UNIVERSITY OF
GLOUCESTERSHIRE

This is a peer-reviewed, final published version of the following document, © 2022 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/). and is licensed under Creative Commons: Attribution 4.0 license:

Bennetts, Rachel J., Gregory, Nicola J., Tree, Jeremy, Di Bernardi Luft, Caroline, Banissy, Michael J., Murray, Ebony ORCID logoORCID: https://orcid.org/0000-0003-4928-5871, Penton, Tegan and Bate, Sarah (2022) Face specific inversion effects provide evidence for two subtypes of developmental prosopagnosia. Neuropsychologia, 174. art:108332. doi:10.1016/j.neuropsychologia.2022.108332

Official URL: http://doi.org/10.1016/j.neuropsychologia.2022.108332
DOI: http://dx.doi.org/10.1016/j.neuropsychologia.2022.108332
EPrint URI: https://eprints.glos.ac.uk/id/eprint/11362

Disclaimer

The University of Gloucestershire has obtained warranties from all depositors as to their title in the material deposited and as to their right to deposit such material.

The University of Gloucestershire makes no representation or warranties of commercial utility, title, or fitness for a particular purpose or any other warranty, express or implied in respect of any material deposited.

The University of Gloucestershire makes no representation that the use of the materials will not infringe any patent, copyright, trademark or other property or proprietary rights.

The University of Gloucestershire accepts no liability for any infringement of intellectual property rights in any material deposited but will remove such material from public view pending investigation in the event of an allegation of any such infringement.

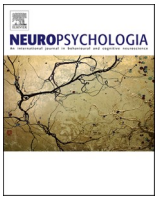
PLEASE SCROLL DOWN FOR TEXT.

Neuropsychologia 174 (2022) 108332

Contents lists available at ScienceDirect

Neuropsychologia

journal homepage: www.elsevier.com/locate/neuropsychologia



Face specific inversion effects provide evidence for two subtypes of developmental prosopagnosia

Rachel J. Bennetts [a,*], Nicola J. Gregory [b], Jeremy Tree [c], Caroline Di Bernardi Luft [d], Michael J. Banissy [e,f], Ebony Murray [g], Tegan Penton [f], Sarah Bate [b]

[a] College of Health, Medicine and Life Sciences, Brunel University, UK
[b] Department of Psychology, Bournemouth University, UK
[c] Department of Psychology, Swansea University, UK
[d] School of Biological and Behavioural Sciences, Queen Mary University of London, UK
[e] School of Psychological Science, University of Bristol, UK
[f] Department of Psychology, Goldsmiths, University of London, UK
[g] Department of Psychological Sciences, University of Gloucestershire, UK

ARTICLE INFO

Keywords:
Prosopagnosia
Face recognition
Face perception
Cluster analysis

ABSTRACT

Many studies have attempted to identify the perceptual underpinnings of developmental prosopagnosia (DP). The majority have focused on whether holistic and configural processing mechanisms are impaired in DP. However, previous work suggests that there is substantial heterogeneity in holistic and configural processing within the DP population; further, there is disagreement as to whether any deficits are face-specific or reflect a broader perceptual deficit. This study used a data-driven approach to examine whether there are systematic patterns of variability in DP that reflect different underpinning perceptual deficits. A group of individuals with DP (N = 37) completed a cognitive battery measuring holistic/configural and featural processing in faces and non-face objects. A two-stage cluster analysis on data from the Cambridge Face Perception Test identified two subgroups of DPs. Across several tasks, the first subgroup (N = 21) showed typical patterns of holistic/configural processing (measured via inversion effects); the second (N = 16) was characterised by reduced or abolished inversion effects compared to age-matched control participants (N = 91). The subgroups did not differ on tasks measuring upright face matching, object matching, non-face holistic processing, or composite effects. These findings indicate two separable pathways to face recognition impairment, one characterised by impaired configural processing and the other potentially by impaired featural processing. Comparisons to control participants provide some preliminary evidence that the deficit in featural processing may extend to some non-face stimuli. Our results demonstrate the utility of examining both the variability between and consistency across individuals with DP as a means of illuminating our understanding of face recognition in typical and atypical populations.

1. Introduction

Prosopagnosia is a condition characterised by a severe, long-lasting deficit in face recognition. In acquired prosopagnosia, face recognition deficits occur after brain injury; whereas in developmental prosopagnosia (DP; also referred to as congenital prosopagnosia), problems in face recognition occur in otherwise typically developing individuals with no history of brain injury, no co-occurring developmental disorders, and typical cognitive, intellectual, and lower level visual skills (Bate and Tree, 2017). DP occurs in roughly 2% of the population (Bennetts et al., 2017; Kennerknecht et al., 2006), and over the past 20 years there has been substantial interest in the cognitive characteristics of DP. Specifically, there have been a large number of neuropsychological studies that have attempted to determine which cognitive processes are impaired in individuals with DP. To date, most studies have either focused on identifying a single underpinning deficit common to all cases of DP (e.g., Biotti et al., 2018; DeGutis et al., 2012; Gerlach et al., 2017; Palermo et al., 2011), or have aimed to describe patterns of deficits in relatively small groups (5–11 participants; e.g., Behrmann et al., 2005; Le Grand et al., 2006; Schmalzl et al., 2008; Ulrich et al.,

* Corresponding author. Dr Rachel Bennetts, Division of Psychology College of Health Medicine and Life Sciences Brunel University, London Kingston Lane Uxbridge UB8 3PH, United Kingdom.
E-mail address: rachel.bennetts@brunel.ac.uk (R.J. Bennetts).

https://doi.org/10.1016/j.neuropsychologia.2022.108332
Received 25 August 2021; Received in revised form 6 July 2022; Accepted 8 July 2022
Available online 14 July 2022
0028-3932/© 2022 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

2017).

One of the perceptual processes that has been commonly studied in relation to face recognition is holistic processing – the tendency to integrate and process information from the entire face as a whole, rather than decomposing faces into individual parts (Maurer et al., 2002; Piepers and Robbins, 2012; Richler et al., 2012). Holistic processing is often measured by examining the effects of rotating the face 180° so that it appears upside-down (inversion). Recognition of faces is disproportionately impaired by inversion compared to recognition of other objects (Rossion, 2008; Valentine, 1988; Yin, 1969) – a phenomenon known as the face inversion effect. This appears to reflect the fact that holistic processing is reduced or abolished for inverted faces (Rossion, 2008). Specifically, some studies suggest that face inversion effects arise because it is particularly difficult to extract information about spatial relationships between facial features from inverted faces (henceforth 'configural processing'; Maurer et al., 2002; Rossion, 2008). Instead, inverted faces must rely on piecemeal processing of individual facial features, which are less affected by inversion than configural or holistic processing mechanisms (Mondloch et al., 2002; Rossion, 2008) (although see McKone and Yovel, 2009 for a discussion of inversion effects for features).

Evidence of reduced or abolished face inversion effects has been used to support the idea that holistic or configural processing deficits underpin DP (e.g., Avidan et al., 2011; Behrmann et al., 2005; DeGutis et al., 2012; Huis in 't Veld et al., 2012). For example, DeGutis et al. (2012) provided an overview of 14 studies which examined face inversion effects in DP, and concluded that "*the majority of DP studies using face inversion show no evidence of holistic processing [in DP]*" (p. 426). However, examining the DP literature more broadly reveals that there remain substantial disagreements about the consistency and specificity of holistic processing impairments in DP, and it was this fundamental issue that motivated our work.

Regarding consistency, there is often substantial variability in the amount of holistic or configural processing (as indexed by face inversion effects) displayed between different individuals with DP. Although it may be the case that group-level comparisons between DPs and typical individuals show reduced holistic processing in DP (Behrmann et al., 2005; DeGutis et al., 2012; Klargaard et al., 2018; Russell et al., 2009), there remains substantial variability on a case-by-case basis. For example in a recent study, Biotti et al. (2018) used the Cambridge Face Perception Test (CFPT; Duchaine et al., 2007a) to assess face perception in 72 individuals with DP. Of these, 17 displayed a substantially reduced face inversion effect compared to controls, but the effect ranged from −4.93 to 1.95 SDs from the control mean (see also Klargaard et al., 2018; Russell et al., 2009).

Similar patterns of variability in DP performance occur for other tasks which are commonly used to assess holistic and configural processing. In the composite task, the bottom half of one face is paired with the top half of another face, and participants are asked to identify or match one half whilst ignoring the other half. When the two face halves are aligned (i.e., form a usual face shape), performance is typically worse than when the two halves are spatially misaligned (e.g., offset to the left or right). This difference, referred to as the 'composite effect', is interpreted as a measure of holistic processing (Murphy et al., 2017; Rossion, 2013).[1] While some studies have found significant reductions in the composite effect in DP (Avidan et al., 2011; Liu and Behrmann, 2014; Palermo et al., 2011), a number of studies have found minimal or no evidence of reduction (e.g., Biotti et al., 2017; Le Grand et al., 2006; Susilo et al., 2010; Ulrich et al., 2017). Further, as with face inversion effects, group level analyses can mask large amounts of heterogeneity amongst DP participants. For example, out of 14 DPs tested in Avidan et al.'s (2011) study, only half demonstrated a composite effect more than 1 SD away from the control mean (see also Schmalzl et al., 2008). There are fewer studies that have explicitly examined sensitivity to configural information in DP, but those that do also find a heterogeneous pattern of results. For example, Le Grand et al. (2006) and Schmalzl et al. (2008) each found that around half of their DP sample showed significantly impaired spacing judgements (see also Duchaine et al., 2007b; Ulrich et al., 2017).

This heterogeneity is further complicated by the fact that holistic processing deficits might depend somewhat on the task used to assess them. For example, Klargaard et al. (2018) found that, on a group level, inversion effects in individuals with DP were reduced for a memory task, but not for a perception task. Further, one of the few studies that has assessed both inversion and composite effects in DP (Avidan et al., 2011) demonstrated little overlap between participants with significant holistic processing deficits using the inversion and composite measures (see also Le Grand et al., 2006; Schmalzl et al., 2008; Ulrich et al., 2017).

