP, providing support that it is a categorical condition that differs non-arbitrarily from fluctuations in face recognition ability that are observed in the typical population. This finding has important implications for the future study of DP: given a categorical outcome also emerged for individuals with a "minor" form of DP, we support recent suggestions for a relaxation of current diagnostic protocols (Burns et al., 2023;

DeGutis & Campbell, 2024; Epihova & Astle, 2024; Gerlach et al., 2024). Finally, given this is one of the first attempts to apply the taxometric approach to a cognitive process, we suggest that other specific processes (e.g., reading ability and dyslexia, see Cilibrasi & Tsimpli, 2020) are also investigated to test existing conceptualisations of dimensionality.

CRediT authorship contribution statement

Sarah Bate: Writing – original draft, Visualization, Validation, Software, Resources, Project administration, Methodology, Investigation, Funding acquisition, Formal analysis, Data curation, Conceptualization. **Emma Portch:** Writing – review & editing, Software, Methodology, Formal analysis, Data curation, Conceptualization. **Rachel J. Bennetts:** Writing – review & editing, Project administration, Methodology, Investigation, Data curation. **Benjamin A. Parris:** Writing – review & editing, Methodology, Investigation, Data curation, Conceptualization.

Ethical statement

Participants gave informed consent in accordance with guidelines set forth by the Ethics Committee at Bournemouth University.

Open practices

The study in this article has earned Open Data badge for transparent practices. The data are available at: https://osf.io/jx9d7/

Declaration of competing interest

All authors declare no competing interests.

REFERENCES

Adams, A., Hills, P., Bennetts, R., & Bate, S. (2020). Coping strategies for developmental prosopagnosia. *Neuropsychological Rehabilitation, 30*, 1996–2015.

Arizpe, J. M., Saad, E., Douglas, A. O., Germine, L., Wilmer, J. B., & DeGutis, J. M. (2019). Self-reported face recognition is highly valid, but alone is not highly discriminative of prosopagnosia-level performance on objective assessments. *Behavior Research Methods, 51*, 1102–1116.

Avidan, G., & Behrmann, M. (2009). Functional MRI reveals compromised neural integrity of the face processing network in congenital prosopagnosia. *Current Biology, 19*, 1146–1150.

Avidan, G., & Behrmann, M. (2014). Impairment of the face processing network in congenital prosopagnosia. *Frontiers in Bioscience (Elite Ed), 6*, 236–257.

Barton, J. J., & Corrow, S. L. (2016). The problem of being bad at faces. *Neuropsychologia, 89*, 119–124.

Bate, S., Adams, A., Bennetts, R., & Line, H. (2019a). Developmental prosopagnosia with concurrent topographical disorientation: A case report and virtual reality training programme. *Neuropsychological Rehabilitation, 29*, 1290–1312.

Bate, S., Bennetts, R. J., Gregory, N., Tree, J. J., Murray, E., Adams, A., Bobak, A. K., Penton, T., Yang, T., & Banissy, M. J. (2019c). Objective patterns of face recognition deficits in 165 adults with self-reported developmental prosopagnosia. *Brain Sciences, 9*, Article E133.

Bate, S., Bennetts, R. J., Tree, J. J., Adams, A., & Murray, E. (2019b). The domain-specificity of face matching impairments in 40 cases of developmental prosopagnosia. *Cognition, 192*, Article 104031.

Bate, S., Dalrymple, K., & Bennetts, R. J. (2022). Face recognition improvements in adults and children with face recognition difficulties. *Brain Communications, 4*, Article fcac068.

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*, 49.

Bate, S., Parris, B., Haslam, C., & Kay, J. (2010). Socio-emotional functioning and face recognition ability in the normal population. *Personality and Individual Differences, 48*, 239–242.

Bate, S., & Tree, J. J. (2017). The definition and diagnosis of developmental prosopagnosia. *The Quarterly Journal of Experimental Psychology: QJEP, 70*, 193–200.

Beauchaine, T. P. (2003). Taxometrics and developmental psychopathology. *Development and Psychopathology, 15*, 501–527.

