This paper is the accepted version for publication at Behavior Research Methods on the 7th of June, 2024. Please cite as follows:

Burns, E. J. (2024). Improving the DSM-5 approach to cognitive impairment: developmental prosopagnosia reveals the need for tailored diagnoses. *Behavior Research Methods.*

Improving the DSM-5 approach to cognitive impairment: developmental prosopagnosia reveals the need for tailored diagnoses

Edwin J. Burns
Department of Psychology, Swansea University

Number of figures and tables: 7 (6 figures, 1 table).
Supplementary materials: Yes (https://osf.io/3x86n/).
Number of pages: 46
Abstract word count: 175
Word count (excluding abstract, figures and reference list): 9355

*Correspondence:
Dr. Edwin Burns
Department of Psychology
Swansea University
UK
Phone: +44 07831 106743

Email: edwinjamesburns@gmail.com

Abstract

The Diagnostic Statistical Manual of Mental Disorders (DSM-5) recommends diagnosing neurocognitive disorders (i.e., cognitive impairment) when a patient scores beyond -1 SD below neurotypical norms on two tests. I review how this approach will fail due to cognitive tests' power limitations, validity issues, imperfect reliabilities, and biases, before summarising their resulting negative consequences. As a proof of concept, I use developmental prosopagnosia, a condition characterised by difficulties recognising faces, to show the DSM-5 only diagnoses 62-70% (*n1* = 61, *n2* = 165) versus 100% (*n1* = 61) through symptoms alone. Pooling the DSM-5 missed cases confirmed the presence of group-level impairments on objective tests, which were further evidenced through meta-analyses, thus validating their highly atypical symptoms. These findings support a paradigm shift towards bespoke diagnostic approaches for distinct cognitive impairments, including a symptom-based method when validated effective. I reject dogmatic adherence to the DSM-5 approach to neurocognitive disorders, and underscore the importance of a data driven, transdiagnostic approach to understanding patients' subjective cognitive impairments. This will ultimately benefit patients, their families, clinicians, and scientific progress.

Keywords: diagnosis; neurocognitive disorders; prosopagnosia; single case analysis; mild cognitive impairment; major; subjective cognitive impairment, MCI; transdiagnostic.

1. Introduction

The Diagnostic and Statistical Manual for Mental Disorders fifth edition (DSM-5, APA, 2014) is considered the gold standard guidance for practitioners diagnosing mental disorders in the United States. One section of the DSM-5 focuses on neurocognitive disorders which can be graded as reflecting mild and major cognitive impairment due to a variety of causes, including Alzheimer's disease, Parkinson's disease, HIV and traumatic brain injury (Sachdev et al., 2014). Impairments are typically characterised by a reduction in cognitive or behavioural functioning within, or across, six key domains, covering perceptual-motor function, language, learning and memory, social cognition, complex attention, and executive function (Sachdev et al., 2014). To diagnose impairment, the DSM-5 developers recommend that a patient must score more poorly than one standard deviation below a neurotypical mean on two cognitive or behavioural tasks (Sachdev et al., 2014).

It could be claimed that the DSM-5 has been helpful in providing formal, and straightforward, standardised guidance to diagnose neurocognitive disorders. However, this approach has been criticised partly because the liberal criterion of -1 SDs below a neurotypical mean will result in mistaken diagnoses (Schultz, 2010, 2013; Wakefield, 2013), i.e., sixteen percent of the normally distributed general population would be diagnosed as abnormal on a single test, despite such individuals being cognitively intact. Thus, the DSM-5 in its current form poses risks to neurotypical patients by diagnosing impairments where none exist.

However, we must not ignore the converse risks absolute cut-offs pose when patients are erroneously rejected as cognitively intact, i.e., missed diagnoses. Such false negatives are rarely highlighted in critiques of the DSM-5 approach to neurocognitive disorders (Schultz, 2010, 2013; Wakefield, 2013), despite missed diagnoses and pathologizing normality often having a common cause, i.e., when patients' diagnostic test performance distributions

substantially overlap with neurotypicals. Numerous conditions associated with cognitive impairments suffer problems with missed diagnoses, including Long COVID (Costas-Carrera et al., 2022; Pihlaja et al., 2023) and dementia (Beishon et al., 2019; Potts et al., 2022), where patients will often appear, at least from their cognitive test results, indistinguishable from those absent of disease. It is therefore important for clinicians and researchers to understand the risks associated with missed diagnoses, and how they can occur. This will help avoid dogmatic thinking that the DSM-5 approach is infallible when assessing whether a patient is impaired.

Can developmental prosopagnosia reveal the DSM-5's limitations?

To illustrate how the DSM-5 approach can result in missed diagnoses, I present the case of developmental prosopagnosia (DP). This is a lifelong condition characterised by severe difficulties recognising facial identity (Avidan & Behrmann, 2008; Bate et al., 2014; Behrmann & Avidan, 2005; Behrmann et al., 2005; Bennetts et al., 2024; De Haan, 1999; De Haan & Campbell, 1991; Duchaine & Nakayama, 2006; Halder et al., 2024; Maw et al., 2024; McConachie, 1976; Thomas et al., 2009), affecting 1.88-6% of the general population (Burns, 2023; Burns et al., 2022; Gray et al., 2017; Kennerknecht et al., 2006; Kennerknecht et al., 2008). It can have a substantial negative impact upon peoples' interpersonal, romantic and professional relationships, causing fear, anxiety and low self-confidence (Dalrymple et al., 2014; Yardley et al., 2008). While the causes of DP are unclear, it does run in families suggesting a possible genetic component (De Haan, 1999; Duchaine, Germine, et al., 2007; Grueter et al., 2007; Kennerknecht et al., 2008; Lee et al., 2010), which may account for the wide range of neural atypicalities they exhibit (Behrmann et al., 2007; Behrmann & Plaut, 2013; Burns et al., 2013; 2014; Fisher et al., 2016, 2017; Fisher et al., 2020; Fox et al., 2011; Furl et al., 2011; Jiahui et al., 2018; Lohse et al., 2016; Manippa et al., 2023; Righart & de

Gelder, 2007; Rivolta et al., 2014; Rosenthal et al., 2017; Song et al., 2015; Thomas et al., 2008; Towler et al., 2012; 2018; 2016; Van den Stock et al., 2008).

I chose this group for the current paper first, because I have experience working with them and the tests used to assess their problems. Second, these individuals suffer an extremely high proportion of potentially missed diagnoses (i.e., up to 85%) when using a cut-off of -2 SDs on two cognitive tasks of face processing (Bate, Bennetts, Gregory, et al., 2019; Burns et al., 2022; Lowes et al., 2023). Owing to this, I had an *a priori* hypothesis (see Burns et al., 2023) that problems in diagnosing would remain, even if we had used the more liberal DSM-5 criteria for neurocognitive disorders.

While neurodevelopmental conditions like DP would not be included in the umbrella term of neurocognitive disorders by the DSM-5 developers (Sachdev et al., 2014), researchers have applied its principles to DP (DeGutis et al., 2023; Stumps et al., 2020). Similarly, even though the updated DSM-5 stresses that clinicians must not rigidly follow its recommended diagnostic cut-offs (DSM-5-TR Neurocognitive Disorders Supplement, APA, 2022), this is what some researchers have suggested the field adopt (DeGutis et al., 2023). Moreover, an argument could be made that acquired prosopagnosia, typically onsetting after an observable brain injury (Barton et al., 2001; Behrmann & Plaut, 2014; Bornstein & Kidron, 1959; Bukach et al., 2006; Bukach et al., 2012; Bukach et al., 2008; de Gelder & Rouw, 2000; Humphreys et al., 2007; Marotta et al., 2001), *would* come under the umbrella term of neurocognitive disorders. Importantly, this form suffers similar problems as DP whereby cases can perform too well on face processing tasks (Burns et al., 2023; Fysh & Ramon, 2022; Josephs & Josephs, 2024). Given acquired cases are exceptionally difficult to recruit in large numbers, and similar cognitive task based diagnostic issues are present in both groups, it seems reasonable to use DP in the present paper to demonstrate the limitations of the DSM-5 approach to diagnosing.

Reasons for Missed Diagnoses

In this section, I present reasons why developmental prosopagnosia may go undiagnosed through the DSM-5, building on prior work (Burns et al., 2023; Epihova & Astle, 2024; Lowes et al., 2023; McIntosh & Rittmo, 2021; Volfart & Rossion, 2024). My intention is to provide a basis upon which professionals dealing with other neurocognitive disorders can reflect on, and scrutinize, the potential issues with cognitive and behavioural tests they employ. This should encourage the adoption of a more accurate, data driven approach for diagnosing and treatment, where researchers and clinicians recognise the limitations of the DSM-5 method, i.e., we should attempt to validate self-reported complaints with objective data, rather than using arbitrary cognitive task cutoffs to reject subjective complaints.

Before I begin, I should mention some researchers believe self-identified DP cases do not have the condition when their individual cognitive test scores miss diagnostic cutoffs. While theoretically possible, I largely reject this hypothesis. In my experience, historically missed cases describe qualitatively similar face recognition failures during interview as those who do meet criteria, and exhibit quantitatively comparable symptoms (Burns et al., 2023). Moreover, it is arguably easy to detect when you fail to recognise a familiar person during a conversation, as it is patently obvious that they know who you are, but you do not recollect them (Burns, 2023; Burns et al., 2023; Tsantani et al., 2021). Consider the types of conversations you have with familiar people; they are intuitively different from those struck up by strangers. As I suspect most, if not all, missed cases have DP, this review focuses on why the DSM-5 fails to diagnose those with the condition.

This first reason why the DSM-5 fails is because researchers and clinicians do not follow the guidance set out by those who develop diagnostic tests. This has been a consistent issue in the DP literature since the Cambridge Face Memory Test (CFMT), an unfamiliar face

memory task widely used to diagnose developmental prosopagnosia, first came into print (Duchaine & Nakayama, 2006). Its developers reported that it failed to detect impairments at the -2 SD level in 25% of DP cases, and 12.5% at -1 SD. As a result, the authors stated that professionals should not solely rely on it for a diagnosis (Duchaine & Nakayama, 2006). Despite this, researchers and clinicians have not heeded these warnings. If you read the literature over the last 10-15 years, you will find impairment on the CFMT was essential for a diagnosis in the majority of papers (Burns et al., 2023; DeGutis et al., 2023), and I must admit to being guilty of this myself (Burns et al., 2017; Burns et al., 2017; Burns et al., 2014; Wilcockson et al., 2020). Thus, even when developers of diagnostic tests highlight their limitations, professionals will fail to acknowledge them. This will result in patients erroneously told they do not have developmental prosopagnosia, simply because the tests and cutoffs we enforce do not capture every patient's impairment.

Why do tasks like the CFMT fail to detect atypicality in every self-identified case? One reason may be that such tests suffer imperfect ecological validity (Burns et al., 2023; Ramon et al., 2019). This occurs if they fail to fully capture the problems a patient experiences in the real world. Alternatively, such tasks may accurately reflect the problems they suffer from, but fail to clearly detect the superior abilities of neurotypicals. In either case, the performance distributions of DP and neurotypicals will overlap to such an extent that they render the DSM-5's -1 SD cutoff ineffective for diagnostic purposes.

To illustrate how tests can potentially lack validity, let us consider the defining characteristics of DP: consistent failures when attempting to recognise personally familiar people, such as co-workers, friends and even family members. In an ideal world, it would seem sensible to use these people in our diagnostic tests. However, this is exceptionally impractical due to the consent requirements of all involved, and the time constraints on researchers and clinicians who must create such tasks. As a solution, a famous faces test

(FFT) is almost always used to aid a diagnosis (Bate & Tree, 2017; Burns et al., 2022; Dalrymple & Palermo, 2016), where patients are required to recognise images of highly familiar celebrities.

However, this test cannot be easily standardised given that people of different cultures, different personal interests, and different age groups will be more familiar with certain famous faces than others. This may partly explain the heterogeneity of 15-35% of those who self-identify as suffering from developmental prosopagnosia failing to score below -1 SD on this task (Bate et al., 2019; Burns et al., 2023; Lowes et al., 2023). When employed, this cutoff simply removes the top end of the homogenous DP performance distribution when plotted with those who do meet criteria (Figure 1). If we assume that they are all part of the same DP group as the distributions suggest, then the -1 SD cutoff will inevitably exclude many from a diagnosis.