The discrepancy between tasks may reflect the fact that the term 'holistic processing' has been interpreted in a number of different ways in the literature (see Piepers and Robbins, 2012; Richler et al., 2012; for an overview) and measures such as inversion and the composite effect may not reflect the same underlying mechanisms. For example, Richler et al. (2012) suggest that face inversion effects likely reflect sensitivity to facial configurations, whereas composite effects likely measure a failure of selective attention. The idea that different measures reflect distinct perceptual mechanisms is supported by the finding that face inversion and composite effects do not correlate with one another in the typical population (Rezlescu et al., 2017). One of the clear implications from these results is that measures of holistic processing – and the conclusions that can be drawn from them – can vary substantially across tasks, and researchers should not assume that evidence of a holistic processing deficit on one task necessarily implies a similar deficit across all tasks. However, to date there have been no larger-scale studies which might allow researchers to draw conclusions about systematic relationships (or lack thereof) between the different measures in DP.

The potential heterogeneity between different measures of holistic processing is further complicated by the fact that individuals with DP may also vary in other skills that contribute to face processing. For example, several studies have attempted to examine featural and configural processing separately, using sets of faces that systematically vary in either their features, or the spacing of their features (known as the 'Jane' task; Mondloch et al., 2002; or the 'Alfred task'; Yovel and Duchaine, 2006). Yovel and Duchaine (2006) and Duchaine et al., 2007b found that DPs as a group showed equal deficits for both spacing (i.e., configural) and feature discriminations, although, as with holistic processing, case series reports suggest that featural processing can vary between individuals (Le Grand et al., 2006; Schmalzl et al., 2008). This is in line with many of the studies of the face inversion effect, which find that DPs often show poorer performance than controls for inverted faces (Behrmann et al., 2005; Biotti et al., 2018; Garrido et al., 2008; Klargaard et al., 2018), which are thought to be processed in a more feature-based manner (Rossion, 2008). These lines of evidence suggest that at least some individuals with DP may show deficits in facial feature processing, independently or alongside deficits in holistic processing. As such, performance on baseline measures or tasks which are thought to isolate "featural" processing (e.g., inverted faces; the "Jane" task, Mondloch et al., 2002) should be taken into account when characterising patterns of impairment in DP.

Finally, there has been much debate over whether any impairment in holistic processing in DP is face-specific, or whether it reflects impairments in more general visual processing mechanisms (Avidan et al., 2011; Duchaine et al., 2007b; Gerlach et al., 2017, 2022). Typically, this question has been examined using the Navon task (Navon, 1977), in which participants are presented with a compound letter stimulus – a large letter constructed from a number of small letters – and asked to

[1] In the traditional form of the task, the distractor (to-be-ignored) half is always different, although other forms of the composite task have been proposed which use more complex designs (see Murphy et al., 2017 for a description; Piepers and Robbins, 2012; Rossion, 2013 for critiques).

respond to either the larger or smaller letters while disregarding the other. In general, people are faster to respond to the larger letter (a global precedence bias) (Avidan et al., 2011; Behrmann et al., 2005; Duchaine et al., 2007b).

On a group level, some studies have found that this global bias is reduced in individuals with DP (Avidan et al., 2011; Behrmann et al., 2005; Bentin et al., 2007; Gerlach et al., 2017), and suggest that this reflects a general deficit in processing global shape information (Avidan et al., 2011). However, other studies have found no difference between DPs and controls on the Navon task (Duchaine et al., 2007a; Duchaine et al., 2007b; Ulrich et al., 2017), and once again, there appears to be substantial variability at an individual level (e.g., Schmalzl et al., 2008). Further, all of the studies that have used the Navon paradigm have analysed relatively small groups of participants (5–14 DPs), so it is possible that the heterogeneity within the sample simply makes it difficult to identify reliable associations between tasks; alternatively, it may be that only a subset of DPs shows general difficulties with holistic processing.

In sum, holistic processing deficits (and, in some case, feature processing deficits) in DP vary substantially, both within and between tasks. The origins and nature of this variability are unclear: there remains some debate as to whether different measures of holistic processing tap into similar perceptual mechanisms, and whether any deficits are (1) specific to holistic processing and (2) specific to faces. Although many studies in this area have sought to identify a single underlying impairment common to individuals with DP, evidence of heterogeneity in DP supports the hypothesis that there may in fact be multiple, separable impairments present in different cases of DP. Studies in DP that have used an in-depth case series approach have often explicitly explored this possibility (Dalrymple et al., 2014; Le Grand et al., 2006; Schmalzl et al., 2008; Ulrich et al., 2017). However, identifying a deficit at the level of individual cases requires a relatively severe impairment (generally at least 1.7–2 SDs from the control mean); consequently, case series analyses are insensitive to more subtle patterns of deficits. On the other hand, group-based analyses are capable of identifying smaller effects, but they obscure important patterns of variability within the data. These patterns of variability are not trivial, as they have the capacity to inform theories of face recognition in typical individuals. For example, identifying subtypes related to holistic and featural face processing would indicate that both contribute to typical face recognition, contributing to the debate around the nature of holistic face processing (Piepers and Robbins, 2012). Likewise, identifying subtypes that show differing levels of generalisability to object recognition could clarify the presence of shared and separate visual processing pathways for faces and objects (Gerlach et al., 2022), and clarify some of the long-standing debates about the specificity of DP (see Geskin and Behrmann, 2018 for a review).

Currently, the presence of multiple subtypes of individuals with DP remains relatively unexplored outside case-series analyses. Consequently, the main aim of the current study was to employ a commonly-used measure of face perception, the Cambridge Face Perception Test (CFPT; Duchaine et al., 2007a), to examine whether there are different "subtypes" of DP which show separable patterns of impairment on holistic and part-based processing. We selected the CFPT as it includes an inverted condition (allowing simple assessment of inversion effects), and it is widely used by researchers to assess face perception in individuals with DP (Bate et al., 2019b; Bate et al., 2019c; Biotti et al., 2018; Corrow et al., 2016; Dalrymple and Palermo, 2016; Gerlach et al., 2022; Klargaard et al., 2018) – as such, the findings of the current study are applicable to (and can be validated in) samples from independent research labs worldwide.

In order to identify whether different subtypes were present in the data, we used cluster analysis: an analytical approach which can identify groups displaying different patterns of performance across multiple measures. Cluster analysis has been used to identify and characterise potential subtypes in a variety of developmental and psychiatric disorders (Barton et al., 2004; Lewandowski et al., 2014; Pacheco et al., 2014; Prior et al., 1998) – for example, Barton et al. (2004) examined a group of individuals with social developmental disorders on their performance on face processing tasks, and identified four clusters which showed differential impairment on tests of face perception and face imagery. To date, this approach has not been employed in cases of DP.

Once clusters had been identified, we examined the pattern of performance of each cluster on a broad battery of measures designed to assess holistic, configural, and featural processing, in both faces and non-face objects. We aimed to investigate (a) whether the patterns of performance we observed in the CFPT clusters would replicate in other conceptually similar tasks (i.e. CFMT and Matching Task); and (b) whether the clusters also differed in their performance on other tasks and measures (i.e. alternative measures of holistic face processing, object processing, and non-face holistic and part-based processing). Each cluster was first compared to other clusters; subsequently, supplementary analyses compared each cluster to a large group of age-matched control participants. These analyses allowed us to characterise the features of each cluster, and determine whether specific differences and deficits in holistic, configural, and featural processing were present. Further, by employing a broad range of tasks, we were also able to examine the task- and face-specificity of the holistic and featural processing deficits present in each cluster. Consequently, this study has the potential to offer new insights into the different perceptual pathways that can lead to face recognition difficulties.

2. Methods

2.1. Participants

Thirty-seven adults with DP (19 female, 18 male; age range = 18–75 years, M = 48.28, SD = 17.18) took part in this study. An additional 10 adults with DP took part in the research but were excluded from the analysis as they did not provide complete data. All participants with DP contacted our laboratory complaining of severe difficulties with face recognition. Subsequently, participants were invited to our lab and completed a battery of tests designed to assess their face recognition abilities, general visual processing, and general cognitive skills. All individuals with DP met the criteria for DP as adhered to by most researchers in the field (see Dalrymple and Palermo, 2016; Murray et al., 2018). Specifically, all participants with DP performed significantly (>2SDs) below published age-matched control cut-offs on the Cambridge Face Memory Test (CFMT; Duchaine and Nakayama, 2006) and a famous faces test (Bate et al., 2019b).

No individual reported a history of socio-emotional, psychiatric or neurological disorders. Participants also completed the Autism Quotient (Baron-Cohen et al., 2001). While the AQ is not a formal diagnostic instrument for Autism Spectrum Conditions, very few age-matched controls score in the extremely high range (>34, as defined by Baron--Cohen et al., 2001). Although face recognition difficulties are common in Autism Spectrum Conditions (Weigelt et al., 2012), it is possible that they are qualitatively distinct from those in DP; therefore, we excluded any participants scoring above 34 from the current analysis (see Dalrymple and Palermo, 2016 for further discussion). General cognitive abilities were estimated using the Wechsler Test of Adult Reading (WTAR; Holdnack, 2001), and participants with an estimated IQ lower than 70 (3 SDs below the mean in the typical population) were excluded. For participants over the age of 65 years, we screened for cognitive decline using the Mini Mental State Examination (Folstein et al., 1975), and excluded any individuals scoring below 26/30 (as per Larner, 2012). To rule out lower-level visual impairments, participants completed assessments of basic visual acuity using a standard Snellen letter chart (3 m), the Hamilton-Veale contrast sensitivity test, and four sub-tests of the Birmingham Object Recognition Battery (BORB; Humphreys and Riddoch, 1993): Line Match, Size Match, Orientation Match, and Position of the Gap Match. To assess basic category recognition, participants

completed the Object Decision Test (hard version) of the BORB. None of the participants with DP performed poorly in either the WTAR or the MMSE, none showed pervasive difficulties with lower-level vision or object categorisation, and none scored in the extreme range on the AQ.

A total of 91 control participants (47 female, age range = 20–75 years, M = 41.97, SD = 16.34) also took part in this study (an additional 17 control participants were recruited but did not complete at least 3 tests in the battery, so their data was excluded from analysis). The control group was IQ-matched to the DP sample (using the WTAR, Holdnack, 2001), and did not report any history of socio-emotional, neurodevelopmental, psychiatric or neurological disorders. No individual reported everyday difficulties in face recognition, and all controls performed within the typical range for the CFMT and famous faces task.[2] Controls were recruited from the departmental participant pool and received a small financial payment in exchange for their time.