Beauchaine, T. P., & Beauchaine, R. J. (2002). A comparison of maximum covariance and k-means cluster analysis in classifying cases into known taxon groups. *Psychological Methods, 7*, 245–261.

Beauchaine, T. P., & Marsh, P. (2006). Taxometric methods: Enhancing early detection and prevention of psychopathology by identifying latent vulnerability traits. In D. Cicchetti, & D. Cohen (Eds.), *Developmental psychopathology* (2nd ed., pp. 931–967). Hoboken, NJ: Wiley.

Bennetts, R. J., Gregory, N. J., Tree, J., Luft, C. D. B., Banissy, M. J., Murray, E., Penton, T., & Bate, S. (2022). Face specific inversion effects provide evidence for two subtypes of developmental prosopagnosia. *Neuropsychologia, 174*, Article 108332.

Bindemann, M., Avetisyan, M., & Rakow, T. (2012). Who can recognize unfamiliar faces? Individual differences and observer consistency in person identification. *Journal of Experimental Psychology. Applied, 18*, 277–291.

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

Bodamer, J. (1947). Die Prosop-Agnosie; die Agnosie des Physiognomieerkennens [Prosop's agnosia; the agnosia of cognition]. *Archiv fur Psychiatrie und Nervenkrankheiten, vereinigt mit Zeitschrift fur die gesamte Neurologie und Psychiatrie, 118*, 6–53.

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*, 423–455.

Bruce, V., Henderson, Z., Newman, C., & Burton, A. M. (2001). Matching identities of familiar and unfamiliar faces caught on CCTV images. *Journal of Experimental Psychology. Applied, 7*, 207–218.

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

Burton, A. M., Wilson, S., Cowan, M., & Bruce, V. (1999). Face recognition in poor-quality video: Evidence from security surveillance. *Psychological Science, 10*, 243–248.