Why might people with DP score above the DSM-5 cutoff on the famous faces test? These tasks almost always use a single, still image of each celebrity. Maybe neurotypicals rely more heavily on movement when recognising familiar faces in the real world than DP cases. This means neurotypical performance when using photographs will be shifted down so that many DP cases land above the -1 SD cutoff, despite their problems in real life. Another reason may be that people with developmental prosopagnosia have more problems in real world settings than is captured by a computer screen based famous faces test[1]. This could occur if the brain processes celebrities, or rather celebrity photographs, to some extent differently from personal acquaintances (Herzmann et al., 2004; Ramon & Gobbini, 2018;

[1] Curiously, video clips of famous faces abolish the recognition impairments that are apparent when still images are used (Bennetts et al., 2015). Does normal performance on what is a more ecologically valid test due to the inclusion of movement, not suggest that cognitive tests fail to capture something with regards to DP cases' complaints? Maybe diagnosing DP cases through still image recognition impairments, we are removing those who are largely intact on this task, but who may suffer difficulties integrating identity related information when faces are moving. Anecdotally, DP cases who perform too well on cognitive tests to acquire a diagnosis have told me that they find moving faces particularly challenging in contrast to photographs.

Taylor et al., 2009; Wiese et al., 2022), where the recognition of the former may be partly intact in DP, while the latter is not. In any case, given that 15-35% of people with self-identified prosopagnosia perform too well on famous faces tests (Bate et al., 2019; Burns et al., 2022; Lowes et al., 2023), we must acknowledge such tasks' validity limitations can theoretically cause missed diagnoses.

Cognitive tests also suffer imperfect test-retest reliabilities too, whereby a patient can acquire a diagnosis one day, but then fail to gain one the next. To illustrate this, I reanalysed data from Murray and Bate (2020) who retested DP cases days to months apart after an initial assessment on the CFMT. Out of their 70 cases, 29% failed to replicate their initial diagnostic status using the -1 SD cutoff, shifting from DP to neurotypical, or from neurotypical to DP (Figure 2). Importantly, the CFMT has been discoverable from internet searches over the years, so those suspecting that they may have the condition could have taken it prior to formal testing. Given 80% of Bate and Murray's (2020) cases that crossed the -1 SD threshold on their second attempt moved from potentially diagnosed to missed should give cause for concern. This is because many DP cases will miss acquiring a diagnosis simply because of their curiosity to seek out an initial online CFMT self-assessment prior to contacting a clinician for testing. Thus, the DSM-5 will fail to diagnose many patients because of imperfect test-retest reliabilities[2].

Problems with two diagnostic tests

It is important to note that scoring below -1 SD on a single cognitive test is not enough to acquire a diagnosis in the DSM-5 (Sachdev et al., 2014). A patient must score below -1 SD on *two* tests (Sachdev et al., 2014). However, this additional requirement is especially problematic as it exacerbates the issue of missed diagnoses: maybe all cognitive

[2] Please note, Murray and Bate (2020) showed changing diagnostic status was also an issue when using the -2 SD cutoff.

tests suffer the validity and reliability issues outlined above. If true, our ability to diagnose any cognitive impairment will be constrained by the statistical power of the weakest test. Power in this case is simply the percentage of our patient sample that scores below any arbitrary cutoff (i.e., alpha) we have chosen for our two tests[3] (McIntosh & Rittmo, 2021). When the famous faces test requires a -1 SD cutoff, analysing data from recent papers suggests its power ranges from 65-85% (e.g., Bate et al., 2019; Burns et al., 2022; Lowes et al., 2023). The CFMT is the most widely used standardised DP diagnostic assessment in combination with the FFT (Burns et al., 2023; DeGutis et al., 2023). However, it too misses many cases, exhibiting power of only 50-79% at the -1 SD cutoff (e.g., Bate et al., 2019; Burns et al., 2023; Lowes et al., 2023). If we require impairment on both the FFT and CFMT, then power will likely decline further, with it impossible to achieve overall power to diagnose higher than that provided by the lowest powered test.

To illustrate this, I plotted the FFT and CFMT data (Figure 1) from a large sample of 165 self-identified DP cases reported in Bate et al. (2019). Eighty-five percent met the -1 SD diagnostic criteria on the famous faces test, while 79% met criteria on the CFMT. When we required impairment on both tests, power fell below the lowest of the two, namely the CFMT, with 70% diagnosed. Many people who self-identify as having DP will therefore never acquire a diagnosis simply because of power constraints. Furthermore, modelling work has shown testing additional control participants will have limited scope for improving power (McIntosh & Rittmo, 2021).

Another issue with requiring impairment on two cognitive tests is that it misses cases who are, objectively speaking, highly abnormal in terms of their single test score. We can see

[3] My discussion of power limitations is comparatively brief compared to prior works on the subject in neuropsychology (McIntosh & Rittmo, 2021) and DP (Burns et al., 2023). I recommend both papers, but especially the first which provides a convincing demonstration of how cognitive tests can lack power.

this in the bottom left panel of Figure 1. The FFT cut-off excludes cases who score more poorly than -2 SDs on the CFMT, which is the two-tailed threshold for an individual's test score being statistically significant in terms of atypicality. Despite these cases suffering severe difficulties when learning new faces, the DSM-5 requirement of impaired on two tests means they will never acquire a diagnosis. These problems appear even more pronounced when we examine cases excluded by the CFMT, with many scoring between -2 to -7 SDs on the FFT. Can we honestly state someone who scores almost -7 SDs on the FFT does not have developmental prosopagnosia? I do not believe so. Especially when such cases may be spared in learning unfamiliar faces over several seconds, as tested by the CFMT, but fail to effectively recollect identity for long-term recognition, as is required in the FFT and daily life. Thus, while adding a DSM-5 diagnostic option of impaired on a single task at the -2 SD level may counter these issues, the data shows the current DSM-5 excludes many objectively atypical DP cases.

Another problem with requiring two tests to diagnose neurocognitive disorders is that only one of them may have perfect, or close to perfect, validity. In DP, this would likely be some form of assessment that captures their failings to recognise personally familiar people in the real world. Even if this hypothetical task could detect impairment at the -1 SD level in all cases, no other task would arguably be as good. This is because the defining problems of the condition are captured by this perfect test, and any other test will likely detect related, albeit imperfect, peripheral aspects of their day-to-day difficulties. As a result, requiring impairment on a second test will inevitably exclude patients from a diagnosis. This is demonstrated in a hypothetical famous faces test (bottom right panel of Figure 1) capable of detecting impairment in 100% of DP cases at the -1 SD level. Despite this perfect assessment, overall power to diagnose a patient will be equal to that provided by the second test, namely the Cambridge Face Memory Test (79%), with many atypical FFT cases missed. Thus, if one

cognitive task is perfect for diagnosing, then we should simply use one, rather than redundantly introducing a second as the DSM-5 recommends.

Relatedly, the DSM-5 only describes in general terms which two tests should be used to diagnose impairments within each of its six cognitive domains (DSM-5, 2014). For example, in the visual-motor category, face perception and/or recognition tasks are referred to as potential candidates with few details beyond that (DSM-5, 2014). This risks clinicians and researchers viewing perception (e.g., what's makes this face unique?), unfamiliar face recognition (e.g., can I recognise a face that I have been briefly exposed to?), and familiar face recognition (e.g., can I identify a personally known or famous face?) tests as equally valid diagnostic tools. However, a famous faces test that controls for familiarity is currently the best for detecting single case atypicality (e.g., Bate et al., 2019), and the closest that can come to measuring the severity of symptoms described by those with DP (e.g., Bate et al., 2019). Thus, if familiar face recognition is the cognitive construct that is impaired in this group, then we should only use diagnostic tasks that measure this construct. While perception and unfamiliar face recognition tasks may be useful for identifying subtypes of DP, they can lack the validity and sensitivity of the famous faces test. Thus, the DSM-5 does not contain sufficient details that enable clinicians and researchers to make informed choices about which two diagnostic tasks are the most valid.

Another issue with the DSM-5 is that the strength of the relationship between the two tests will introduce a unique bias into the types of patients we can diagnose. This is illustrated in Figure 2, with our 100% powered hypothetical FFT, and a second hypothetical test that only has around 50% power. When the correlation between the two is perfect (i.e., $r = +1$), we will exclusively diagnose the lowest scoring participants on the FFT test. By contrast, if there is no correlation between the two (i.e., $r = 0$), we will sample cases from throughout the FFT distribution. It is important to note that the further the relationship moves from $r = +1$ to

$r = 0$, the probability of excluding cases who are more severely impaired on the FFT increases. Thus, while we are diagnosing a sample that reflects the breadth of the FFT distribution when $r = 0$, we will also exclude many highly atypical cases from a diagnosis too. The DSM-5 requirement of impaired on two tests will therefore introduce a correlation dependent bias in the types of patients we can ever diagnose, study, and treat.

In summary, the DSM-5 approach to diagnosing cognitive impairment is extremely limited due to issues in cognitive and behavioural testing. These include imperfect validities, test-retest reliabilities, and a failure to acknowledge diagnostic tests' limitations. Similarly, requiring impairments on a second task will further exclude objectively atypical cases. By highlighting these pitfalls in DP, I encourage clinicians and researchers working in other neurocognitive disorders to reevaluate their diagnostic methods. Only by doing so can we hope to develop more sensitive, patient-centred, diagnostic approaches.

Consequences of missed diagnoses

While it is important to acknowledge why missed diagnoses occur, it is equally important to recognise the myriad of negative consequences they create. For example, patients can question their own lived-in experiences and sanity after being told there is nothing wrong with them by a medical practitioner (Au et al., 2022; Eyal, 2022; Wise, 2022). This is a frequent occurrence in diverse conditions linked to problems in cognition, such as Long COVID (Au et al., 2022; Eyal, 2022), dementia (Nelson & O'Connor, 2008; Rentz et al., 2000; Rentz et al., 2004), electroconvulsive therapy patients (Rose, 2022; Rose et al., 2003), and DP (Burns et al., 2022). Without a formal diagnosis, patients will not be able to move forward with insurance claims, impacting their ability to acquire support and treatment. This will be particularly problematic if treatments are only effective at an early stage of disease, as those missed from a diagnosis will not be treated in time. Also, an absence of a diagnosis will block patients from legally protected, workplace related, reasonable adjustments to

accommodate their loss of function. This means that missed diagnoses will severely impact the lives of patients with neurocognitive disorders and their families.

The above could be described as patient focused issues, but missed diagnoses will also negatively impact science. For example, if only the most extremely impaired cases who score below -1 SD are diagnosed, then prevalence rates and effect sizes of impairments will never be accurate, owing to missed cases being excluded from the top end of the disorder's performance distribution. Such exclusions will undermine any epidemiological work that seeks to assess the aetiology and outcomes of neurocognitive disorders, as missing cases will potentially bias results and waste vast resources.

Missed cases will also impact neurocognitive models, because their absence risks altering dissociations and associations to such an extent, they render the model's underlying evidence base meaningless. To illustrate this, imagine a hypothetical diagnostic test of face recognition has power of roughly 50% at the -1 SD level, and is correlated with an object recognition test (Figure 3). The effect size of impairment in potentially diagnosed DP cases on our faces test averages -1.8 SDs below a neurotypical mean, with a comorbid impairment of -.62 SD in object recognition. These group level impairments suggest the two processes are to some extent not dissociable. Thus, a cognitive model derived from this data shows overlapping functionality between face and object recognition.

However, our 50% powered diagnostic test has excluded many DP cases. If we include them, then we will find we overestimated our DP group's impairment on this task, as it is now reduced to -1 SD. Importantly, the impairment in object recognition disappears. The resulting cognitive model now favours a dissociation between these processes[4]. Moreover, given object recognition is now intact, the correlation between faces and objects may be

[4] Please note, excluded cases will only affect neurocognitive models if there is a relationship between face and object recognition. If there is no correlation (i.e., $r = 0$), then missed diagnoses will only alter the severity of group level face processing impairments in DP. By contrast, object recognition will not be affected.

explained by another process. For example, attention is frequently cited as the domain general cause of such associations because face and object recognition are believed by many to be reliant upon dissociable brain networks (Kanwisher, 2017; Kanwisher & Yovel, 2006; McKone et al., 2007; although see Burns, Arnold, et al., 2019; Gauthier et al., 2017) Thus, if attention were the cause of the relationship between face and object recognition, then adding data from an attention task as a covariate should abolish the link. Alternatively, the correlation could be due to domain general processes being utilised, but a dissociable face-specific component being impaired, hence the lack of group level object recognition impairments. Either way, exclusions similarly affect our cognitive model's assumptions.

The same problem would also be apparent when identifying abnormalities in structural or functional MRI work. Imagine replacing object recognition in Figure 3 with the neural activation in a brain region. This region's BOLD response is correlated with performance on our hypothetical face recognition test used to diagnose DP. Again, as we excluded many of our DP cases, we find this brain region exhibits a lower-than-normal BOLD response, suggesting it is atypical in DP. When we add the missed cases, this atypicality disappears. Our neural model therefore changes from this region seeming essential for face recognition, to being of limited importance and appearing intact in DP. Remarkably, a recent study showed DP cases exhibited reduced neural responsiveness across much of the brain (Guo et al., 2018). However, as these atypicalities were shown in only a minority of the most extremely impaired DP cases (i.e., diagnosed -2 SDs on two face processing tests), they may as a population appear intact if we had included the full range of potential cases. Missed diagnoses will therefore impact virtually every area of cognitive and neurocognitive science, undermining any trust we have in the literature.