Demographic and screening task data for control and DP participants is available at https://osf.io/3gzyr/?view_only=42eee9c2277f4e1a9295cb738da1a1bf.

Ethical approval for this experiment was granted by the institutional Ethics Committee, and all participants provided informed consent according to the Declaration of Helsinki.

2.2. Materials

2.2.1. CFPT

The CFPT (Duchaine et al., 2007a) is a standardised test of face perception. In each trial, participants simultaneously view a single target face and six comparison faces. Each of the comparison faces has been morphed to resemble the target face to a different degree. Participants are given up to 60 s to sort the comparison faces in order of similarity to the target face. Each face measured approximately 3.5–4.0 cm in width on the screen. For an example of the trial layout for the CFPT, see Duchaine et al., 2007a.

The CFPT contains 16 trials (eight each upright and inverted, presented in a fixed pseudo-random order), and traditionally performance is scored by summing deviations from the correct order (e.g., if a face is three spaces from its correct location, it would add three to the deviation score), so that a higher score equates to worse performance. To aid in the analysis, the present study converted scores for the CFPT into percentage correct using formula [100 × (1-(deviation score/maximum score))] as per Rezlescu et al. (2012). Separate scores were calculated for upright and inverted trials.

In addition to the percentage correct measure, we also calculated an inversion index, normalised for baseline levels of performance, using the formula [(upright − inverted)/(upright + inverted)] (Avidan et al., 2011; Ulrich et al., 2017). A positive inversion index indicates an inversion effect (i.e., better performance for upright than inverted faces).

Within our sample of individuals with DP, reliability for the CFPT (estimated using Guttman's λ2, as in DeGutis et al., 2013; Rezlescu et al., 2017) was λ = 0.613 for upright trials and λ = 0.551 for inverted trials. Reliability for the inversion effect (calculated as per DeGutis et al., 2013) was λ = 0.525.

2.2.2. Face and object matching

This task used a sequential same/different matching procedure, involving matching of upright and inverted faces, houses, and hands. An extended version of the procedure has been described in detail elsewhere (Bate et al., 2019c; Bobak et al., 2016). In brief, for the face condition, in each trial participants viewed a single image of a face for 250 msec, followed by a 1000 msec ISI, then a second image of a face (presented until a response was recorded). All stimuli measured approximately 8 cm in width on the screen. Examples of the stimuli used in this task can be viewed in Bobak et al. (2016) and Bate et al., 2019c.

Participants used the "z" and "m" keys on a keyboard to indicate whether the two images showed the same identity or two different identities (assignment of response keys was counterbalanced between participants). Participants were instructed to respond as quickly and as accurately as possible. Trials for houses and hands followed exactly the same procedure.

For each stimulus, there were 32 pairs of images (16 same identities; 16 different identities), and images within each pair differed slightly to minimise image-matching strategies. For faces and houses, the first image showed a frontal viewpoint; the second image showed the face/house from around 30°; for hands, the images showed different hand positions (e.g., fingers together and fingers splayed). All image pairs were presented twice upright and twice inverted. Presentation of object categories was blocked, and the order randomised between participants. Within each block, the order of trials was randomised. Each block was preceded by six practice trials (containing different stimuli to the main experiment).

Accuracy was calculated separately for same-identity and different-identity trials in each condition, and combined into a single bias-free measure of sensitivity *d'* (Macmillan and Creelman, 2005). Reaction times were also analysed. RTs for trials with incorrect responses; any RTs less than 150 msec, and any RTs more than 3SDs from the participant's mean RT were excluded from analysis.

Given there has been much debate about the specificity of DP (see e.g., Bate et al., 2019c; Geskin and Behrmann, 2018), we were interested in examining whether either of the subtypes identified in the cluster analysis was associated with more general abnormalities with orientation-specific processing. Consequently, the follow-up analyses included data from the face and house matching condition (hands were excluded as they do not show a canonical orientation, rendering inversion effects difficult to interpret; for a complete analysis of all conditions of this task see Bate et al., 2019c).

2.2.3. Composite task

This study used a modified version of the composite task employed by Robbins and McKone (2007) (see also Bobak et al., 2016). The task uses a sequential same-different matching procedure to examine the composite effect for faces and Labrador dogs (full-body photographs displaying a side view). In each trial, participants viewed a single composite image (created from the top half of one identity and the bottom half of a second identity) for 600 msec, followed by an ISI of 300 msec, then a second composite image (presented until participants made a response). Subsequently, participants were required to press the space bar to move onto the next trial. Stimuli were offset by 25% of screen size, to prevent matching based on the size or location of the features. The top half of the two images could show either the same face/dog, or a different face/dog. The bottom halves of the two images were always drawn from two different identities, as is typical in the standard or original composite design (Robbins and McKone, 2007; Rossion, 2013; also referred to as the "partial design"; e.g., Richler and Gauthier, 2013).[3] Aligned face images measured approximately 4 cm in width on screen; aligned dog images measured approximately 6 cm in width on

[2] A minority of controls (9/91) did not complete the famous faces task; however, all performed in the typical range for the CFMT and CFPT, and were therefore retained in the sample.

[3] Although there has been much debate over the use of the standard design in comparison to a longer version, referred to as the "congruency/interference" or "complete" design (see e.g., Richler and Gauthier, 2013; Rossion, 2013), this study utilised the standard design with these stimuli primarily because there is minimal evidence for a composite effect in the control conditions (inverted faces and dogs) using the standard design; whereas studies using the congruency design often find composite effects for inverted faces and non-face objects. For a more comprehensive comparison of the two designs, including the perceptual and neural locus of effect in each case, see Rossion (2013).

the screen. Examples of all categories of stimuli can be viewed in Robbins and McKone (2007).

Participants used the "z" and "m" keys on a keyboard to indicate whether the top half of the two images (i.e., the section containing the eyes) showed the same identity (same face or dog) or two different identities (assignment of response keys was counterbalanced between participants). Participants were instructed to respond as quickly and as accurately as possible.

Precise details of stimulus creation can be found in Robbins and McKone (2007), but in brief, the top and bottom halves of the stimuli were either aligned so that the edges joined up relatively neatly (aligned condition); or the bottom half of the stimulus was offset by approximately one quarter the width of the stimulus (misaligned condition). The original set of stimuli contained 30 same-identity pairings and 180 different-identity pairings each for dogs and faces. For the current study, we randomly extracted 30 same-identity pairs and 30 different-identity pairs (15 pairs each for faces and dogs) from the original stimulus set, and presented the same pairings in each of the conditions (upright, inverted; aligned, misaligned). Trials were blocked by object and orientation. Within each block, the order of trials (same/different; aligned/misaligned) was randomised. Each block was preceded by six practice trials (drawn from a different stimulus set than the rest of the trials).

Accuracy for the same-identity trials only was included in the follow-up analyses (McKone and Robbins, 2007; Rossion, 2013). As participants had an unlimited time to respond to the second stimulus, there was a potential for a trade-off between response speed and accuracy. Consequently, reaction times for correct responses to same-identity trials were also included in the follow-up analyses.

2.2.4. Navon task

In this task, participants are presented with a composite letter stimulus (many small letters arranged in the shape of a larger letter) and asked to identify either the larger letter or the smaller letter (Navon, 1977). In the current study, the stimuli consisted of the letters H and S (see Bobak et al., 2016; Duchaine et al., 2007b), which were constructed so the large and small letters were the same (congruent) or different (incongruent). Each trial began with a 600 msec fixation cross; subsequently, the stimuli were presented onscreen in one of four different positions, to prevent participants focusing on a single spot on the screen. Large letters measured approximately 4 cm in width; while small latters measured approximately 0.4 cm in width. Examples of Navon stimuli used in this study can be found in Duchaine et al., 2007b and Berhmann et al., 2005.

Participants responded by pressing the "s" or "h" keys on their keyboard. Participants were instructed to respond as quickly and as accurately as possible, and the stimuli remained onscreen until participants responded.

In this study, the task was divided into four sections, each with 48 trials. In two of the sections, participants were required to identify the larger letter (global trials); in the other two, participants were required to identify the smaller letter (local trials). The sections were presented in a fixed semi-random order, and the order of trials within each section was randomised.

Given that participants were very accurate in performing this task, only RT was considered in the follow-up analysis. In order to examine the level of holistic processing bias for non-face stimuli, a global bias index was calculated by dividing the average RT for correct responses to global trials by the average RT for correct responses to local trials using the formula [((Global congruent RT + Global incongruent RT)/2)/(Local congruent RT + Local incongruent RT)/2] (Bobak et al., 2016; Duchaine et al., 2007a). A score below 1 indicates that participants were faster identifying the larger letters (a global or more holistic bias); a score above 1 indicates that participants were faster at identifying the smaller letters (a local or more piecemeal bias). As this index incorporates multiple mechanisms which can contribute to global shape processing (Gerlach et al., 2017), we also examined global precedence (global congruent – local congruent), local-to-global interference (global congruent – global incongruent), and global-to-local interference (local congruent – local incongruent) (Duchaine et al., 2007b).

2.2.5. Jane task

This study used a modified version of the Jane task (Mondloch et al., 2002). The task uses a sequential same-different matching task to examine participants' sensitivity to changes in facial features, facial feature configurations, or facial contours. In each trial, participants viewed a single image of a Caucasian female face for 200 msec, followed by a 300 msec ISI, then a second image of the face (presented until participants made a response). Stimuli were offset by 25% of the screen width, to prevent matching based on the size or location of the features. The first and second image in each trial were either identical to the first image ("same identity" trials) or varied subtly ("different identity" trials). The faces could vary in three different ways: the eyes and mouth were replaced with different features (feature change trials); the spacing of the eyes and mouth were changed (i.e., moving the eyes up/down/-further apart/closer together and the mouth up/down; spacing trials), or the shape of the face outline was altered (contour trials) (for precise details of stimulus variations and examples of the stimuli, see Mondloch et al., 2002). Faces measured approximately 6 cm in width on screen.

Participants were asked to judge whether the images were the same or different. Participants used the "s" and "n" keys on a keyboard to indicate whether the images showed the same image or two different images, respectively. Participants were instructed to respond as quickly and as accurately as possible.