Cilibrasi, L., & Tsimpli, I. (2020). Categorical and dimensional diagnoses of dyslexia: Are they compatible? *Frontiers in Psychology, 11*, 2171.
Cleland, C. M., Rothschild, L., & Haslam, N. (2000). Detecting latent taxa: Monte Carlo comparison of taxometric, mixture model, and clustering procedures. *Psychological Reports, 87*, 37–47.
Cleveland, W. S. (1979). Robust locally-weighted regression and smoothing scatterplots. *The Journal of the Acoustical Society of America, 74*, 829–836.
Corrow, J. C., Corrow, S. L., Lee, E., Pancaroglu, R., Burles, F., Duchaine, B., Iaria, G., & Barton, J. J. (2016). Getting lost: Topographic skills in acquired and developmental prosopagnosia. *Cortex; a Journal Devoted To the Study of the Nervous System and Behavior, 76*, 89–103.
Dalrymple, K. A., Garrido, L., & Duchaine, B. (2014). Dissociation between face perception and face memory in adults, but not children, with developmental prosopagnosia. *Developmental Cognitive Neuroscience, 10*, 10–20.
Dalrymple, K. A., & Palermo, R. (2016). Guidelines for studying developmental prosopagnosia in adults and children. *Wiley Interdisciplinary Reviews: Cognitive Science, 7*, 73–87.
De Renzi, E., Faglioni, P., Grossi, D., & Nichelli, P. (1991). Apperceptive and associative forms of prosopagnosia. *Cortex; a Journal Devoted To the Study of the Nervous System and Behavior, 27*, 213–221.
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; a Journal Devoted To the Study of the Nervous System and Behavior, 161*, 51–64.
DeGutis, J., & Campbell, A. (2024). Measure twice, cut once: Moving toward more inclusive, principled criteria for diagnosing developmental prosopagnosia. *Cortex; a Journal Devoted To the Study of the Nervous System and Behavior*. https://doi.org/10.1016/j.cortex.2024.06.006
DeGutis, J., & Evans, T. (2023). *What is the prevalence of developmental prosopagnosia? An empirical assessment of different diagnostic cutoffs*. January 18. Retrieved from osf.io/5469z.
Dollaghan, C. A. (2004). Taxometric analyses of specific language impairment in 3- and 4-year-old children. *Journal of Speech, Language, and Hearing Research: JSLHR, 47*, 464–475.
Dollaghan, C. A. (2011). Taxometric analyses of specific language impairment in 6-year-old children. *Journal of Speech, Language, and Hearing Research: JSLHR, 54*, 1361–1371.
Duchaine, B., Germine, L., & Nakayama, K. (2007). Family resemblance: Ten family members with prosopagnosia and within-class object agnosia. *Cognitive Neuropsychology, 24*, 419–430.
Duchaine, B., & Nakayama, K. (2006). The Cambridge Face Memory Test: Results for neurologically intact individuals and an investigation of its validity using inverted face stimuli and prosopagnosic participants. *Neuropsychologia, 44*, 576–585.
Eimer, M., Gosling, A., & Duchaine, B. (2012). Electrophysiological markers of covert face recognition in developmental prosopagnosia. *Brain: a Journal of Neurology, 135*, 542–554.
Epihover, G., & Astle, D. E. (2024). What is developmental about developmental prosopagnosia? *Cortex; a Journal Devoted To the Study of the Nervous System and Behavior, 173*, 333–338.
Frazier, T. W., Youngstrom, E. A., Sinclair, L., Kubu, C. S., Law, P., Rezai, A., Constantino, J. N., & Eng, C. (2010). Autism spectrum disorders as a qualitatively distinct category from typical behavior in a large, clinically ascertained sample. *Assessment, 17*, 308–320.