Failing to diagnose patients correctly also means treatments can erroneously appear effective due to statistical artefacts. This is because when we only diagnose and treat a

sample that perform the lowest on cognitive assessments, post-treatment improvements identified through retests risk being nothing more than a regression to the mean (Barnett et al., 2005; Finney, 2008; Morton & Torgerson, 2005). That is, patients who score exceptionally poorly in the first instance to gain a diagnosis, will typically perform better on a retest due to the simple fact that there is only one direction their scores can go post treatment. Conversely, patients with the mildest impairments that fail to acquire a diagnosis may be the most responsive to interventions due to their residual cognitive abilities. Unfortunately, as they have been blocked from acquiring a diagnosis, they will never be included in clinical trials that assess a treatment's efficacy. In summary, failed diagnoses will have a substantial, negative impact upon patients, their families, and science.

Validating a Symptom-Based Approach to Diagnosing

Given the shortcomings of cognitive tests as diagnostic tools, we must explore viable alternatives when substantial overlaps exist between the performance distributions of neurotypicals and neurocognitive disorders, i.e., when the DSM-5 approach proves ineffective at distinguishing between impaired versus intact cognitive abilities at the level of the individual patient. One solution might be a symptom questionnaire if it were shown more effective at differentiating a patient's complaints from neurotypicals. If so, we must validate it. One way of accomplishing this would be to pool missed cases' cognitive task data together to enhance power for detecting objective impairments at the group level. This will transform missed patients' atypical levels of symptoms from mere subjective complaints, into validated indices of underlying cognitive deficits. Moreover, the remaining patients' self-reported complaints will have been validated at the -1 SD level on two cognitive tasks via the DSM-5 approach. Given these dual approaches to detecting impairment across the whole patient sample, a symptom-based method, if more effective, would become a viable diagnostic alternative to the DSM-5.

As a proof of concept, I assessed a large sample (n = 61) of self-identified developmental prosopagnosia cases. I first quantified the proportion of these individuals that would fail to acquire a diagnosis through the DSM-5, then sought to identify their group-level impairments on multiple tasks to validate a symptom-based approach to diagnosing. Finally, I replicated these deficits in a separately collected sample (Bate et al., 2019), before assessing all data using meta-analyses. If missed cases exhibit deficits in face processing, it will confirm the DSM-5 fails to diagnose objectively impaired DP. This would mean the one-size-fits-all approach recommended by the DSM-5 does not always work, as it blocks people with objective impairments from acquiring a diagnosis. Instead, we should tailor diagnostic approaches to the specific cognitive impairments we are assessing.

2. Methods

2.1. Transparency and Openness

All data required to replicate our results is available on the Open Science Framework (https://osf.io/3x86n/). As I do not own the copyright for the tasks, I do not make them available. The PI20 is freely available in the original paper (Shah et al., 2015).

2.2.Participants

I report how I determined my sample size, all data exclusions, all manipulations, and all measures in the study. The first DP sample comprised 62 cases whose ages ranged between 18-72 years old (M = 41.5, SD = 14), with nine males, and three who identified as neither male nor female. All cases reported severe, lifelong troubles with faces with no obvious historical reason for it being acquired (Burns et al., 2023). Due to COVID-19 restrictions, all tests were carried out online. One DP case was excluded for failing two out of two attention checks throughout the tasks. Another failed to move any faces during the Cambridge Face Perception Test: i.e., 62-year-old female who made 96 errors on upright and inverted (Table 1). Presumably, she had a problem with her mouse on this task, so included her other data

where possible. This left 61 DP cases for all assessments, except for the CFPT related measures (n = 60). Data was not collected on participants' other cultural backgrounds (e.g., socioeconomic status; Applebaum et al., 2018).

Fifty-two controls participated: ages ranged 22-68 years old (M = 38.8, SD = 11.3) with 31 males. While there were many more males than females in this group, and women typically recognise faces better than men (Herlitz & Lovén, 2013), there is a concern this may result in an underestimation of face recognition difficulties in our DP sample. I ran independent t-tests on the control data from all measures and found no significant gender related effects [all $ps > .08$, BF10 ≤ 1], thus do not include gender in any of my analyses.

Three control males reported lifelong troubles with faces, and scored highly atypical (i.e., significant at the one-tailed alpha threshold) on the prosopagnosia index questionnaire (74, 68 and 63), suggesting they likely have DP. I therefore excluded them from all analyses.

Power was difficult to estimate *a priori*, as it was unclear just how many cases would fail to acquire a diagnosis using the DSM-5 approach, or what their level of impairment might be. Given the DSM-5 diagnostic criterion is -1 SD on two cognitive tests, and the average level of cognitive impairment in the self-identified DP cases is just below -2 SDs on the CFMT (Burns et al., 2023), I presumed the group-based deficits in the DSM-5 missed cases would be close to -1 SD. This is because their distribution was potentially skewed (i.e., being the high performing tail-end of the normally distributed DP group), with most cases congregating around this level. An effect size (*Cohen's d* = .8) was chosen based on this hypothesis. This required 26 participants in the excluded DP group (i.e., those who failed to meet DSM-5 criteria) and 26 in the control group, to detect effects with an alpha of .05 and power of 80%. A *post hoc* power analysis based on the 23 DP cases who were missed by the DSM-5 approach, and 48 controls, suggested that power was 87%, with 80% power to detect down to *Cohen's d* = .72.

To avoid reducing power further, I do not make any corrections for multiple comparisons, especially given these are difficult to recruit cases. However, to reassure readers the results were not false positives, I largely replicated my findings in an additional DP group (Bate et al., 2019). Moreover, I ran *p*-curve meta-analyses to ensure any impairments detected in cases excluded from a DSM-5 diagnosis had evidential value (Simonsohn et al., 2014).

None of our experiments or hypotheses were formally pre-registered. While low-level object recognition difficulties were not tested (e.g., Birmingham Object Recognition Battery: BORB; Riddoch & Humphreys, 2022), no DP cases disclosed general problems with vision, and when I historically included the BORB in DP testing, those with the condition did not exhibit problems. This has been objectively confirmed across 200 cases in recent papers (Bate et al., 2019; Lowes et al., 2023), with only one DP case exhibiting possible difficulty (Lowes et al., 2023). Thus, despite some concerns about omitting such testing (Nørkær et al., 2024), low-level vision problems are no more prevalent in DP than neurotypicals, meaning that such tasks are unnecessarily onerous for DP testing in the absence of patient complaints.

2.3. Procedure and Materials

A battery of cognitive tests that have historically been used to diagnose developmental prosopagnosia (Burns et al., 2023; DeGutis et al., 2023) were completed by all participants. These included a 72-trial assessment of unfamiliar face memory (the Cambridge Face Memory Test: CFMT; Duchaine & Nakayama, 2006), an unfamiliar face perception test which included eight upright trials and eight inverted (the Cambridge Face Perception Test: CFPT; Duchaine et al., 2007), and a validated, 30-trial familiar face memory test (i.e., Famous Faces Test: FFT) developed by my own lab (Burns et al., 2023). FFT scores were not corrected for participants' familiarity (i.e., we did not provide celebrities' names afterwards to adjust participants' possible scores based on these responses). Participants also completed

the Prosopagnosia Index questionnaire (Shah et al., 2015). All DP cases' individual scores and full test details are presented in Supplementary Information here (https://osf.io/3x86n/).

The Bate et al. (2019) sample completed their own FFT, the CFMT and only the eight CFPT upright trials. Bate et al.'s (2019) FFT differed from mine in a few ways. First, they used two versions, one for participants aged under 35 years old, and one for participants aged ≥35; faces were selected based on pilot work with these age groups. Both versions used 60 faces, and in contrast to my FFT, each participant's total possible correct scores were adjusted by removing celebrities they were not familiar with by name. Full details of Bate et al.'s (2019) methods can be found in their open access paper.

Ethical approval was granted by Edge Hill University Ethics Review Board, with all work carried out in accordance with the 1964 Helsinki Declaration on human testing. All participants gave informed consent, and for their anonymised data to be published.

3. Results

3.1. The DSM-5 diagnoses 62% of DP cases, the symptom-based approach 100%

The DSM-5 approach to neurocognitive disorders requires participants to score more poorly than -1 SD from the neurotypical mean on two cognitive tests. DeGutis et al. (2023) recommended that these should be two tests of face memory when diagnosing DP. I used this guidance with the CFMT and FFT because they have historically been the two most widely used tests to diagnose DP (Burns et al., 2023; DeGutis et al., 2023). This revealed a striking 38% of self-identified DP cases were excluded from a diagnosis. By contrast, 100% were classified as atypical using the prosopagnosia index via a Crawford's *t*-test (Crawford & Howell, 1998).

We wanted to assess whether the DSM-5 approach diagnosed DP cases who reported more severe symptoms than those that were excluded. To ensure higher power to detect potential symptom differences between the groups, and because all DP cases were highly

abnormal in their self-reported symptoms at the individual level (i.e., more than -2 SDs below a neurotypical mean), I performed a between participants *t*-test on the DSM-5 diagnosed and excluded DP cases' PI20 scores, i.e., I did not include control data. This revealed excluded cases reported fewer problems with faces [M = 79, SD = 8.17] than those who acquired a diagnosis using the DSM-5 approach [M = 83.61, SD = 6.84, $t(59) = 2.37$, $p = .021$, *Cohen's* $d = .63$]. The DSM-5 approach therefore seems to capture some of the symptom differences between those who are diagnosed and those that are not. Figure 4 illustrates the mean level of z-score impairment for both the DP groups across all measures.

3.2.Excluded DP cases are impaired in unfamiliar face memory

To validate excluded cases' highly abnormal symptom complaints, we compared their CFMT scores to the neurotypical group (Figure 4). This confirmed they were impaired in unfamiliar face memory abilities [Excluded DP M = 51.52 trials correct, SD = 8.39; Neurotypical M = 59.7, SD = 9.8, $t(69) = 3.43$, $p < .001$, *Cohen's* $d = .87$]. Thus, the DSM-5 approach to diagnosing neurocognitive disorders excludes DP cases who exhibit objective difficulties with unfamiliar faces.

3.3. Excluded DP cases are impaired in upright face perception

The Cambridge Face Perception Test (Duchaine, Germine & Nakayama, 2007) is often used in neuropsychological assessment to identify DP cases who suffer from the apperceptive form of the disorder (Biotti & Cook, 2018), i.e., those that have difficulties telling faces apart from one another. I used the CFPT to assess whether the DSM-5 approach would exclude DP cases who were objectively impaired in face perception. Confirming this, they made significantly more errors on this task [M = 41.9 errors, SD = 12.02] than the control group [M =33.33, SD = 14.16, $t(69) = 2.5$, $p = .015$, *Cohen's* $d = .64$]. This means we have validated excluded cases' self-reported complaints in unfamiliar face memory and face perception.

3.4.Excluded DP cases' holistic perception abilities are impaired

Holistic perception is characterised by the brain encoding an interaction between a face's features to create a salient, unitary percept (Burns & Wilcockson, 2019; Dal Lago et al., 2023; 2024; Luo et al., 2017; Maw et al., 2023), with reduced abilities a suggested cause of face processing difficulties in DP (Avidan et al., 2011; DeGutis et al., 2012; although see: Biotti et al., 2019). I assessed whether excluded DP cases were impaired in holistic perception using the corrected inversion scores from the CFPT as an index (Supplementary Information). This revealed that excluded cases [M = .69, SD = .53] exhibited smaller inversion effects in comparison to our controls [M = 1.16, SD = .9, t(66.04) = 2.74, p = .008, *Cohen's d* = .58]. We have therefore validated excluded DP cases' symptom complaints via three objective measures: CFMT, CFPT upright and CFPT Holistic Perception.

3.5. Excluded DP cases exhibit impairments when judging famous faces

DSM-5 excluded DP cases exhibited deficits in face perception and unfamiliar face recognition. However, familiar faces, such as friends and celebrities, are thought processed in a partially distinct way from unfamiliar faces (Ellis et al., 1979; Johnston & Edmonds, 2009; Megreya & Burton, 2006). I therefore assessed the presence of familiar face impairments in excluded cases using famous faces. This confirmed those who failed to acquire a diagnosis through the DSM-5 [M = 17.87 trials correct, SD = 4.75] identified fewer celebrity faces in comparison to controls [M = 20.65 correct, SD = 6.2, t(55.34) = 2.08, p = .042, *Cohen's d* = .48]. This means excluded DP cases exhibited objective impairments on all four measures of face processing.