Each condition (feature, spacing, and contour trials) contained 30 pairs of images (15 same; 15 different identity trials). All image pairs were presented both upright and inverted. Upright and inverted trials were presented in separate blocks, with upright trials always presented first. Within each block, the order of trials was randomised. Each block was preceded by six practice trials.

Accuracy was calculated separately for same-identity and different-identity trials in each condition, and combined into a single bias-free measure of sensitivity *d'* (Macmillan and Creelman, 2005). Reaction times were also analysed. RTs for trials with incorrect responses and any RTs less than 150 msec or more than 4500msec were excluded from analysis.

As the primary focus of the follow-up analysis was holistic and facial feature processing (and there was no a priori empirical or theoretical reason to predict differences based on sensitivity to facial contours), only the results from the upright and inverted feature and spacing change conditions were included in the follow-up analyses.

2.3. General procedure

All participants were tested individually in a lab-based setting. Each test was presented on a laptop computer, with a screen measuring 15 inches. Participants completed the tests from a viewing distance of approximately 60 cm. The exact order of the tests was varied between participants, although the screening tests were always presented at the beginning of the session. The full battery of tests (including basic DP screening measures and the experimental measures) was completed over two separate sessions.

Due to time constraints and some computer errors, not all participants completed all tests – 17 control participants were missing data from a single test in the battery.

2.4. Analysis

All three measures from the CFPT (percentage correct for upright and inverted faces, and the inversion index) were transformed into z-scores to be entered into the cluster analysis (z-scores were based on DP data only: i.e., control scores were not taken into account at this stage).

For the cluster analysis itself, we first applied agglomerative hierarchical cluster analysis to explore how many clusters were present in the DP data. We then applied k-means in order to separate the clusters and analyse their characteristics. Only data from DP participants was entered into the cluster analysis. The use of a two-stage cluster analysis, with a hierarchical method followed by a k-means method, has been recommended as the most appropriate strategy for the assessment and analysis of clusters within a dataset (Lange et al., 2002). Furthermore, the use of a two-stage strategy allows researchers to establish the internal validity of the cluster solution (Lange et al., 2002). Cluster analyses were carried out in MATLAB version R2018b.

To determine the characteristics of the clusters that were identified, we carried out ANCOVAs comparing performance between the members of different clusters (with cluster membership defined in the k-means analysis) on each of the tests in the battery, controlling for participant age. All ANCOVA analyses were carried out in JASP 0.10.2.0 (JASP Team, 2019). The data and code used in the cluster analyses, along with participants' performance on each task in the cognitive battery, can be accessed at https://osf.io/3gzyr/?view_only=42eee9c2277f4e1a9295cb738da1a1bf.

3. Results

First, we present the results of the agglomerative hierarchical cluster analysis used to determine the number of clusters present in the data. Second, we present a k-means cluster analysis based on the number of clusters observed. The k-means cluster analysis enabled us to characterise the observed clusters (Lange et al., 2002). Third, we examine the differences in performance between the clusters on a variety of face and non-face tasks and measures of holistic processing.

3.1. Cluster analysis

3.1.1. Agglomerative hierarchical cluster analysis

We adopted a single-link hierarchical cluster technique (Jain et al., 1999). This analysis followed three main steps. First, we calculated the distance matrix based on the correlations (proximity measure) between elements and tested their linkage based on those distances. The linkages were presented as a hierarchical cluster tree (dendogram). Second, we calculated the cophenetic correlation coefficient in order to verify if the clustering based on the distance vector was valid as we expect that the linkage of the elements in the cluster tree should be strongly correlated with the actual distances between them. This validates our distance measure. Third, we determined the number of distinguishable clusters in the data by using the inconsistency coefficient which compares the separation between different links within the cluster. The inconsistency coefficient compares the height of a link in a cluster hierarchy with the average height of links below it. A link with a similar height than the links below it shows that there are no distinct divisions between the elements joined at its level of hierarchy. The linkage inconsistency was calculated for a given link minus the average height of all other links. The depth of links' average was determined by the number of links observed in the dendogram (depth = 5). The threshold for the clustering was determined as:

$$ClustThreshold = Hmax(IC) - 0.05 \quad \text{Eq.1}$$

where *Hmax* is the maximum height of the linkage with the highest inconsistency coefficient (*IC*). The value 0.05 was subtracted in order to make sure the threshold is just slightly lower than the maximum height.

Our hierarchical cluster results can be visualised on the dendogram on Fig. 1. The dendogram shows the presence of two distinguishable clusters highlighted in red and blue colours as well as the threshold as calculated via Eq. (1). The presence of two clusters is clear given that the heights of the linkages just below the threshold are substantially higher than the heights of the links below them. The *cophenet* correlation coefficient observed was 0.957 which demonstrates the consistency between our distance measures and the clustering observed (i.e. linkages). After thresholding on the estimated value (threshold = 1.72), the cluster solution gives us two clusters, one with 21 participants and another with 16. These groupings were validated using the k-means algorithm with 2 groups as described below.

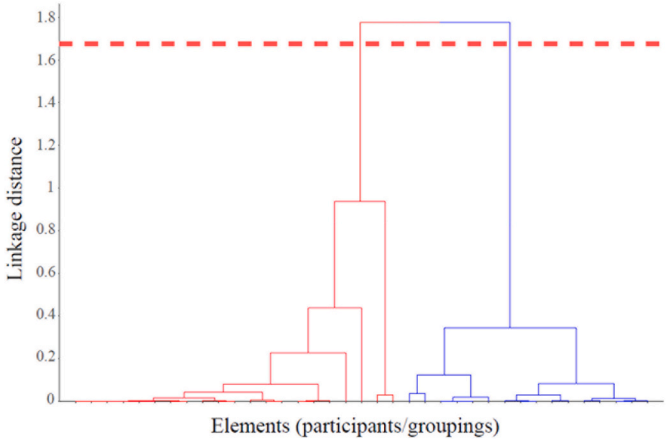


Fig. 1. Dendogram demonstrating the presence of two clusters in the data. A smaller cluster in blue (n = 16) and a larger cluster (n = 21) represented in red. The cluster threshold is represented as the dashed red line and it was determined as described in Eq. (1). (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

3.1.2. K-means cluster analysis

In order to both validate and further explore the two clusters we observed in the hierarchical cluster analysis, we carried out a cluster analysis using k-means partitioning the data into 2 mutually-exclusive clusters using correlation as a measure of distance. We observed two clusters of the same size as in the hierarchical analysis, one with 21 participants and another with 16. Importantly, we observed a total agreement between the two methods: the same participants were allocated to the same clusters in both analyses (hierarchical and k-means).[4]

The cluster characteristics in relation to the input variables can be observed in Fig. 2. The presence of two groups is clear in the figure, which shows a larger group (in red, henceforth DP1) with participants who presented a higher inversion index ($M = 0.67$, $SE = 0.16$) and lower performance on inverted faces ($M = 0.74$, $SE = 0.11$); and a second group (in blue, henceforth DP2) who showed a lower inversion index ($M = -0.70$, $SE = 0.17$) and better performance on inverted faces ($M = 0.99$, $SE = 0.17$). The groups did not appear to differ in their upright face perception (DP1: $M = 0.14$, $SE = 0.21$; DP2: $M = 0.05$, $SE = 0.23$) (all scores reported as z-scores, based on DP data only).

Comparisons between the control group and the two clusters support the idea that the larger, red cluster (DP1) presents with a "typical" face inversion effect – in other words, the inversion effect for the DP1 cluster does not differ significantly from the inversion effect for controls, $t(104) = 1.32$, $p = .187$, Hedges' $g = 0.32$. However, the DP1 cluster performed significantly worse than controls in the inverted version of the CFPT, $t(71.88) = -9.75$, $p < .001$, Hedges' $g = 1.54$. On the other hand, the smaller, blue cluster (DP2) showed a reduced inversion effect compared to control participants, $t(99) = -5.32$, $p < .001$, Hedges' $g = 1.45$, but typical levels of performance in the inverted version of the CFPT, $t(99)$

[4] Typically, following the two-stage clustering process, the internal validity of the cluster solution would be assessed via a multi-profile multi-method correlation matrix (in essence, correlating the profiles of the clusters derived from each method) (Lange et al., 2002). However, as both solutions identified exactly the same cases, suggesting perfect internal validity across the two stages, this step was omitted in the current analysis.

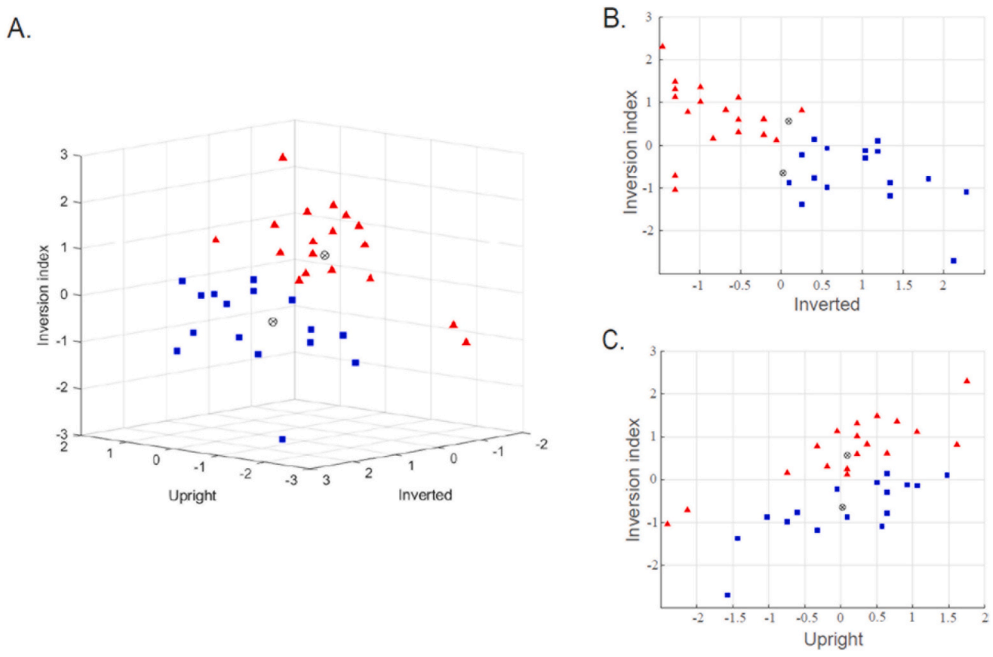


Fig. 2. Performance of the two clusters on the CFPT. The red triangles represent participants in cluster 1 (DP1; n = 21) whereas the blue squares represent participants in cluster 2 (DP2; n = 16), and black circles show the centroids of each cluster. A. 3D scatterplot showing the performance of the clusters in the CFPT three measures: upright, inverted and inversion index; B. 2D scatterplot showing the performance of the two clusters in the inverted faces and the inversion index; C. 2D scatterplot showing the performance of the two clusters in the upright faces and the inversion index. All axes are z-scores based on data from DP cases only. . (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

= 0.48, p = .631, Hedges' g = 0.15 (see Table 1).[5]

3.1.3. Control cluster analyses

To address the possibility that the outcome of the main cluster analysis reflected task-specific noise, we carried out a number of additional cluster analyses which incorporated measures from different tasks.