Furl, N., Garrido, L., Dolan, R. J., Driver, J., & Duchaine, B. (2011). Fusiform gyrus face selectivity relates to individual differences in facial recognition ability. *Journal of Cognitive Neuroscience, 23*, 1723–1740.
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; a Journal Devoted To the Study of the Nervous System and Behavior*. https://doi.org/10.1016/j.cortex.2024.03.001
Grove, W. M. (1991). Validity of taxonic inferences based on cluster analysis stopping rules. In D. Cicchetti, & W. M. Grove (Eds.), *Thinking clearly about psychology* (pp. 313–329). Minneapolis: University of Minnesota Press.
Grove, W. M., & Meehl, P. E. (1993). Simple regression-based procedures for taxometric investigations. *Psychological Reports, 73*, 707–737.
Hancock, P. J., Bruce, V., & Burton, A. M. (2000). Recognition of unfamiliar faces. *Trends in Cognitive Sciences, 4*, 330–337.
Haslam, N., McGrath, M. J., Viechtbauer, W., & Kuppens, P. (2020). Dimensions over categories: A meta-analysis of taxometric research. *Psychological Medicine, 50*, 1418–1432.
James, R. J., Dubey, I., Smith, D., Ropar, D., & Tunney, R. J. (2016). The latent structure of autistic traits: A taxometric, latent class and latent profile analysis of the adult autism spectrum Quotient. *Journal of Autism and Developmental Disorders, 46*, 3712–3728.
Jenkins, R., White, D., Van Montfort, X., & Burton, A. M. (2011). Variability in photos of the same face. *Cognition, 121*, 313–323.
Johnston, R. A., & Edmonds, A. J. (2009). Familiar and unfamiliar face recognition: A review. *Memory, 17*, 577–596.
Kanwisher, N., McDermott, J., & Chun, M. M. (1997). The fusiform face area: A module in human extrastriate cortex specialized for face perception. *The Journal of Neuroscience, 17*, 4302–4311.
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, 140*, 1617–1622.
Kennerknecht, I., Pluempe, N., & Welling, B. (2008). Congenital prosopagnosia - a common hereditary cognitive dysfunction in humans. *Frontiers in Bioscience, 13*, 3150–3158.
Kramer, R. S., Young, A. W., & Burton, A. M. (2018). Understanding face familiarity. *Cognition, 172*, 46–58.
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; a Journal Devoted To the Study of the Nervous System and Behavior, 46*, 949–964.
McCaffery, J. M., Robertson, D. J., Young, A. W., & Burton, A. M. (2018). Individual differences in face identity processing. *Cognitive Research: Principles and Implications, 3*, 1–15.
McCarthy, G., Puce, A., Gore, J. C., & Allison, T. (1997). Face-specific processing in the human fusiform gyrus. *Journal of Cognitive Neuroscience, 9*, 605–610.
McConachie, H. R. (1976). Developmental prosopagnosia. A single case report. *Cortex; a Journal Devoted To the Study of the Nervous System and Behavior, 12*, 76–82.
Meehl, P. E. (1992). Factors and taxa, traits and types, differences of degree and differences in kind. *Journal of Personality, 60*, 117–174.
Meehl, P. E. (1995). Bootstraps taxometrics: Solving the classification problem in psychopathology. *American Psychologist, 50*, 266–275.
Meehl, P. E., & Yonce, L. J. (1994). Taxometric analysis: I. Detecting taxonicity with two quantitative indicators using means above and below a sliding cut (MAMBAC procedure). *Psychological Reports, 74*, 1059–1274.
Meehl, P. E., & Yonce, L. J. (1996). Taxometric analysis: II. Detecting taxonicity using covariance of two quantitative