3.6. Replication: DSM-5 excluded DP cases' exhibit face memory impairments

Bate et al. (2019) used the CFPT upright, CFMT and their own FFT to test 165 self-identified DP cases. Strikingly, 30% of these individuals failed to meet the DSM-5 criteria (Table 1). While slightly lower than the 38% I found with my sample, it largely replicated

this figure, i.e., a substantial number of potential DP cases will fail to acquire a diagnosis using the DSM-5 method. Table 1 shows the mean z-scores for my data and Bate et al. (2019) side by side for easy comparison. Importantly, analyses of Bate et al. (2019) replicated my finding of missed cases exhibiting significant unfamiliar and familiar face memory impairments [*ps* < .013], but did not corroborate the upright CFPT difficulties [*p* = .24].

3.7. Meta-analysis support the existence of objective impairments in missed cases

It is increasingly common for researchers to assess multi-experiment evidence presented within their papers through a meta-analysis (e.g., Alves et al., 2017; Van Kuijk et al., 2018). This is an important antidote to the replication crisis psychology has faced in recent years, as it can provide support that any studied effect is real (Simonsohn et al., 2014). The *p*-curve is a widely used (Simonsohn et al., 2014) and effective (Lakens, 2023) meta-analysis technique that only uses significant *p*-values. It is based on the fact that when true effects exist, *p*-values will exhibit a right skew when plotted together; with the highest frequency of values congregating around .01 (Burns & Bukach, 2023; Simonsohn et al., 2014). By contrast, if the null hypothesis is true, then *p*-values should appear flat (i.e., uniformly distributed).

I performed a *p*-curve using all significant *p*-values that confirmed missed cases' objective cognitive impairments from mine and the Bate et al. (2019) samples. This meta-analysis was significant [Full curve: Z = -3.08, p = .001; Half curve: Z = -2.74, p = .003], thus supporting the proposal that DSM-5 missed cases' cognitive impairments contain evidential value. Figure 5 illustrates the right-skewed distribution of *p*-values indicating they reflect a real effect. As the holistic perception related *p*-value may not be entirely independent from the CFPT upright *p*-value, I ran the analysis again with the former excluded: the *p*-curve remained significant [Full curve: Z=-2.92, p < .002; Half curve: Z = -2.83, p < .003]. This was replicated when I replaced the CFPT upright *p*-value with the CFPT holistic perception *p*-value [Full curve: Z = -3.13, p < .001; Half curve: Z = -3.18, p < .001]. In summary, meta-

analyses confirm DSM-5 missed cases' objective impairments in face processing.

4. Discussion

The DSM-5 approach to diagnosing cognitive impairment requires a patient to score beyond -1 SD on two objective tests of behaviour or cognition (Sachdev et al., 2014). I outlined how this approach will result in missed diagnoses, using developmental prosopagnosia as a proof of concept. Testing two large, independently collected samples, I showed the DSM-5 excluded between 30-38% self-identified DP cases. Importantly, both groups of missed cases exhibited impairments in familiar and unfamiliar face memory, with one sample displaying face perception difficulties too[5]. These problems were further confirmed through meta-analyses. The prosopagnosia index questionnaire proved more effective as a diagnostic tool than the DSM-5, with the former identifying significant atypicality in 100% of cases at the individual level, in contrast to the latter that ranged between 62-70%. It is important to recognise that a diagnosis of DP based on a case's symptoms is no longer a purely subjective measure, but reflects multiple, underlying objective impairments. These deficits were found using the DSM-5 approach for individual cases at the -1 SD threshold on two tests, or when that failed, at the level of the group in excluded DP, thus validating the prosopagnosia index as a diagnostic tool (Figure 6).

This study is an important demonstration of how cognitive tests can fail to diagnose a substantial proportion of objectively impaired patients. While I have shown this in one clinically relevant population, it is frequent to find substantial overlaps in the performance distributions of those who report cognitive impairments in other disorders versus those who are neurotypical (Beishon et al., 2019; Burns et al., 2022; Costas-Carrera et al., 2022; Eyal,

[5] There are a few potential reasons why the excluded DP cases from Bate et al. (2019) failed to exhibit perceptual impairments in contrast to my own: their DP group was several years older and contained proportionally fewer women, they excluded cases who were high in autistic traits, and they corrected famous face scores for familiarity which would have changed the types of cases diagnosed. Any, or all, of these differences may have contributed to the absence of perceptual impairments in the Bate et al. (2019) sample, although the direction of effect, while much smaller, was the same.

2022; Nelson & O'Connor, 2008; Potts et al., 2022; Rentz et al., 2000; Rentz et al., 2004). This should pose as a warning to any clinician or researcher who tries to rigidly enforce a diagnostic criterion, such as the DSM-5 cutoff of -1 SD on two cognitive tests, when it clearly does not fit the patient population. Instead, I advocate the use of a transdiagnostic, data driven approach, in which the whole range of the possible patient sample self-reporting complaints are tested (Astle et al., 2022; Burns et al., 2023; Epihova & Astle, 2024), to identify, and correct, deficiencies in diagnostic tests and cutoffs. Only by doing so, and including them in our work, can we hope to improve our methods (Burns et al., 2023; Epihova & Astle, 2024).

If a symptom-based approach is more effective at identifying a patient's atypicality in daily life, then I recommend validating it, and using it, to replace the cognitive or behavioural test-based approach proposed by the DSM-5. We can validate a symptom questionnaire, such as the prosopagnosia index, by identifying group based, objective impairments in missed cases on multiple tasks. By contrast, the remaining DP cases are already validated via the DSM-5 approach at the level of the individual patient. One benefit of using a symptom-based approach to diagnosing DP is that it may not be susceptible to the biases that face recognition tests suffer from. For example, women are typically better at identifying still images of faces in cognitive tests than men (Herlitz & Lovén, 2013; Wright & Sladden, 2003), and people are typically better at recognising faces of their own ethnicity (Bate, Bennetts, Hasshim, et al., 2019; Burns, Tree, et al., 2019; Childs et al., 2021; Estudillo et al., 2020; Meissner & Brigham, 2001), and ages (Rhodes & Anastasi, 2012). Given that standardized cognitive tests (e.g., CFMT, CFPT) almost always contain images of young adults, they will underestimate face processing abilities of participants furthest away from these age groups (Burns, 2023). While some may have concerns that a symptom-based approach is susceptible to malingering patients (i.e., they can be easily faked), these issues are equally true of cognitive tests (Suhr et

al., 2008; Suhr et al., 2021). Moreover, an argument could be made that DP cases failing to meet DSM-5 criteria are highly unlikely to be malingerers, given the fact that they perform *above* diagnostic cutoffs on cognitive tasks, i.e., malingerers would likely exaggerate problems on the Famous and Cambridge Face Memory Tests.

Of course, it is theoretically possible a single cognitive test may detect atypicality in every patient. Or maybe when using two tests, at least one will always detect a patient's cognitive impairment at the -2 SD level. In such cases, I would recommend loosening the DSM-5 criteria. For example, if we added an additional option to diagnose based on a single test (i.e., -2 SD cutoff), then we would remedy the DSM-5's problems in missing highly atypical cases. Similarly, incorporating patients' response times, in contrast to the historical reliance on accuracy rates, improves cognitive tests' diagnostic sensitivity (Lowes et al., 2023). Thus, small modifications can result in improvements to the current DSM-5 method. Simply put, I would recommend using the most effective test for providing a diagnosis. This would prove useful, in contrast to the self-reported symptom approach, when patients are lacking awareness of their cognitive difficulties. The important message here for readers is that not every cognitive impairment will be easily diagnosed by the one-size-fits-all approach endorsed by the DSM-5, i.e., impaired at -1 SD on two tests. When you consider the heterogeneity of patients' impairments, and the heterogeneity in cognitive tests' validities and reliabilities, it seems implausible that such a rigid method will work ubiquitously. We must tailor how impairments are diagnosed by assessing our best options within patient samples. Only by using bespoke, rather than general, approaches can we improve patient support and science.

I should add that there is some merit in the DSM-5 method. While I reject it as a diagnostic approach in the context of DP, it did distinguish between self-reported symptoms in those it diagnosed, versus those that it missed. This shows the DSM-5 can reflect cases'

differing levels of symptoms. There is a great deal of interest in the face recognition literature as to whether people have insights into their cognitive abilities (Bobak et al., 2019; Estudillo & Wong, 2021; Gehdu et al., 2023; Gray et al., 2017; Livingston & Shah, 2018; Matsuyoshi & Watanabe, 2021; Nørkær et al., 2023; Oishi et al., 2024; Palermo et al., 2016; Shah et al., 2015; Ventura et al., 2018). I have shown those with developmental prosopagnosia exhibit accurate insights into the existence of their objective impairments, and their severity, given the graded symptom levels between the DP cases on either side of the DSM-5 cutoff. This rejects suggestions that excluded cases are misinterpreting their face recognition abilities, or that they are suffering from a failure of meta-cognition (Arizpe et al., 2019; De Haan, 1999; DeGutis et al., 2023).

Another benefit of the symptom-based approach is that it is much shorter (i.e., a couple minutes) than the battery of cognitive tests we typically ask DP cases to complete (e.g., at least 40-60 minutes). This means a symptom questionnaire, once validated as a diagnostic tool, can save patients and clinicians valuable time in clinical settings. Also, there are no standardised cognitive assessments for DP in all ethnicities as cognitive tests used to diagnose DP are frequently geared towards Caucasian samples (e.g., Duchaine & Nakayama, 2006; Duchaine, Yovel, et al., 2007). This makes them exclusionary given there are substantial performance variations when recognising faces from other ethnicities, e.g., Caucasians will often exhibit problems recognising Asian faces (e.g., Bate et al., 2019; Childs et al., 2021; Meissner & Brigham, 2001). A symptom-based approach should, in theory, negate these issues to some extent.

It is important to note some believe symptoms alone should never be used when diagnosing cognitive impairment (DeGutis et al., 2023; DeHaan, 1999; Nørkær et al., 2024), and that the solution to missed diagnoses is to develop more sensitive experimental tests (DeGutis et al., 2023). I have shown here that pooling missed cases' data reveals group level

impairments that are not otherwise detectable using the DSM-5. Thus, when a patient is diagnosed through the symptom-based approach, we can be confident their atypical symptoms reflect underlying objective impairments on multiple experimental measures (Figure 6). I agree cognitive tests need to be improved in their sensitivity, validity, and reliability, but such improvements will further validate the symptom-based approach. Imagine we develop a cognitive testing battery that matches the PI20's sensitivity, i.e., identifying all cases as atypical at the -2 SD level. Why would we use such time-consuming tasks in overstretched clinical practices when we have a validated, and rapid, symptom questionnaire at our disposal? If we must wait for cognitive tests to improve, we will only perpetuate the problems outlined in the Introduction, and block 30-38% of objectively impaired developmental prosopagnosia cases from a diagnosis. Without the symptom-based approach, such individuals will be unable to access essential treatments, support, and legal protections in the workplace.

I must acknowledge there may be limitations to a symptom-based approach. For example, prior work has shown, albeit not with the prosopagnosia index, that symptom questionnaires can be susceptible to pathologizing normal behaviours in one culture over the other (Norbury & Sparks, 2013). However, given the vast numbers of cases missed when using cognitive tests in DP, potential cross-cultural issues in symptoms are, in my opinion, likely to have a much smaller impact in terms of missed diagnoses. Also, the PI20 (Sun et al., 2021) outperforms the CFMT (Murray & Bate, 2020; Wilmer et al., 2010) in test-retest reliabilities, meaning that a patient's diagnostic status is less susceptible to changing from one day to the next, in contrast to a symptom-based approach. Despite this, it is recommended that a team independent of the scale developers assess such questionnaires, to remove potentially redundant items (Boateng et al., 2018). While this has been done to some extent with the PI20, the new scale was designed to improve the detection of neurotypical face recognition

abilities, not developmental prosopagnosia symptoms (Bobak et al., 2019). This means there may be some benefits from further PI20 refinement. However, in our sample of 61 DP cases, it performed exceptionally well, identifying 100% of cases as suffering atypical levels of prosopagnosia symptoms beyond the neurotypical -2 SD cutoff, i.e., no self-identified DP cases were erroneously reporting symptoms in the neurotypical range.