3.1.3.1. Clustering with multiple measures. In order to test whether the clustering was consistent when the data for each participant represented their performance on multiple measures, we averaged the z-scores for upright, inverted and inversion index across three different tasks (CFPT, CFMT, Matching Task) and submitted them to the same cluster analyses procedures described above (hierarchical and k-means).

Table 1
Demographic characteristics and mean (SD) scores on diagnostic test battery and CFPT for control participants and two DP clusters.

Measure	Controls	DP1	DP2
N	91 (47 F)	21 (12 F)	16 (7 F)
Age	41.97 (16.34)	49.86 (17.50)	46.25 (17.08)
CFMT	57.87 (8.04)	37.29 (3.64)	32.94 (6.36)
Famous faces (%)	92.87 (5.75)	50.24 (16.79)	50.96 (13.07)
CFPT (%)			
Upright	73.38 (10.08)	58.80 (9.91)	57.94 (9.16)
Inverted	53.97 (9.92)	39.75 (4.51)	55.21 (6.03)
Inversion index	0.16 (0.10)	0.19 (0.09)	0.02 (0.09)
AQ	15.46 (7.03)	17.38 (9.48)	19.93 (7.70)
Estimated IQ	113.13 (8.85)	116.52 (6.55)	117.53 (4.55)
BORB			
Length decision	26.49 (1.56)	26.19 (1.91)	26.80 (1.42)
Size decision	27.63 (1.57)	26.33 (1.77)	26.87 (1.81)
Orientation judgement	26.14 (2.25)	26.10 (1.58)	25.60 (2.44)
Location of the gap	36.79 (2.16)	34.67 (3.60)	35.27 (4.33)
Object decision	55.20 (3.78)	54.24 (3.40)	55.20 (3.78)

For the hierarchical cluster analysis, we observed a robust cophenet correlation coefficient of 0.921 and the presence of two clusters (threshold of 1.66) as demonstrated in the dendogram (Fig. 3A). The two clusters were of similar size: the larger one containing 23 participants and the smaller one with 14. We tested the agreement between the two analyses (CFPT-based cluster vs. Multi-measure cluster) and the agreement was 81.08% (only 7 participants changed cluster – all of whom switched from the DP1 cluster to the DP2 cluster).

The k-means cluster analysis, based on 2 clusters, revealed the exact same clusters shown by the hierarchical cluster analysis (separated via threshold on the dendogram). The pattern of the clusters (Fig. 3B) was the same as in the analysis of the CFPT data: the DP2 cluster (n = 23) showing a smaller inversion index (M = −0.42, SE = 0.16) compared to the DP1 cluster (n = 14), which demonstrated a larger inversion index (M = 0.90, SE = 0.0.19). As the previous cluster revealed, the pattern seems to show that these clusters do not differ in terms of performance in the upright faces in any of the tasks, including the CFPT, $t(35) = 0.693$, p = .493, Hedges' g = 0.23, the CFMT, $t(35) = 0.900$, p = .374, Hedges' g = 0.30, and the matching task, $t(35) = 1.248$, p = .220, Hedges' g = 0.41. However, the DP2 cluster showed a significantly smaller inversion index than the DP1 cluster in all tasks, including the CFPT, $t(35) = 5.110$, $p < .001$, Hedges' g = 1.70, the CFMT, $t(35) = 3.505$, $p < .001$, Hedges' g = 1.16, and the matching task, $t(35) = 3.778$, $p < .001$, Hedges' g = 1.25.[6] Further, the DP1 cluster performed significantly worse on the inverted version of all tests than the DP2 cluster; CFPT inverted: $t(35) = 5.66$, $p < .001$, Hedges' g = 1.88, CFMT inverted: $t(35) = 4.13$, $p < .001$, Hedges' g = 1.37, matching inverted: $t(35) = 3.23$, p = .003, Hedges' g = 1.07.

3.1.3.2. Clustering with different measures. Next, we repeated the k-means process from the main cluster analysis; however, we included measures from different tasks (the CFPT, CFPT, and Matching task) within each analysis. For example, while the main analysis used measures of upright face processing, inverted face processing, and the inversion index from the CFPT only (cluster analysis 1 in Table 2), in this

[5] Identical t-tests were carried out on the inversion effect calculated by simple subtraction (upright – inverted). These revealed the same pattern of results for all comparisons.

[6] Identical t-tests were carried out on the inversion effect calculated by simple subtraction (upright – inverted). These revealed the same pattern of results for all comparisons.

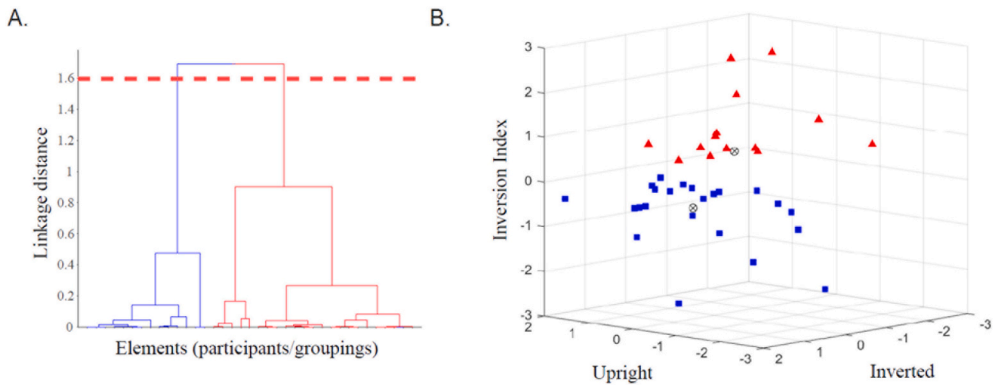


Fig. 3. Cluster Analysis using the average of CFPT, CFMT and Matching Task. A. Dendogram for multi-measure cluster analysis based on hierarchical cluster analysis; B. 3D scatterplots showing the K-means clusters patterns.

Table 2
Scores on the CFPT upright, CFPT inverted, and the CFPT inversion index for DP clusters derived using different input measures.

Cluster analysis	CFPT upright			CFPT inverted			CFPT inversion index			Agreement[a]
	DP1	DP2	*t*-test	DP1	DP2	*t*-test	DP1	DP2	*t*-test	
1: CFPT upright; CFPT inverted; CFPT inversion index	58.80	57.94	$t(35) = 0.27$, $p = .79$	39.75	55.21	$t(35) = -8.93$, $p < .001$	0.19	0.02	$t(35) = 5.69$, $p < .001$	–
2: CFPT upright; CFPT inverted; matching inversion index	57.34	59.85	$t(35) = -0.79$, $p = .43$	40.28	54.51	$t(35) = -7.09$, $p < .001$	0.17	0.04	$t(35) = 3.67$, $p < .001$	89.2%
3: Matching upright; matching inverted; CFPT inversion index	59.86	56.74	$t(35) = 1.00$, $p = .32$	41.25	52.53	$t(35) = -4.58$, $p < .001$	0.18	0.04	$t(35) = 4.58$, $p < .001$	86.5%
4: CFPT upright; CFMT inverted; matching inversion index	59.57	57.35	$t(35) = 0.71$, $p = .48$	42.67	50.00	$t(35) = -2.57$, $p = .01$	0.17	0.07	$t(35) = 2.82$, $p < .001$	64.9%
5: Matching upright; CFPT inverted; CFMT inversion index	58.04	58.83	$t(35) = -0.25$, $p = .80$	38.96	54.32	$t(35) = -9.00$, $p < .001$	0.19	0.04	$t(35) = 4.98$, $p < .001$	94.6%
6: Matching upright; CFMT inverted; CFPT inversion index	59.80	56.98	$t(35) = 0.90$, $p = .37$	41.59	51.54	$t(35) = -3.81$, $p < .001$	0.18	0.05	$t(35) = 4.01$, $p < .001$	73.0%

Note. The groups were assigned the labels DP1/DP2 based on which group had the lower mean score on the CFPT inverted (labelled DP1 for consistency with the main manuscript). The cluster analysis reported in the main text is presented in bold.

[a] Indicates the percentage of cases assigned to the same DP cluster as the main cluster analysis (cluster analysis 1).

step we repeated the k-means analysis using the upright and inverted face processing score from the CFPT but the inversion index from the matching task (cluster analysis 2 in Table 2); the upright and inverted face processing scores from the matching task and the inversion index from the CFPT (cluster analysis 3 in Table 2), and several analyses that incorporated measures from the CFMT,[7] CFPT, and the matching task (cluster analyses 4–6 in Table 2).

To examine whether the clusters derived from these analyses were similar, we adopted two approaches. First, we examined the pattern of performance of each cluster on the CFPT (Table 2). All cluster analyses showed the same pattern of performance: for the CFPT upright, there was no significant difference in performance between the DP1 and DP2 clusters; for the CFPT inverted, there was a significant difference between the DP1 cluster and the DP2 cluster (clusters were coded so that the cluster with lower mean performance on the CFPT inverted was always classed as DP1); and for the inversion index, the DP1 cluster showed a significantly larger inversion index than the DP1 cluster.