indicators in successive intervals of a third indicator (MAXCOV procedure). *Psychological Reports, 78*, 1091–1227.
Milligan, G. W., & Cooper, M. C. (1985). An examination of procedures for determining the number of clusters in a data set. *Psychometrika, 50*, 159–179.
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*, Article 108067.
Mishra, M. V., Likitlersuang, J., Wilmer, J. B., Cohan, S., Germine, L., & DeGutis, J. M. (2019). Gender differences in familiar face recognition and the influence of sociocultural gender inequality. *Scientific Reports, 9*, Article 17884.
Murray, E., & Bate, S. (2020). Diagnosing developmental prosopagnosia: Repeat assessment using the Cambridge face memory test. *Royal Society Open Science, 7*, Article 200884.
Murray, E., Bennetts, R., Tree, J., & Bate, S. (2022). An update of the Benton facial recognition test. *Behavior Research Methods, 54*, 2318–2333.
Murray, E., Hills, P. J., Bennetts, R. J., & Bate, S. (2018). Identifying hallmark symptoms of developmental prosopagnosia for non-experts. *Scientific Reports, 8*, 1690.
Palermo, L., Foti, F., Ferlazzo, F., Guariglia, C., & Petrosini, L. (2014). I find my way in a maze but not in my own territory! Navigational processing in developmental topographical disorientation. *Neuropsychology, 28*, 135–146.
Peretz, I., Gosselin, N., Tillmann, B., Cuddy, L. L., Gagnon, B., Trimmer, C. G., Paquette, S., & Bouchard, B. (2008). On-line identification of congenital amusia. *Music Perception: An Interdisciplinary Journal, 25*, 331–343.
Portch, E., Wignall, L., & Bate, S. (2023). Why can people with developmental prosopagnosia recognise some familiar faces? *Peer J, 11*, Article e15497.
Pozo, E. T., Germine, L., Scheuer, L., & Strong, R. W. (2023). Evaluating the reliability and validity of the Famous Faces Doppelgangers Test, a novel measure of familiar face recognition. *Assessment, 30*, 1200–1210.
Rivolta, D., Woolgar, A., Palermo, R., Butko, M., Schmalzl, L., & Williams, M. A. (2014). Multi-voxel pattern analysis (MVPA) reveals abnormal fMRI activity in both the "core" and "extended" face network in congenital prosopagnosia. *Frontiers in Human Neuroscience, 8*, 925.
Ruscio, J., Haslam, N., & Ruscio, A. M. (2006). *Introduction to the taxometric method: A practical guide*. Mahwah, NJ: Lawrence Erlbaum.
Ruscio, J., & Ruscio, A. M. (2004). Clarifying boundary issues in psychopathology: The role of taxometrics in a comprehensive program of structural research. *Journal of Abnormal Psychology, 113*, 24.
Ruscio, J., Ruscio, A. M., & Carney, L. M. (2011). Performing taxometric analysis to distinguish categorical and dimensional variables. *The Journal of Economic Perspectives: a Journal of the American Economic Association, 2*, 170–196.
Ruscio, J., & Walters, G. D. (2009). Using comparison data to differentiate categorical and dimensional data by examining factor score distributions: Resolving the mode problem. *Psychological Assessment, 21*, 578–594.
Ruscio, J., & Walters, G. D. (2011). Differentiating categorical and dimensional data with taxometric analysis: Are two variables better than none? *Psychological Assessment, 23*, 287–299.
Ruscio, J., Walters, G. D., Markus, D. K., & Kaczetow, W. (2010). Comparing the relative fit of categorical and dimensional latent variable models using consistency tests. *Psychological Assessment, 22*, 5–21.
Ruscio, J., & Wang, S. B. (2017). *Rtaxometrics: Taxometric analysis*. Available at https://CRAN.R-project.org/package=RTaxometrics (deposited 9th September 2017).
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*, 634–642.
Schmalzl, L., Palermo, R., & Coltheart, M. (2008). Cognitive heterogeneity in genetically based prosopagnosia: A family study. *Journal of Neuropsychology, 2*, 99–117.
Stantic, M., Brewer, R., Duchaine, B., Banissy, M., Bate, S., Susilo, T., Catmur, C., & Bird, G. (2022). The Oxford face matching task: A novel, non-biased, test of individual differences in face perception. *Behavior Research Methods, 54*, 158–173.
Steinke, A., & Kopp, B. (2020). Relex: An Excel-based software tool for sampling split-half reliability coefficients. *Methods in Psychology, 2*, Article 100023.
Susilo, T., & Duchaine, B. (2013). Advances in developmental prosopagnosia research. *Current Opinion in Neurobiology, 23*, 423–429.
Tardif, J., Morin Duchesne, X., Cohan, S., Royer, J., Blais, C., Fiset, D., Duchaine, B., & Gosselin, F. (2019). Use of face information varies systematically from developmental prosopagnosics to super-recognizers. *Psychological Science, 30*, 300–308.
Tian, X., Wang, R., Zhao, Y., Zhen, Z., Song, Y., & Liu, J. (2020). Multi-item discriminability pattern to faces in developmental prosopagnosia reveals distinct mechanisms of face processing. *Cerebral Cortex, 30*, 2986–2996.
Tonidandel, S., & Overall, J. E. (2004). Determining the number of clusters by sampling with replacement. *Psychological Methods, 9*, 238–249.
Tsantani, M., & Cook, R. (2020). Normal recognition of famous voices in developmental prosopagnosia. *Scientific Reports, 10*, Article 19757.
Verhallen, R. J., Bosten, J. M., Goodbourn, P. T., Lawrance-Owen, A. J., Bargary, G., & Mollon, J. (2017). General and specific factors in the processing of faces. *Vision Research, 141*, 217–227.
Walters, G. D., & Ruscio, J. (2009). To sum or not to sum: Taxometric analysis with ordered categorical assessment items. *Psychological Assessment, 21*, 99–111.
Walters, G. D., & Ruscio, J. (2010). Where do we draw the line? Assigning cases to subsamples for MAMBAC, MAXCOV, and MAXEIG taxometric analyses. *Assessment, 17*, 321–333.
Wang, R., Li, J., Fang, H., Tian, M., & Liu, J. (2012). Individual differences in holistic processing predict face recognition ability. *Psychological Science, 23*, 169–177.
White, D., & Burton, A. M. (2022). Individual differences and the multidimensional nature of face perception. *Nature Reviews Psychology, 1*, 287–300.
Yin, R. K. (1969). Looking at upside-down faces. *The Journal of Economic Perspectives: a Journal of the American Economic Association, 81*, 141–145.
Young, A. W., Hellawell, D., & Hay, D. C. (1987). Configurational information in face perception. *Perception, 16*, 747–759.
Yovel, G., & Kanwisher, N. (2004). Face perception: Domain specific, not process specific. *Neuron, 44*, 889–898.
Yovel, G., Wilmer, J. B., & Duchaine, B. (2014). What can individual differences reveal about face processing? *Frontiers in Human Neuroscience, 8*, 562.
Zhang, J., Liu, J., & Xu, Y. (2015). Neural decoding reveals impaired face configural processing in the right fusiform face area of individuals with developmental prosopagnosia. *Journal of Neuroscience, 35*, 1539–1548.