In summary, I have shed light on the limitations of the DSM-5 approach to diagnosing neurocognitive disorders, using developmental prosopagnosia as a compelling case in point. The conventional DSM-5 criterion of scoring below -1 SD on two objective tasks excludes a significant percentage (i.e., 30-38%) of individuals who report severe problems in daily life. By introducing a symptom-based approach, we have identified excluded DP cases' complaints as significantly atypical in all instances, and validated them through their underlying objective impairments. This offers a more comprehensive and patient-centred perspective on diagnoses, acknowledging the limitations of cognitive tests. These findings, although focused on developmental prosopagnosia, serve as a crucial reminder to clinicians and researchers that diagnostic criteria must be tailored to the unique characteristics of the patient population. Embracing a data-driven approach through such cases' suspected issues can lead to a more effective diagnostic method and improve the accuracy of assessments. If a symptom-based approach demonstrates superior effectiveness in identifying atypicality in daily life, as we have shown here, then it should replace the DSM-5 method. By doing so, we can enhance the diagnostic process, making it more inclusive, unbiased, and ultimately, more reflective of real-world cognitive functioning.

5. Acknowledgements:

I would like to express my deepest gratitude to the developmental prosopagnosia cases who gave up their free time to make this research possible, and two anonymous reviewers for

their helpful comments that improved this paper. I would also like to thank Nicola van Rijsbergen, John Towler, and Jeremy Tree for discussions and feedback on earlier versions of this manuscript.

6. Funding

No funding to report.

7. Conflicts of interest/Competing interests

No conflicts/competing interests to report.

8. Ethics approval

Ethical approval was granted by Edge Hill University Ethics Review Board.

9. Consent to participate

All participants provided informed consent to participate.

10. Consent for publication

All participants gave consent for their anonymized data to be published.

11. Availability of data and materials

The data required to replicate my Results is available on the Open Science Framework (https://osf.io/3x86n/)

12. Code availability

I do not own the copyright for many of the tests used so do not make them available, but I thank Brad Duchaine for providing the CFMT and CFPT. The PI20 can be found in the paper it was first reported on (Shah et al., 2016).

13. Authors' contributions:

I am responsible for all aspects of this paper.

14. References

Alves, H., Koch, A., & Unkelbach, C. (2017). The "common good" phenomenon: Why similarities are positive and differences are negative. *Journal of Experimental Psychology: General*, *146*(4), 512.

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.

Au, L., Capotescu, C., Eyal, G., & Finestone, G. (2022). Long covid and medical gaslighting: Dismissal, delayed diagnosis, and deferred treatment. *SSM-Qualitative Research in Health*, *2*, 100167.

Avidan, G., & Behrmann, M. (2008). Implicit familiarity processing in congenital prosopagnosia. *J Neuropsychol*, *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.

Barnett, A. G., Van Der Pols, J. C., & Dobson, A. J. (2005). Regression to the mean: what it is and how to deal with it. *International journal of epidemiology*, *34*(1), 215-220.

Barton, J. J., 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

Bate, S., Bennetts, R., Hasshim, N., Portch, E., Murray, E., Burns, E., & Dudfield, G. (2019). The limits of super recognition: An other-ethnicity effect in individuals with extraordinary face recognition skills. *Journal of Experimental Psychology: Human Perception and Performance*, *45*(3), 363.

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.

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., & Tree, J. J. (2017). The definition and diagnosis of developmental prosopagnosia. *Q J Exp Psychol (Hove)*, *70*(2), 193-200. https://doi.org/10.1080/17470218.2016.1195414

Behrmann, M., & Avidan, G. (2005). Congenital prosopagnosia: face-blind from birth. *Trends Cogn Sci*, *9*(4), 180-187. https://doi.org/10.1016/j.tics.2005.02.011

Behrmann, M., Avidan, G., Gao, F., & Black, S. (2007). Structural imaging reveals anatomical alterations in inferotemporal cortex in congenital prosopagnosia. *Cerebral Cortex*, *17*(10), 2354-2363.

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

Behrmann, M., & Plaut, D. C. (2013). Distributed circuits, not circumscribed centers, mediate visual recognition. *Trends in cognitive sciences*, *17*(5), 210-219.

Behrmann, M., & Plaut, D. C. (2014). Bilateral hemispheric processing of words and faces: evidence from word impairments in prosopagnosia and face impairments in pure alexia. *Cerebral Cortex*, *24*(4), 1102-1118.

Beishon, L. C., Batterham, A. P., Quinn, T. J., Nelson, C. P., Panerai, R. B., Robinson, T., & Haunton, V. J. (2019). Addenbrooke's Cognitive Examination III (ACE-III) and mini-ACE for the detection of dementia and mild cognitive impairment. *Cochrane Database of Systematic Reviews*(12).

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.

Biotti, F., & Cook, R. (2018). Impaired perception of facial emotion in developmental prosopagnosia: A reply to Van den Stock's commentary. *Cortex*, *101*, 298-299. https://doi.org/10.1016/j.cortex.2017.11.013

Boateng, G. O., Neilands, T. B., Frongillo, E. A., Melgar-Quiñonez, H. R., & Young, S. L. (2018). Best practices for developing and validating scales for health, social, and behavioral research: a primer. *Frontiers in public health*, *6*, 149.

Bobak, A. K., Mileva, V. R., & Hancock, P. J. (2019). Facing the facts: Naive participants have only moderate insight into their face recognition and face perception abilities. *Q J Exp Psychol (Hove)*, *72*(4), 872-881.

Bornstein, B., & Kidron, D. (1959). Prosopagnosia. *Journal of Neurology, Neurosurgery & Psychiatry*, *22*(2), 124-131.

Bukach, C. M., Bub, D. N., Gauthier, I., & Tarr, M. J. (2006). Perceptual expertise effects are not all or none: Spatially limited perceptual expertise for faces in a case of prosopagnosia. *Journal of cognitive neuroscience*, *18*(1), 48-63.

Bukach, C. M., Gauthier, I., Tarr, M. J., Kadlec, H., Barth, S., Ryan, E., Turpin, J., & Bub, D. N. (2012). Does acquisition of Greeble expertise in prosopagnosia rule out a domain-general deficit? *Neuropsychologia*, *50*(2), 289-304.

Bukach, C. M., Le Grand, R., Kaiser, M. D., Bub, D. N., & Tanaka, J. W. (2008). Preservation of mouth region processing in two cases of prosopagnosia. *Journal of Neuropsychology*, *2*(1), 227-244.

Burns, E. (2023). A symptom guided diagnosis of prosopagnosia is valid: a commentary on DeGutis et al.(2023).

Burns, E. J., Arnold, T., & Bukach, C. M. (2019). P-curving the fusiform face area: Meta-analyses support the expertise hypothesis. *Neuroscience & Biobehavioral Reviews*, *104*, 209-221.

Burns, E. J., Bennetts, R. J., Bate, S., Wright, V. C., Weidemann, C. T., & Tree, J. J. (2017). Intact word processing in developmental prosopagnosia. *Scientific Reports*, *7*(1), s41598-41017. https://doi.org/10.1038/s41598-017-01917-8

Burns, E., & Bukach, C. (2023). Meta-analyses reveal the occipital face area is altered in visual expertise.

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*, 1-24.

Burns, E. J., Gaunt, E., Kidane, B., Hunter, L., & Pulford, J. (2023). A new approach to diagnosing and researching developmental prosopagnosia: Excluded cases are impaired too. *Behavior Research Methods*, *55*(8), 4291-4314.

Burns, E. J., Martin, J., Chan, A. H. D., & Xu, H. (2017). Impaired processing of facial happiness, with or without awareness, in developmental prosopagnosia. *Neuropsychologia*, *102*, 217-228. https://doi.org/10.1016/j.neuropsychologia.2017.06.020

Burns, E. J., Tree, J., Chan, A. H., & Xu, H. (2019). Bilingualism shapes the other race effect. *Vision research*, *157*, 192-201.

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

Burns, E., Tree, J., & Weidemann, C. (2013, January). Recognition memory in developmental prosopagnosia: Behavioural and electrophysiological evidence for an impairment of recollection of faces. In Perception (Vol. 42, pp. 78-79). 207 BRONDESBURY PARK, LONDON NW2 5JN, ENGLAND: PION LTD.

Burns, E. J., & Wilcockson, T. D. (2019). Alcohol usage predicts holistic perception: A novel method for exploring addiction. *Addictive Behaviors*, *99*, 106000.

Childs, M. J., Jones, A., Thwaites, P., Zdravković, S., Thorley, C., Suzuki, A., Shen, R., Ding, Q., Burns, E., & Xu, H. (2021). Do individual differences in face recognition ability moderate the other ethnicity effect? *Journal of Experimental Psychology: Human Perception and Performance*.

Costas-Carrera, A., Sánchez-Rodríguez, M. M., Canizares, S., Ojeda, A., Martín-Villalba, I., Primé-Tous, M., Rodríguez-Rey, M. A., Segú, X., Valdesoiro-Pulido, F., & Borras, R. (2022). Neuropsychological functioning in post-ICU patients after severe COVID-19 infection: the role of cognitive reserve. *Brain, Behavior, & Immunity-Health*, *21*, 100425.

Crawford, J. R., & Howell, D. C. (1998). Comparing an individual's test score against norms derived from small samples. *The Clinical Neuropsychologist*, *12*(4), 482-486.

Dal Lago, D., Burns, E., Gaunt, E., Peers, E., Jackson, R. C., & Wilcockson, T. D. (2023). Alcohol Use Predicts Face Perception Impairments and Difficulties in Face Recognition. *Substance Use & Misuse*, *58*(13), 1734-1741.

Dal Lago, D., Burns, E., Jackson, R. C., & Wilcockson, T. D. (2024). Are alcohol-related attentional biases and holistic perception independent processes?. Experimental and Clinical Psychopharmacology.

Dalrymple, K. A., Fletcher, K., Corrow, S., das Nair, R., Barton, J. J., 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.

Dalrymple, K. A., & Palermo, R. (2016). Guidelines for studying developmental prosopagnosia in adults and children. *Wiley Interdiscip Rev Cogn Sci*, *7*(1), 73-87. https://doi.org/10.1002/wcs.1374

de Gelder, B., & Rouw, R. (2000). Configural face processes in acquired and developmental prosopagnosia: evidence for two separate face systems? *Neuroreport, 11*(14), 3145-3150.

De Haan, E. H. (1999). A familial factor in the development of face recognition deficits. *Journal of Clinical and Experimental Neuropsychology, 21*(3), 312-315.

De Haan, E. H., & Campbell, R. (1991). A fifteen year follow-up of a case of developmental prosopagnosia. *Cortex*, *27*(4), 489-509.

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., 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.

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.

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., & Nakayama, K. (2007). No global processing deficit in the Navon task in 14 developmental prosopagnosics. *Social cognitive and affective neuroscience*, *2*(2), 104-113.

Ellis, H. D., Shepherd, J. W., & Davies, G. M. (1979). Identification of familiar and unfamiliar faces from internal and external features: Some implications for theories of face recognition. *Perception*, *8*(4), 431-439.

Epihova, G., & Astle, D. E. (2024). What is developmental about developmental prosopagnosia? *Cortex*.

Estudillo, A. J., Lee, J. K. W., Mennie, N., & Burns, E. (2020). No evidence of other-race effect for Chinese faces in Malaysian non-Chinese population. *Applied Cognitive Psychology*, *34*(1), 270-276.

Estudillo, A. J., & Wong, H. K. (2021). Associations between self-reported and objective face recognition abilities are only evident in above-and below-average recognisers. *PeerJ*, *9*, e10629.

Eyal, G. (2022). Long Covid and Medical Gaslighting.

Finney, J. W. (2008). Regression to the mean in substance use disorder treatment research. *Addiction*, *103*(1), 42-52.

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

Fisher, K., Towler, J., & Eimer, M. (2017). Face identity matching is selectively impaired in developmental prosopagnosia. *Cortex*, *89*, 11-27. https://doi.org/10.1016/j.cortex.2017.01.003

Fisher, K., Towler, J., Rossion, B., & Eimer, M. (2020). Neural responses in a fast periodic visual stimulation paradigm reveal domain-general visual discrimination deficits in developmental prosopagnosia. *Cortex*, *133*, 76-102. https://doi.org/10.1016/j.cortex.2020.09.008

Fox, C. J., Hanif, H. M., Iaria, G., Duchaine, B. C., & Barton, J. J. (2011). Perceptual and anatomic patterns of selective deficits in facial identity and expression processing. *Neuropsychologia*, *49*(12), 3188-3200.

Furl, N., Garrido, L., Dolan, R. J., Driver, J., & Duchaine, B. (2011). Fusiform Gyrus Face Selectivity Relates to Individual Differences in Facial Recognition Ability [Article]. *Journal of cognitive neuroscience*, *23*(7), 1723-1740. http://search.ebscohost.com/login.aspx?direct=true&db=aph&AN=60657117&site=ehost-live

Fysh, M. C., & Ramon, M. (2022). Accurate but inefficient: Standard face identity matching tests fail to identify prosopagnosia. *Neuropsychologia*, *165*, 108119.