Second, we examined the agreement between clusters (i.e., the percentage of individuals who were classified in the same cluster in the main analysis and in the subsequent analyses; see Table 2). Overall, agreement was robust ($M = 81.64\%$, $SD = 12.28\%$), indicating that the analyses consistently classified most participants into the same clusters.

Given the substantial agreement between the cluster analysis on the CFPT alone and the cluster analyses incorporating the CFPT, CFMT, and matching task, we elected to conduct the remaining analyses on the clusters derived from the CFPT analysis.

3.2. Characteristics of DP clusters

Table 1 shows the demographic characteristics of each cluster and the control group, along with their mean scores on the diagnostic test battery and CFPT.

To determine how the DP clusters differed from each other, we carried out a series of follow-up analyses. First, we examined whether the DP clusters significantly differed from each other in regard to severity of face recognition problems (measured via their CFMT and famous faces scores), and their holistic and featural processing (measured via performance on the inverted CFMT, face and object matching task, composite task, and Navon task). Mixed ANCOVAs were carried out on each task, with cluster (DP1, DP2) as a between-subjects factor and age as a covariate. Pairwise comparisons and follow-up ANCOVAs were conducted to investigate significant effects. For brevity, only significant and/or theoretically relevant interactions and main effects are reported (all $ps > .05$ for non-significant results). The Greenhouse-Geisser correction is applied where relevant, and multiple comparisons are Bonferroni corrected. Analyses were conducted on unstandardised data.

As the primary focus of this research was to identify and characterise patterns of performance *within* DP, the results presented here focus on

[7] We did not use scores from the CFMT upright in any of the cluster analyses reported here, as the CFMT was used as a screening measure. This restricts the range of scores that the DP group exhibited on this task and could potentially bias the clustering. However, as performance on the inverted version of the CFMT and the inversion index were not used for screening, these were incorporated into cluster analyses 4–6.

participants with DP only. Additional analyses comparing the DP clusters to age-matched controls are presented in the supplementary materials (S1). We also examined whether any of the measures of holistic processing were associated with face recognition abilities by examining the correlations between different upright and inverted face processing and different measures of holistic processing (e.g., inversion indices) for the entire DP sample (independent of clusters). These correlations are also presented in the supplementary materials (S2).

3.2.1. CFMT and famous faces

Fig. 4 displays the individual scores for the two DP clusters (and, for comparison purposes, control participants) on the main DP screening tests: the CFMT and famous face recognition. The ANCOVA comparing DP1 and DP2 clusters revealed that, after controlling for age, the DP1 cluster ($M = 37.29$, $SD = 3.63$) performed significantly more accurately on the CFMT than the DP2 cluster ($M = 32.94$, $SD = 6.36$), $F(1, 34) = 7.56$, $p = .010$, $\eta_p^2 = 0.18$. An identical analysis, after removing an extreme outlier in the DP2 group, also showed a significant difference between the DP1 and DP2 clusters. However, the clusters did not display a significant difference in their famous face recognition, $F(1, 34) = 0.00$, $p = .959$, $\eta_p^2 = 0.00$. This suggests that the DP2 cluster shows a more severe impairment in face learning and/or short term recognition than the DP1 cluster, but the clusters do not discriminate between individuals with higher or lower levels of impairment with long-term memory for faces.

3.2.2. CFMT inversion

Data for the inverted version of the CFMT was used to compare memory for inverted faces across both DP clusters. An ANCOVA on the inverted version of the CFMT revealed a significant difference between the two clusters, with the DP1 cluster ($M = 31.76$, $SD = 5.74$), performing worse than the DP2 cluster ($M = 35.88$, $SD = 4.84$), $F(1, 34) = 4.92$, $p = .033$, $\eta_p^2 = 0.13$.

To examine the effect of inversion, normalised for baseline performance, we calculated an inversion index for each individual, using the formula [(upright − inverted)/(upright + inverted)] (Avidan et al., 2011; Ulrich et al., 2017). An ANCOVA on the inversion index revealed a significant difference between the two groups of participants, $F(1, 34) = 13.02$, $p = < .001$, $\eta_p^2 = 0.28$. The DP1 cluster ($M = 0.87$, $SD = 0.10$) showed a larger effect of inversion than the DP2 cluster ($M = -0.05$, $SD = 0.13$). As in the analysis of the CFMT upright, removal of one outlier from the DP2 group did not change the pattern of results.[8] A violin plot displaying the inversion index for each group is included in Fig. 5.

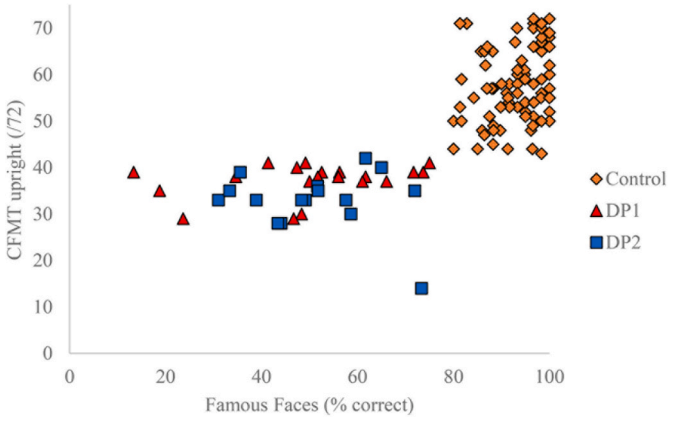


Fig. 4. Scatterplot showing scores on the CFMT (upright) and famous faces task for the two DP clusters and control participants.

3.2.3. Face and object matching

The *d'* results for each cluster are displayed in Fig. 6. A 2 (orientation: upright; inverted) x 2 (object: faces; houses) x 2 (cluster: DP1; DP2) ANCOVA on sensitivity (*d'*) revealed that both within-subjects main effects were significant: orientation: $F(1, 34) = 17.62$, $p < .001$, $\eta_p^2 = 0.34$; object: $F(1, 34) = 154.88$, $p < .001$, $\eta_p^2 = 0.82$; but, averaged across conditions, there was no significant difference between DP clusters $F(1, 34) = 0.19$, $p = .668$, $\eta_p^2 = 0.00$. On average, DP participants performed better on upright than inverted trials and with faces than with houses. Neither of the two-way interactions with group were significant, orientation x cluster $F(1, 34) = 2.39$, $p = .131$, $\eta_p^2 = 0.07$; object x cluster $F(1, 34) = 0.60$, $p = .442$, $\eta_p^2 = 0.02$. However, there was a significant three-way interaction between object, orientation, and cluster, $F(1, 34) = 5.05$, $p = .031$, $\eta_p^2 = 13$. Simple main effects revealed that the DP1 cluster showed a significant difference between upright faces ($M = 1.51$, $SD = 0.33$) and inverted faces ($M = 0.53$, $SD = 0.92$), $p < .001$, whereas the DP2 cluster did not, (upright: $M = 1.34$, $SD = 0.80$; inverted: $M = 0.97$, $SD = 0.78$), $p = .095$. Neither cluster showed significant differences in performance between upright and inverted houses, p's > 0.46, and there was no significant difference between the clusters on any individual condition (faces or houses), p's > 0.96.

An ANCOVA examining the RT results for the two DP clusters revealed no significant main effect of cluster or interactions involving cluster, all p's $> .07$.

Additional analyses on the inversion indices were carried out, and show a similar pattern of results: a significant main effect of object, $F(1, 34) = 16.06$, $p < .001$, $\eta_p^2 = 0.32$, superceded by a significant interaction between object and cluster, $F(1, 34) = 5.32$, $p = .027$, $\eta_p^2 = 0.14$. The DP1 cluster showed a significantly larger inversion index for faces ($M = 0.09$, $SD = 0.10$) than the DP2 cluster ($M = 0.03$, $SD = 0.09$), $p = .034$; there was no significant difference between the inversion indices for houses (DP1: $M = 0.00$, $SD = 0.03$; DP2: $M = 0.01$, $SD = 0.03$; $p > .99$).[9] A violin plot displaying the inversion index for faces and houses across each group is included in Fig. 4.

In sum, the two clusters of DPs did not differ in their ability to match upright and inverted houses. Furthermore, the clusters did not differ in their ability to match upright and inverted faces. However, the effects of face inversion differed between the two clusters. The DP1 cluster showed a significant inversion effect for faces, whereas the DP2 cluster did not. These effects were apparent in the *d'* (sensitivity) analysis, but not in the RT analysis, suggesting that the differences between clusters do not reflect different a speed-accuracy trade-off.

3.2.4. Composite task

Accuracy on the composite task for each cluster is shown in Fig. 6. A 2 (alignment: aligned; misaligned) x 2 (object: faces; dogs) x 2 (orientation: upright, inverted) x 2 (cluster: DP1; DP2) ANCOVA was carried out to investigate performance on the composite task in each DP cluster. In the composite task, the key effect of interest is an interaction between alignment, orientation, and object (typically, this reflects a stronger composite effect for upright faces than for inverted faces, but not for upright compared to inverted objects). This three-way interaction was significant, $F(1, 34) = 17.59$, $p < .001$, $\eta_p^2 = 0.34$. Follow-up simple main effects analyses revealed a significant effect of alignment for upright

[8] As for the analysis of the CFPT, an identical analysis was conducted on inversion effects calculated by subtraction. The ANCOVA showed the same pattern of results as the inversion index.

[9] An identical analysis on inversion effects calculated by subtraction revealed a similar pattern of results, except that the difference between clusters for face inversion only approached significance, $p = .057$.