Gehdu, B. K., Press, C., Gray, K. L., & Cook, R. (2023). Autistic adults have insight into their relative face recognition ability.

Gray, K. L., 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.

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. https://doi.org/http://dx.doi.org/10.1016/S0010-9452(08)70502-1

Halder, T., Ludwig, K., & Schenk, T. (2024). Binocular rivalry reveals differential face processing in congenital prosopagnosia. *Scientific Reports*, *14*(1), 6687.

Herlitz, A., & Lovén, J. (2013). Sex differences and the own-gender bias in face recognition: A meta-analytic review. *Visual Cognition*, *21*(9-10), 1306-1336.

Herzmann, G., Schweinberger, S. R., Sommer, W., & Jentzsch, I. (2004). What's special about personally familiar faces? A multimodal approach. *Psychophysiology*, *41*(5), 688-701.

Humphreys, K., Avidan, G., & Behrmann, M. (2007). A detailed investigation of facial expression processing in congenital prosopagnosia as compared to acquired prosopagnosia. *Exp Brain Res*, *176*(2), 356-373. https://doi.org/10.1007/s00221-006-0621-5

Jiahui, G., Yang, H., & Duchaine, B. (2018). Developmental prosopagnosics have widespread selectivity reductions across category-selective visual cortex. *Proceedings of the National Academy of Sciences*, *115*(28), E6418-E6427.

Johnston, R. A., & Edmonds, A. J. (2009). Familiar and unfamiliar face recognition: A review. *Memory*, *17*(5), 577-596.

Josephs, K. A., & Josephs, K. A., Jr. (2024). Prosopagnosia: face blindness and its association with neurological disorders. *Brain communications*, *6*(1). https://doi.org/10.1093/braincomms/fcae002

Kanwisher, N. (2017). The quest for the FFA and where it led. *Journal of Neuroscience*, *37*(5), 1056-1061.

Kanwisher, N., & Yovel, G. (2006). The fusiform face area: a cortical region specialized for the perception of faces. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *361*(1476), 2109-2128.

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*(15), 1617-1622.

Kennerknecht, I., Ho, N. Y., & Wong, V. C. (2008). Prevalence of hereditary prosopagnosia (HPA) in Hong Kong Chinese population. *American Journal of Medical Genetics Part A*, *146*(22), 2863-2870.

Kennerknecht, I., Pluempe, N., & Welling, B. (2008). Congenital prosopagnosia--a common hereditary cognitive dysfunction in humans. *Front Biosci*, *13*, 3150-3158. https://doi.org/10.2741/2916

Lakens, D. (2023). P-curves are better at effect size estimation than trim-and-fill (and Michael Jordan is better at free throws than I am). *Authorea Preprints*.

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

Livingston, L. A., & Shah, P. (2018). People with and without prosopagnosia have insight into their face recognition ability. *Q J Exp Psychol (Hove)*, *71*(5), 1260-1262.

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.

Lowes, J., Hancock, P. J., & Bobak, A. K. (2023). Balanced Integration Score: A new way of classifying Developmental Prosopagnosia.

Luo, C., Burns, E., & Xu, H. (2017). Association between autistic traits and emotion adaptation to partially occluded faces. *Vision research*, *133*, 21-36.

Manippa, V., Palmisano, A., Ventura, M., & Rivolta, D. (2023). The Neural Correlates of Developmental Prosopagnosia: Twenty-Five Years on. *Brain sciences*, *13*(10), 1399.

Marotta, J., Genovese, C., & Behrmann, M. (2001). A functional MRI study of face recognition in patients with prosopagnosia. *Neuroreport*, *12*(8), 1581-1587.

Matsuyoshi, D., & Watanabe, K. (2021). People have modest, not good, insight into their face recognition ability: a comparison between self-report questionnaires. *Psychological Research*, *85*(4), 1713-1723.

Maw, K., Burns, E., & Beattie, G. (2023). Prosopagnosia is highly comorbid in developmental coordination disorder (DCD). PsyArxiv

Maw, K. J., Beattie, G., & Burns, E. J. (2024). Cognitive strengths in neurodevelopmental disorders, conditions and differences: A critical review. *Neuropsychologia*, 108850.

McConachie, H. R. (1976). Developmental Prosopagnosia. A Single Case Report. *Cortex*, *12*(1), 76-82. https://doi.org/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.

McKone, E., Kanwisher, N., & Duchaine, B. C. (2007). Can generic expertise explain special processing for faces? *Trends in cognitive sciences*, *11*(1), 8-15.

Megreya, A. M., & Burton, A. M. (2006). Unfamiliar faces are not faces: Evidence from a matching task. *Memory & Cognition*, *34*(4), 865-876.

Meissner, C. A., & Brigham, J. C. (2001). Thirty years of investigating the own-race bias in memory for faces: A meta-analytic review. In: American Psychological Association.

Morton, V., & Torgerson, D. J. (2005). Regression to the mean: treatment effect without the intervention. *Journal of Evaluation in Clinical Practice*, *11*(1), 59-65.

Murray, E., & Bate, S. (2020). Diagnosing developmental prosopagnosia: repeat assessment using the Cambridge Face Memory Test. *Royal Society open science*, *7*(9), 200884.

Nelson, A. P., & O'Connor, M. G. (2008). Mild cognitive impairment: a neuropsychological perspective. *CNS spectrums*, *13*(1), 56-64.

Norbury, C. F., & Sparks, A. (2013). Difference or disorder? Cultural issues in understanding neurodevelopmental disorders. *Developmental psychology*, *49*(1), 45.

Nørkær, E., Guðbjörnsdóttir, E., Roest, S. B., Shah, P., Gerlach, C., & Starrfelt, R. (2023). The Danish Version of the 20-Item Prosopagnosia Index (PI20): Translation, Validation and a Link to Face Perception. *Brain sciences*, *13*(2), 337.

Oishi, Y., Aruga, K., & Kurita, K. (2024). Relationship between face recognition ability and anxiety tendencies in healthy young individuals: A prosopagnosia index and state-trait anxiety inventory study. *Acta Psychologica*, *245*, 104237.

Palermo, R., Rossion, B., Rhodes, G., Laguesse, R., Tez, T., Hall, B., Albonico, A., Malaspina, M., Daini, R., & Irons, J. (2016). Do people have insight into their face recognition abilities? *The Quarterly Journal of Experimental Psychology*, 1-16.

Pihlaja, R. E., Kauhanen, L.-L. S., Ollila, H. S., Tuulio-Henriksson, A. S., Koskinen, S. K., Tiainen, M., Salmela, V. R., Hästbacka, J., & Hokkanen, L. S. (2023). Associations of subjective and objective cognitive functioning after COVID-19: A six-month follow-up of ICU, ward, and home-isolated patients. *Brain, Behavior, & Immunity-Health*, *27*, 100587.

Potts, C., Richardson, J., Bond, R., Price, R., Mulvenna, M., Zvolsky, P., Harvey, M., Hughes, C., & Duffy, F. (2022). Reliability of Addenbrooke's Cognitive Examination III in differentiating between dementia, mild cognitive impairment and older adults who have not reported cognitive problems. *European Journal of Ageing*, *19*(3), 495-507.

Ramon, M., & Gobbini, M. I. (2018). Familiarity matters: A review on prioritized processing of personally familiar faces. *Visual Cognition*, *26*(3), 179-195.

Rentz, D. M., Calvo, V. L., Scinto, L. F., Sperling, R. A., Budson, A. E., & Daffner, K. R. (2000). Detecting early cognitive decline in high-functioning elders. *Journal of Geriatric Psychiatry*.

Rentz, D. M., Huh, T. J., Faust, R. R., Budson, A. E., Scinto, L. F., Sperling, R. A., & Daffner, K. R. (2004). Use of IQ-adjusted norms to predict progressive cognitive decline in highly intelligent older individuals. *Neuropsychology*, *18*(1), 38.

Rhodes, M. G., & Anastasi, J. S. (2012). The own-age bias in face recognition: a meta-analytic and theoretical review. *Psychological bulletin*, *138*(1), 146.

Riddoch, J. M., & Humphreys, G. W. (2022). *BORB: Birmingham object recognition battery*. Psychology Press.

Righart, R., & de Gelder, B. (2007). Impaired face and body perception in developmental prosopagnosia. *Proc Natl Acad Sci U S A*, *104*(43), 17234-17238. https://doi.org/10.1073/pnas.0707753104

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. *Front Hum Neurosci*, *8*, 925. https://doi.org/10.3389/fnhum.2014.00925

Rosenthal, G., Tanzer, M., Simony, E., Hasson, U., Behrmann, M., & Avidan, G. (2017). Altered topology of neural circuits in congenital prosopagnosia. *Elife*, *6*. https://doi.org/10.7554/eLife.25069

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.

Schultz, I. Z. (2010). Neurocognitive disorders in DSM-V: Forensic perspective. *Psychological Injury and Law*, *3*, 271-288.

Schultz, I. Z. (2013). DSM-5 neurocognitive disorders: Validity, reliability, fairness, and utility in forensic applications. *Psychological Injury and Law*, *6*, 299-306.

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.

Simonsohn, U., Nelson, L. D., & Simmons, J. P. (2014). P-curve: a key to the file-drawer. *Journal of Experimental Psychology: General*, *143*(2), 534.

Song, S., Garrido, L., Nagy, Z., Mohammadi, S., Steel, A., Driver, J., Dolan, R. J., Duchaine, B., & Furl, N. (2015). Local but not long-range microstructural differences of the ventral temporal cortex in developmental prosopagnosia. *Neuropsychologia*, *78*, 195-206.

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.

Suhr, J., Hammers, D., Dobbins-Buckland, K., Zimak, E., & Hughes, C. (2008). The relationship of malingering test failure to self-reported symptoms and neuropsychological findings in adults referred for ADHD evaluation. *Archives of Clinical Neuropsychology*, *23*(5), 521-530.

Suhr, J. A., Bryant, A., & Cook, C. (2021). Malingering of Executive Functioning in Head Injury Litigation. *Detection of Malingering during Head Injury Litigation*, 257-287.

Sun, W., Wang, Y., Wang, J., & Luo, F. (2021). Psychometric Properties of the Chinese version of the 20-item Prosopagnosia Index (PI20). E3S Web of Conferences,

Taylor, M. J., Arsalidou, M., Bayless, S. J., Morris, D., Evans, J. W., & Barbeau, E. J. (2009). Neural correlates of personally familiar faces: parents, partner and own faces. *Human brain mapping*, *30*(7), 2008-2020.

Thomas, C., Avidan, G., Humphreys, K., Jung, K.-j., Gao, F., & Behrmann, M. (2008). Reduced structural connectivity in ventral visual cortex in congenital prosopagnosia.

Thomas, C., Avidan, G., Humphreys, K., Jung, K. J., Gao, F., & Behrmann, M. (2009). Reduced structural connectivity in ventral visual cortex in congenital prosopagnosia. *Nat Neurosci*, *12*(1), 29-31. https://doi.org/10.1038/nn.2224

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

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

Towler, J., Parketny, J., & Eimer, M. (2016). Perceptual face processing in developmental prosopagnosia is not sensitive to the canonical location of face parts. *Cortex*, *74*, 53-66. https://doi.org/10.1016/j.cortex.2015.10.018

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.

Van den Stock, J., Van De Riet, W., Righart, R., & de Gelder, B. (2008). Neural correlates of perceiving emotional faces and bodies in developmental prosopagnosia: an event-related fMRI-study. *PLoS One*, *3*(9), e3195-e3195.

Van Kuijk, I., Verkoeijen, P., Dijkstra, K., & Zwaan, R. A. (2018). The effect of reading a short passage of literary fiction on theory of mind: A replication of Kidd and Castano (2013). *Collabra: Psychology*, *4*(1).

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. *Q J Exp Psychol (Hove)*, *71*(12), 2677-2679.

Volfart, A., & Rossion, B. (2024). The neuropsychological evaluation of face identity recognition. *Neuropsychologia*, 108865.

Wakefield, J. C. (2013). DSM-5: An overview of changes and controversies. *Clinical Social Work Journal*, *41*, 139-154.

Wiese, H., Hobden, G., Siilbek, E., Martignac, V., Flack, T. R., Ritchie, K. L., Young, A. W., & Burton, A. M. (2022). Familiarity is familiarity is familiarity: Event-related brain potentials reveal qualitatively similar representations of personally familiar and famous faces. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *48*(8), 1144.

Wilcockson, T. D., 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.

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*, *107*(11), 5238-5241.

Wise, J. (2022). Sixty seconds on... medical gaslighting. In: British Medical Journal Publishing Group.

Wright, D. B., & Sladden, B. (2003). An own gender bias and the importance of hair in face recognition. *Acta Psychologica*, *114*(1), 101-114.

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.

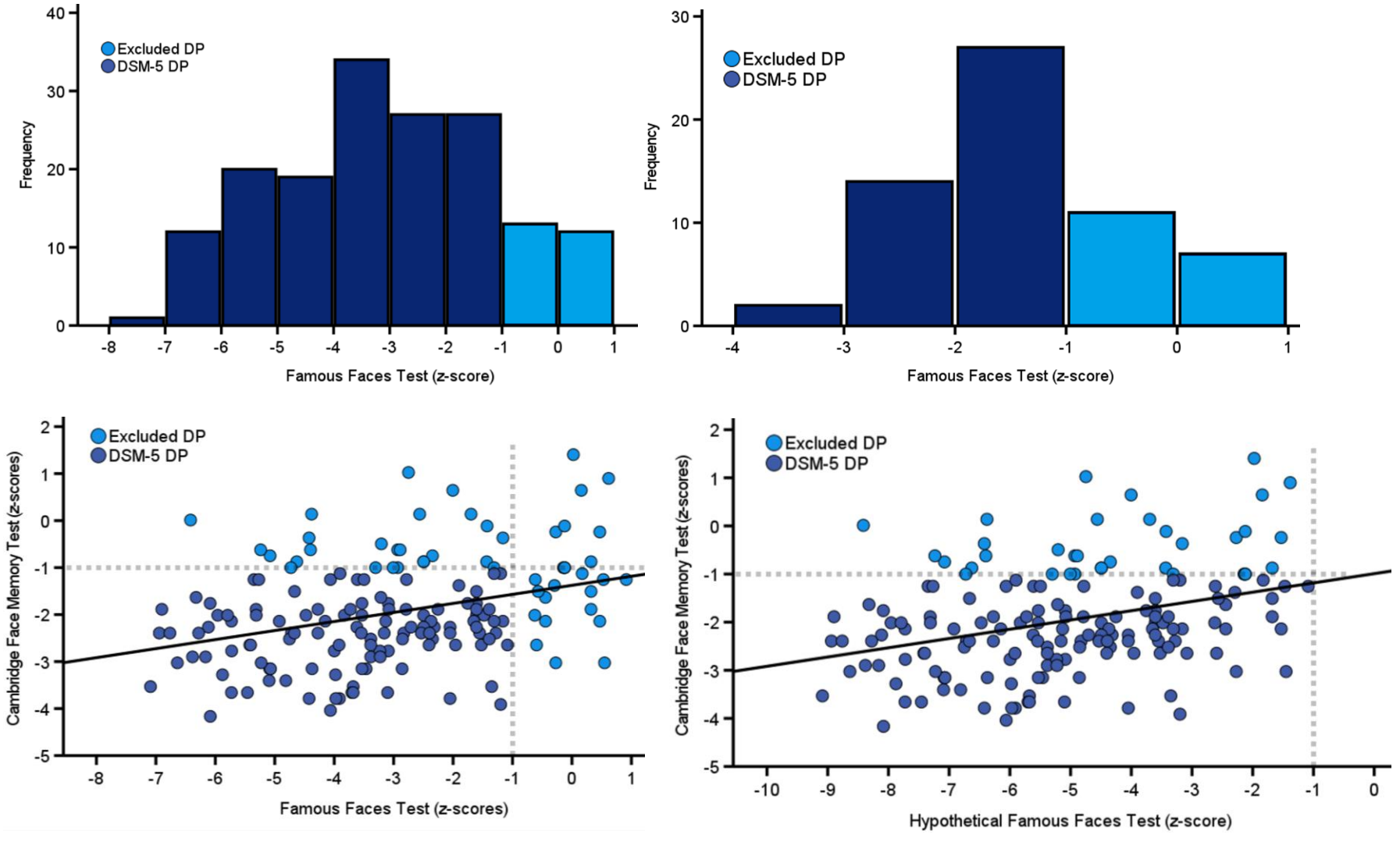

Figure 1. The top left and right panels demonstrate the -1 SD cutoff removes the top end (light blue) of the DP FFT performance distribution, with 15% (top left: Bate et al., 2019) and 30% (top right: Burns et al., 2022) failing to meet criteria. In the bottom left panel, using Bate et al. (2019) data, the -1 SD cutoff on FFT (x-axis) has power to detect impairment in 85% of cases, while CFMT (y-axis) has 79% power. Requiring deficits on both means power can never be greater than the weakest of the two, with only 70% of self-identified DP cases (dark blue) meeting DSM-5 criteria and 30% excluded (light blue). In the bottom right panel, we present a hypothetical famous faces test that has perfect sensitivity, i.e., 100% power at the -1 SD level. Unfortunately, as we require impairment on a second imperfect test that has 79% power (i.e., the CFMT), we will only ever diagnose at this rate. This is despite some of the DP cases scoring almost -9 SDs on the FFT. Please note, I simply subtracted -2 from the Bate et al. (2019) famous faces test data to create this hypothetical data.

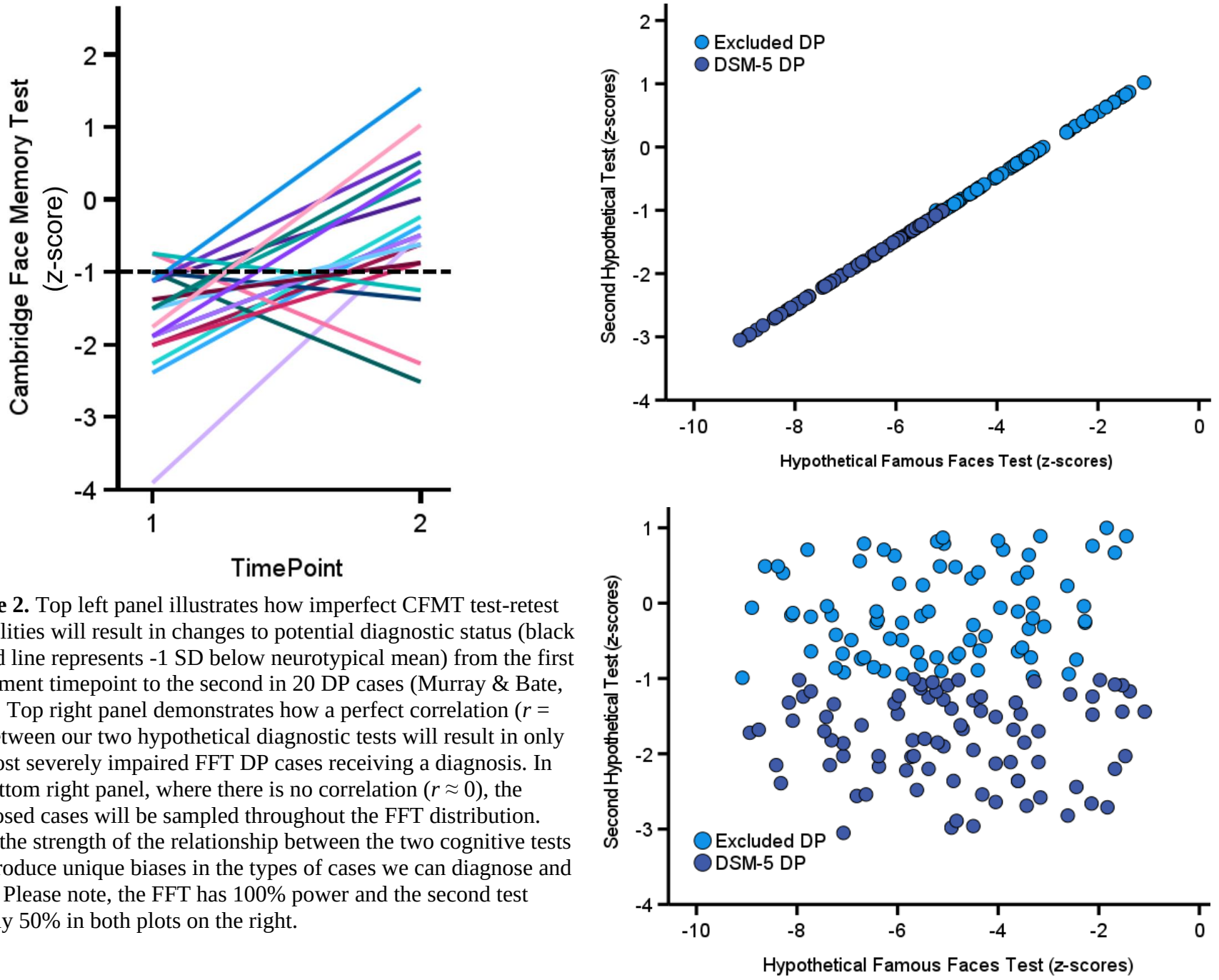

Figure 2. Top left panel illustrates how imperfect CFMT test-retest reliabilities will result in changes to potential diagnostic status (black dashed line represents -1 SD below neurotypical mean) from the first assessment timepoint to the second in 20 DP cases (Murray & Bate, 2020). Top right panel demonstrates how a perfect correlation (r = +1) between our two hypothetical diagnostic tests will result in only the most severely impaired FFT DP cases receiving a diagnosis. In the bottom right panel, where there is no correlation ($r \approx 0$), the diagnosed cases will be sampled throughout the FFT distribution. Thus, the strength of the relationship between the two cognitive tests will produce unique biases in the types of cases we can diagnose and study. Please note, the FFT has 100% power and the second test roughly 50% in both plots on the right.

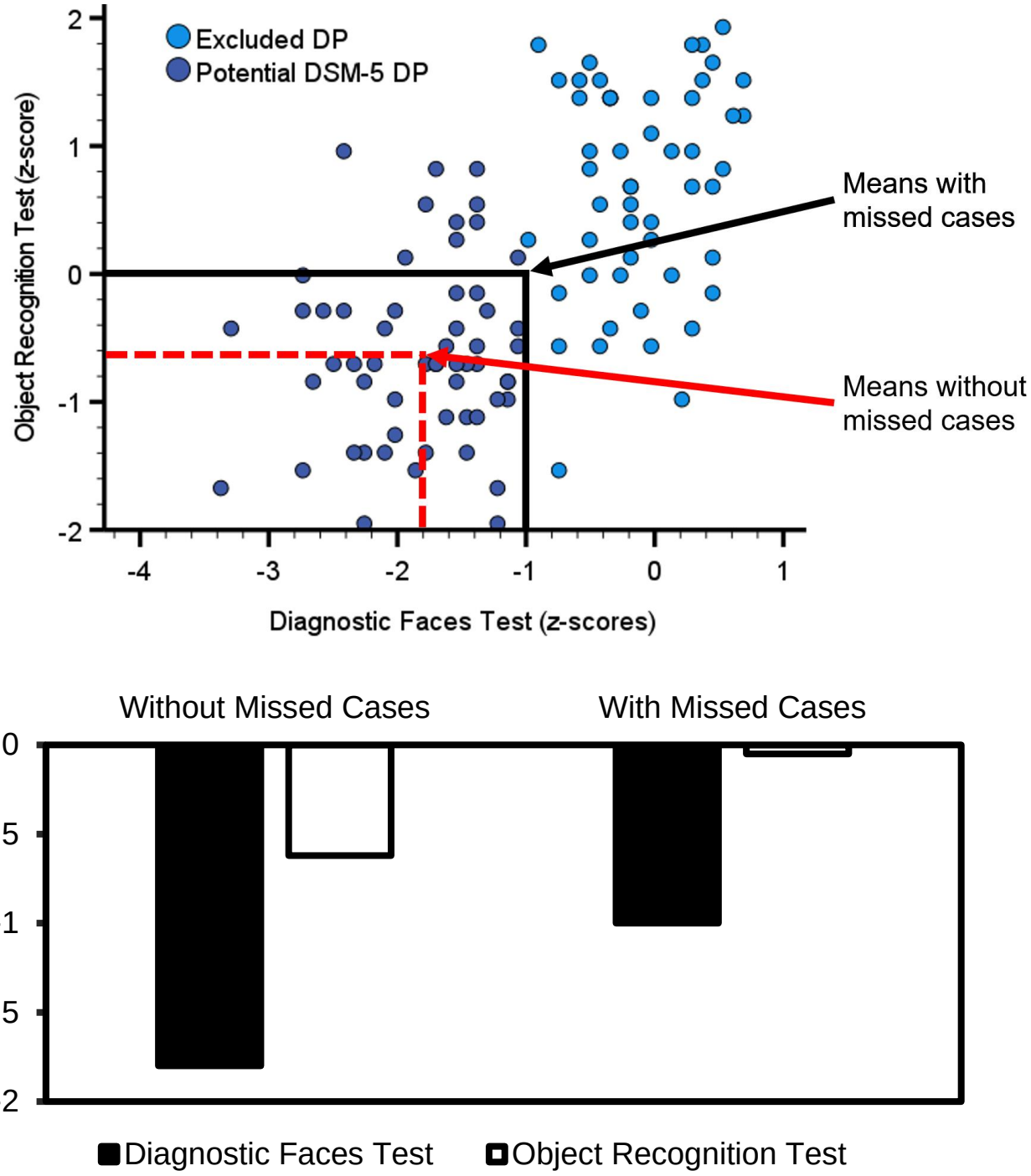

Figure 3. In the top panel, we present a hypothetical diagnostic faces test (y-axis) that is correlated with a hypothetical object recognition test (x-axis). The -1 SD cutoff excludes roughly 50% of potential DP cases on this single test (light blue circles). When we plot the mean z-scores in the bottom panel, you can see that without the missed cases, DP is associated with face *and* object recognition impairments. This supports a cognitive model of shared underlying processes for the two abilities. However, when we include the DP cases who have been excluded from a potential DSM-5 diagnosis due to the -1 SD cutoff, we find estimates of face recognition impairments in this group becoming milder, and object recognition deficits disappear. We now create a cognitive model that dissociates these two abilities. These issues will also be apparent in neurocognitive models if we use replace object recognition with neuroimaging measures, such as an fMRI BOLD response. Please note, I only used one diagnostic test here. These problems will likely become more pronounced with the introduction of a second diagnostic test as the DSM-5 requires.