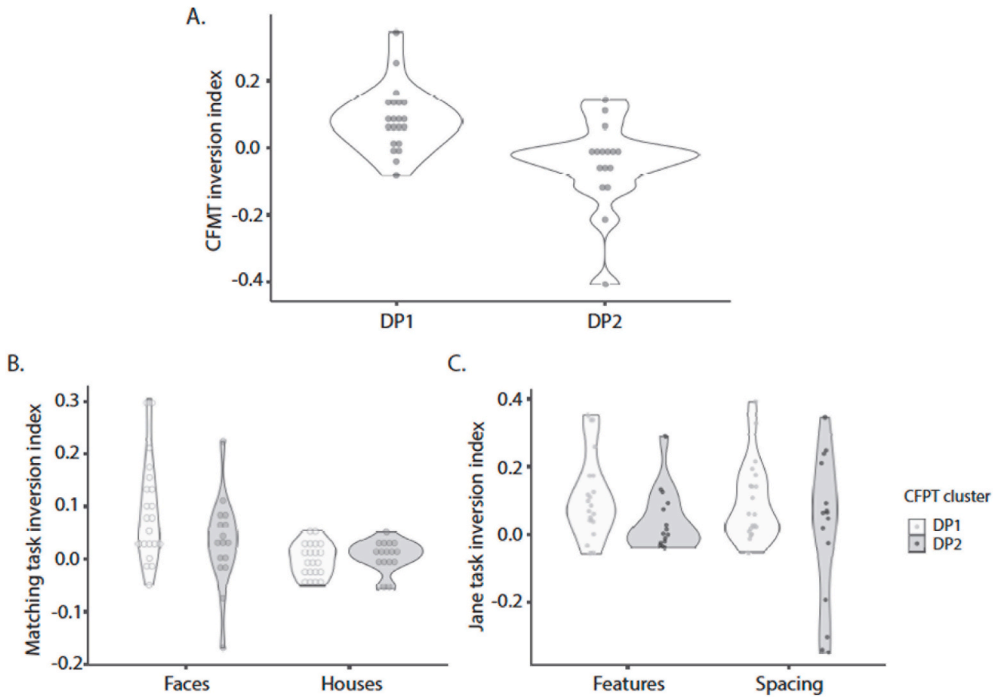


Fig. 5. Violin plots illustrating inversion indices for the two DP clusters. A. CFMT inversion index; B. face and object matching task inversion indices; C. Jane task inversion indices.

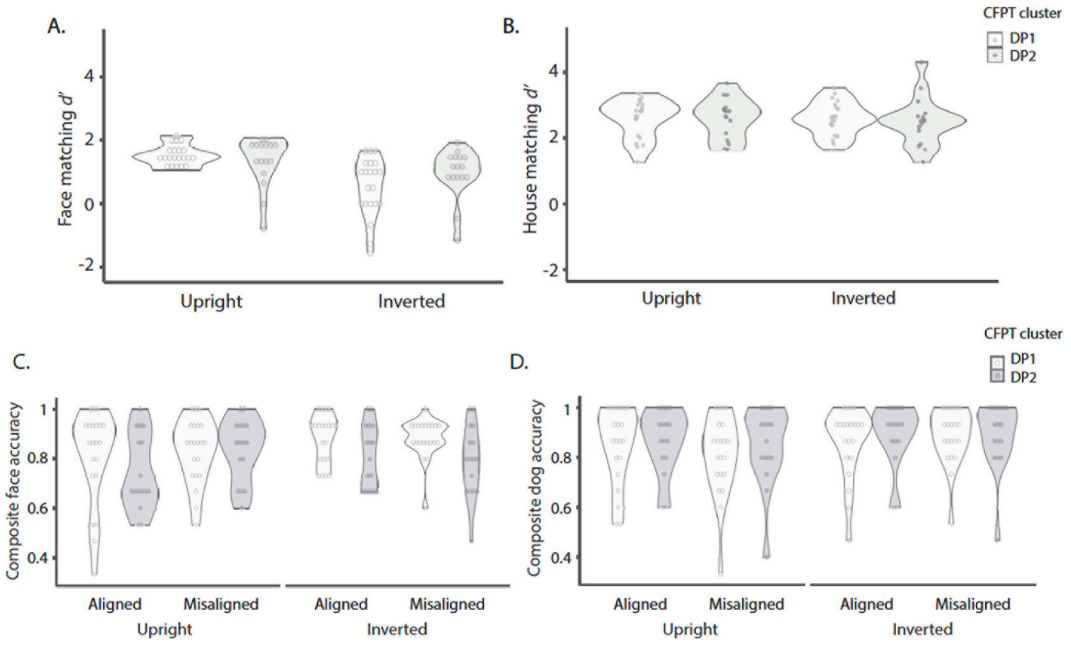


Fig. 6. Violin plots illustrating performance on face and object processing tasks for the two DP clusters. A. d' (sensitivity) in the matching task for faces; B. d' (sensitivity) in the matching task for houses; C. Accuracy in the composite task for faces; D. Accuracy in the composite task for dogs.

faces, $p = .025$, but not for inverted faces, $p = .258$, nor for upright or inverted dogs, p's > 0.15. In short, the DP participants as a group displayed a typical pattern of composite performance. There were no significant three or four-way interactions with cluster, all p's $> .14$, indicating that, unlike the inversion effects for the CFMT and face matching task, the composite effect was not significantly different across the two DP clusters.

There was a significant two-way interaction between cluster and object, $F(1, 34) = 4.52$, $p = .041$, $\eta_{\rho}^2 = 0.12$, reflecting the fact that, averaged across alignment and orientation, the DP1 cluster performed similarly for both faces ($M = 0.85$, $SD = 0.12$) and dogs ($M = 0.85$, $SD = 0.12$), $p = .919$, whereas the DP2 cluster performed significantly better with dogs ($M = 0.88$, $SD = 0.10$) than with faces ($M = 0.80$, $SD = 0.10$), $p = .043$. However, there were no significant differences between the two clusters in their overall accuracy for faces or dogs, p's > 0.8.

An identical ANCOVA was carried out on RT. The key three-way

object x orientation × alignment interaction was not significant, $F(1, 34) = 0.92$, $p = .344$, $\eta_p^2 = 0.03$; nor was any interaction or main effect involving the DP clusters, all p's > 0.14.[10]

Additional analyses comparing the DP clusters to control participants confirm that neither cluster showed a significantly different pattern of performance compared to controls, although the DP1 cluster (but not the DP2 cluster) performed worse than controls when matching dogs, averaged over alignment and orientation (see supplementary materials for full details). This suggests that the DP clusters demonstrate a typical pattern of performance in the composite task.

In sum, the results from the composite task suggest that the two DP clusters do not show a significantly different pattern of composite effects for faces or dogs; therefore it is not possible to conclude that one cluster shows a more generalised deficit in holistic processing for objects compared to the other. However, comparisons to controls (reported in full in the Supplementary materials) provide some preliminary evidence that the deficits in the DP1 cluster may also affect object (dog) recognition.

3.2.5. Navon task

Most participants displayed extremely high accuracy in the Navon task (both clusters showed >95% mean accuracy in all conditions). Given the high levels of performance, and previous suggestions that DP's deficits in the Navon task represent *delayed* processing of global shape information (Gerlach et al., 2017; Gerlach and Starrfelt, 2021), the analysis focused on RT. A 2 (level: global; local) x 2 (congruency: congruent; incongruent) x 2 (cluster: DP1; DP2) ANCOVA on mean correct RTs was carried out to investigate performance on the Navon task in each DP cluster.[11] The ANCOVA revealed a significant effect of congruency, $F(1, 33) = 77.66$, $p < .001$, $\eta_p^2 = 0.70$, with faster responses to congruent than incongruent trials; but no significant main effect of level, $F(1, 33) = 0.99$, $p = .328$, $\eta_p^2 = 0.03$. There was also a significant two-way interaction between congruency and level, $F(1, 33) = 6.60$, $p = .015$, $\eta_p^2 = 0.17$. Simple main effects analyses revealed significant differences between congruent and incongruent trials for both global and local judgements, p's < 0.001; these comparisons indicate that the DPs, as a group, demonstrated both global interference (a significant difference between local congruent and local incongruent trials) and local interference (a significant difference between global congruent and global incongruent trials). However, there were no significant differences between levels for either the congruent or incongruent trials, p's $> .15$, suggesting that, as a group, the DPs did not show a global precedence effect. None of the main effects or interactions involving cluster were significant, all p's $> .23$, indicating that the patterns of performance in the Navon task did not differ significantly across the DP clusters.

To explore the congruency × level interaction further, we directly compared global and local interference for the DP clusters in a 2 (interference: global interference; local interference) x 2 (cluster: DP1; DP2) ANCOVA. There was a significant main effect of interference, $F(1, 33) = 6.60$, $p = .015$, $\eta_p^2 = 0.17$, reflecting significantly greater interference of incongruent local information in global trials ($M = -116.97$, $SE = 13.07$) than global information in incongruent local trials ($M = -62.80$, $SE = 13.07$). No other main effects or interactions in the analysis were significant, p's > 0.10; as such, we found no evidence that this tendency differed between DP clusters.

One-way ANCOVAs with cluster (DP1; DP2) as the between-subjects variable were carried out on the global bias index. There was no significant main effect of cluster, $F(1, 34) = 0.09$, $p = .771$, $\eta_p^2 = 0.00$. A violin plot displaying the Navon task performance for each cluster is included in Fig. 7.

3.2.6. Jane task

The d' results for the Jane task for each cluster are displayed in Fig. 8.[12] A 2 (orientation: upright; inverted) x 2 (condition: spacing; features) x 2 (cluster: DP1; DP2) ANCOVA on sensitivity (d') in the Jane task revealed significant main effects for condition, $F(1,33) = 7.40$, $p = .010$, $\eta_p^2 = 0.18$, and orientation, $F(1, 33) = 12.80$, $p = .001$, $\eta_p^2 = 0.28$. Overall, upright faces were discriminated better than inverted faces, and feature changes were discriminated better than spacing changes. There was no overall difference between the two DP clusters, $F(1,33) = 3.13$, $p = .086$, $\eta_p^2 = 0.09$, but the interaction between condition and cluster was significant, $F(1,33) = 5.42$, $p = .026$, $\eta_p^2 = 0.14$. For the DP1 cluster, there was no significant difference in performance between spacing ($M = 0.96$, $SE = 0.15$) and feature changes, ($M = 1.04$, $SE = 0.19$) $p > .9$; however, participants in the DP2 cluster performed significantly better in the feature ($M = 1.80$ $SE = 0.22$) than spacing discrimination condition ($M = 0.95$, $SE = 0.18$), $p = .011$. The clusters did not differ significantly in their performance on spacing discrimination trials, $p > .9$, however the DP2 cluster performed significantly better than the DP1 cluster in feature discrimination trials, $p = .038$. No other interactions were significant, all p's > 0.05; nor was the main effect of group, $F(1, 33) = 3.13$, $p = .086$, $\eta_p^2 = 0.09$.