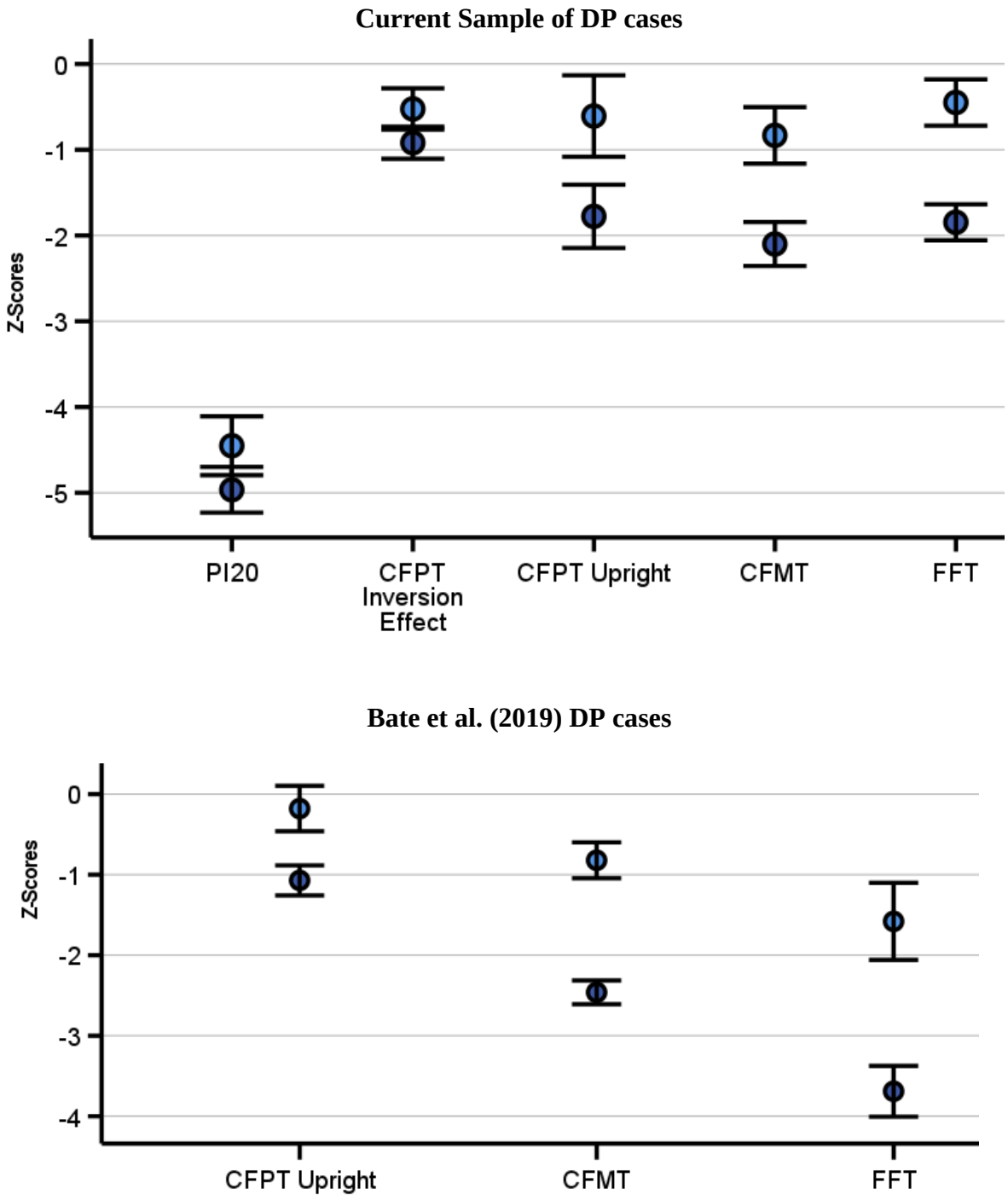

Figure 4. DSM-5 missed DP cases are objectively impaired on multiple cognitive measures. In the top panel, DP cases excluded from a DSM-5 diagnosis (light blue circles) exhibited highly atypical prosopagnosia symptoms (i.e., PI20) and objective impairments on every cognitive task related measure. These deficits validate their subjective complaints. Please note, the DSM-5 DP group's mean z-scores are plotted for context (dark blue): all were significantly milder on all measures [all *ps* < .022]. In the bottom panel, the excluded DP cases from Bate et al. (2019) were impaired on the CFMT and FFT. While the CFPT upright z-scores were not significantly different from controls, they were in the same direction (i.e., impaired) as my DP sample. Please note, Bate et al. (2019) FFT impairments were likely much larger than my sample because the former controlled for participants' familiarity with the faces, while I did not.

DSM-5 Missed Cases	Current Sample	Bate et al. (2019)
% of Total DPs	38%	30%
CFPT Upright	-.6* (.85)	-.18 (.91)
CFPT Holistic	-.52* (.52)	N/A
CFMT	-.83* (.86)	-.82* (.94)
FFT	-.45* (.77)	-1.58* (1.91)
PI20	-4.45* (.91)	N/A

Table 1. Percentage of DSM-5 missed DP cases (top row) and their mean impairment z-scores from the current sample and Bate et al. (2019). While my DSM-5 missed cases were impaired on all measures, Bate et al.'s (2019) cases were only impaired on CFMT and FFT. Please note, Bate et al. (2019) DPs did not complete the PI20 or inverted portion of the CFPT. Asterisks indicate significant impairments relative to neurotypical controls at <.05, SDs are in brackets.

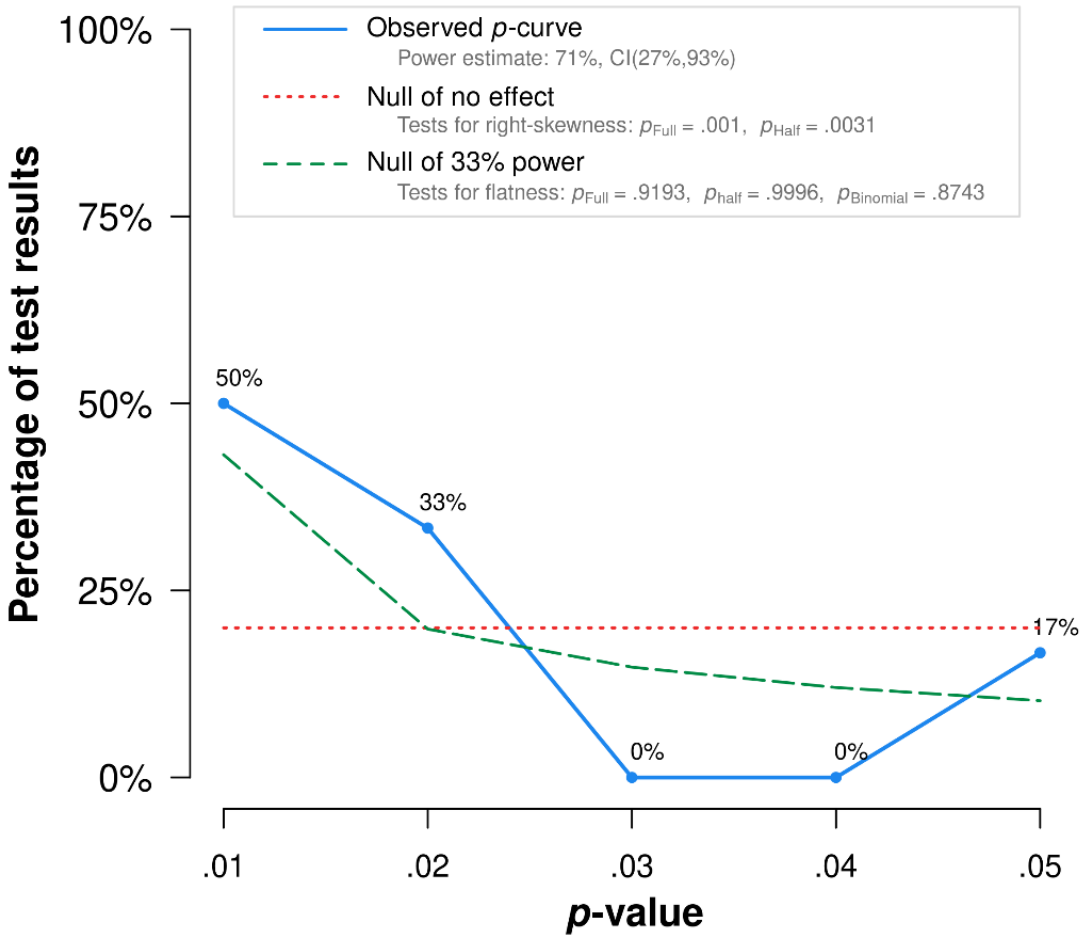

Figure 5. **Meta-analyses support cognitive impairments in excluded DP cases.** All *p*-values in support of missed cases' objective cognitive impairments from my sample and Bate et al. (2019) exhibit a right-skew distribution, with the *p*-curve analysis statistically significant. This means that missed cases' complaints are validated by objective impairments in face processing.

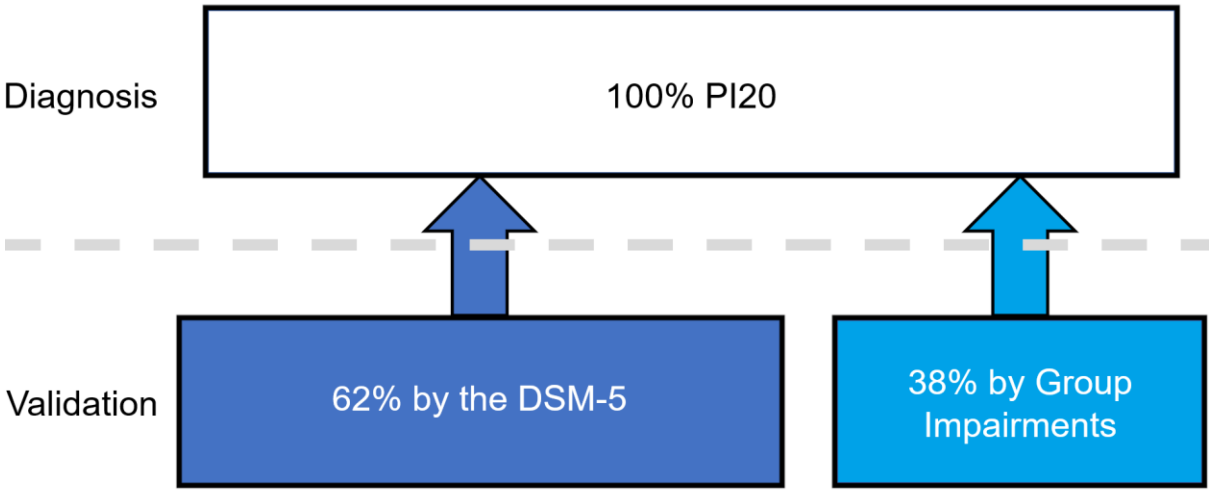

Figure 6. How the symptom-based approach to diagnosing is validated by objective impairments in DP. The PI20 diagnoses 100% of DP cases as atypical in their symptoms (top). Sixty-two percent of these self-identified DP cases were validated as objectively impaired at the -1 SD level on the FFT and CFMT, with the remaining 38% validated through deficits at the group level in their FFT, CFMT, CFPT upright and Holistic Perception measures. Thus, when a DP case reports atypical symptoms via the prosopagnosia index questionnaire, we can be confident that they reflect underlying cognitive impairments.