A priori Bonferroni-corrected comparisons were carried out to examine the presence of inversion effects for each cluster in each condition. The DP1 cluster showed a significant inversion effect for feature discrimination $p = .012$, but not spacing discrimination, $p = .065$; the DP2 cluster did not show a significant inversion effect for either condition, p's > 0.900.

An identical ANCOVA on RT in the Jane task revealed no significant main effects or interactions, all p's > 0.1.

The inversion index for each condition is displayed in Fig. 5. Additional analyses on the inversion indices and the inversion effect measured by subtraction were carried out; these suggest that the DP1 cluster shows a slightly larger inversion effect than the DP2 cluster overall. These are reported in full (along with comparisons to control participants) in the supplementary materials.

3.3. Results summary

Summing up, the cluster analysis, based on performance on the CFPT (upright, inverted, and inversion index), identified two separable clusters of individuals with DP. Subsequent analyses confirmed that the two clusters did not differ on most measures of upright face recognition (famous faces, face matching, spacing discriminations, composite task performance) or inverted face recognition (face matching, Jane task). The clusters also did not differ in their performance on the composite and Navon tasks, nor on their object (house) matching performance. The key difference between the clusters was the presence of a significant face inversion effect in the larger cluster (DP1), but not in the smaller cluster (DP2). This difference appeared in both the CFMT and face matching tasks, and in the feature discrimination condition of the Jane task. In the Jane task, the smaller cluster (DP2) performed significantly better than the larger cluster (DP1) when discriminating feature changes.

Comparisons with the control group indicated that the larger (DP1) cluster performed poorly in most conditions involving inverted faces (inverted CFMT, inverted face matching, inverted conditions in the Jane task), and showed some preliminary evidence of broader perceptual deficits (dog perception, local-to-glocal interference in the Navon task).

[10] Analyses were repeated after the removal of one outlier in the DP1 cluster. This did not change the pattern of results, so we report results from the full sample here.

[11] The Navon task analyses were repeated after the removal of one participant in the DP2 group who was an outlier in accuracy – this did not alter the pattern of results for any measure. One participant in the DP2 group was an extreme influential outlier on the local-to-global interference measure, and was removed from the analysis.

[12] One DP participant's data was excluded from analysis as an outlier.

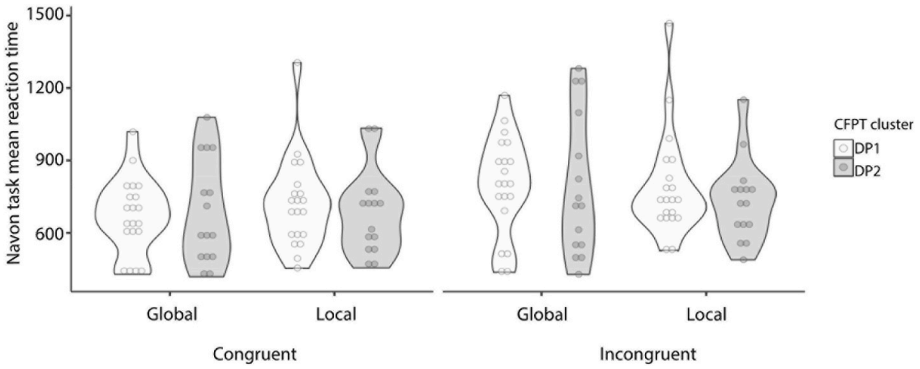


Fig. 7. Violin plots illustrating mean reaction time on the Navon task for the two DP clusters.

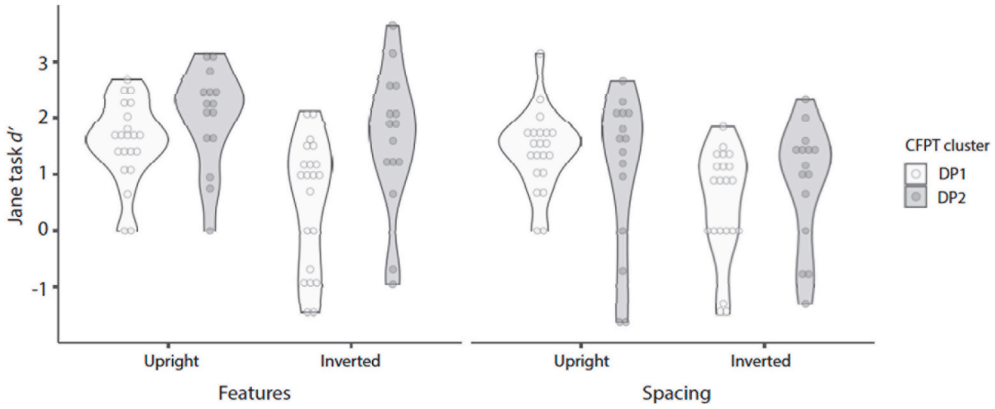


Fig. 8. Violin plots illustrating performance on the Jane task for the two DP clusters.

In contrast, the smaller (DP2) cluster performed similarly to controls in several tasks involving inverted faces (face matching, Jane task), but showed a reduced inversion effect in several tasks (CFMT, face matching, Jane task) and did not differ from controls in any non-face tasks or conditions.

4. Discussion

The perceptual underpinnings of DP have been the topic of substantial debate in the literature. Many studies have investigated whether impairments in holistic or configural processing (either face-specific or more global impairments) can account for the pattern of deficits in DP, with generally conflicting results. This study used cluster analysis to examine whether there are separate patterns of holistic or configural processing impairments present within the DP population. A two-stage cluster analysis on a commonly used, standardised task of face perception (the CFPT) revealed two separate clusters within our group of individuals with DP. Subsequent analyses tested this cluster solution on tasks unseen in the clusters – this allowed us to determine the robustness and generalisability of the cluster solution. The two clusters of DPs were not differentiated by their performance on upright face perception tasks, which suggests that the clusters do not simply reflect individuals with more or less severe perceptual deficits in general. Nor were the clusters differentiated by their performance on non-face processing (house matching, recognition of composite dogs, global/local bias on the Navon task). Rather, this pattern of performance suggests at least two potential pathways to face processing deficits, only one of which involves deficits in holistic processing (as measured by inversion effects).

The initial cluster solution appeared to discriminate between individuals who showed different levels of face inversion effects on the CFPT (see Fig. 2). This was supported by follow-up analyses which confirmed that one cluster (DP1) showed significant negative effects of inversion on face processing (across the CFMT and face matching tasks), whereas the second cluster (DP2) did not. Supplementary analyses comparing the clusters to controls showed that the DP2 cluster demonstrated a significantly reduced inversion index across multiple tasks (CFPT, CFMT, Jane task). Therefore, it is likely that some cases of DP – the minority in our sample – are characterised by a focal deficit in the face-specific processing indexed by inversion effects. This finding is in accord with a large amount of previous work which has shown significantly reduced inversion effects in DP (e.g., DeGutis et al., 2012; Klargaard et al., 2018), however, it indicates that reduced inversion effects are not common to all individuals with DP. Using a clustering approach, as adopted in the current study, may allow researchers to discriminate between groups of individuals with DP, and offer a clearer insight into the perceptual mechanisms that are intact and impaired within each group.

There remains some debate as to what exactly is disrupted in typical observers when a face is inverted, and, by extension, the processes that may be abnormal in the DP2 cluster. While some authors suggest that inversion impairs the processing of configural information (Richler et al., 2012), others suggest that inversion narrows the perceptual field when viewing a face, so that features must be processed in a piecemeal manner, and only spatial relationships within a very small field can be processed efficiently (Rossion, 2008; see Tanaka and Gordon, 2011 for a review). Thus, it is possible that the group of DPs with reduced inversion effects (DP2) in the current study may present with deficits in configural processing, or they may show an abnormally small perceptual field during face processing which prevents integration of information from different areas of the face. The current battery did not include any tasks that were designed to explicitly discriminate between these possibilities, but it is noteworthy that we did not observe any significant differences between the clusters in the spacing discrimination condition of the Jane task, which assesses configural processing. However, the DP2 cluster,

who did not show significant face inversion effects, performed better at feature-based discrimination than spacing-based discrimination in the Jane task (this pattern was not observed for the DP1 cluster) – this provides some support for the idea that the DP2 cluster may be characterised by more piecemeal processing, perhaps due to a smaller perceptual field.

The other, larger, cluster of individuals with DP in this study (DP1) showed a typical pattern of face inversion effects – that is, their performance with inverted faces was significantly lower than their performance with upright faces on both the CFMT and face matching tasks, as well as the feature discrimination condition in the Jane task. While this suggests that some holistic processing remains intact in this cluster, it is unclear whether this reflects "typical" levels of holistic processing for faces. Supplementary analyses comparing the inversion indices for the DP1 cluster to control participants indicate that this group of individuals with DP did not show a significant reduction in inversion indices when compared to controls across several face perception tasks (CFPT, face matching, Jane task), suggesting that the processes that differentiate upright and inverted face perception in the typical population are somewhat intact in this group. However, compared to controls, this group showed a significantly reduced inversion effect for face memory (the CFMT). These findings are in line with those of Klaargard et al., 2018, who reported reduced inversion effects for individuals with DP on the CFMT, but not the CFPT.

It is unclear why this difference between perception and memory tasks arises – one possibility is that both studies suffered from restriction-of-range effects, as the inclusion criteria for both studies means that DP participants' scores on the upright version of the CFMT were limited to a maximum of 42/72. Other possibilities are that memory tasks are more sensitive to inversion effects, or that inversion effects for memory and perception are dissociable. While it is not possible to distinguish between these possibilities on the basis of our data, the fact that the two DP clusters showed similar patterns of inversion effects (i.e., DP1 showed a significantly larger inversion effect than DP2) for both memory and perceptual tasks indicates that the differences between the clusters were somewhat consistent across different measures, and argues against a complete dissociation between inversion effects for memory and perception. Nonetheless, our results reinforce the idea that face inversion effects are partially task-dependent, and conclusions about the degree of holistic processing in individuals with DP should be based on an examination of multiple tasks.

Given the results from the CFMT inversion effect, it is possible that the first cluster shows more subtle deficits in holistic processing than the second cluster. The DP1 cluster also performed better than the DP2 cluster in the CFMT, raising the possibility that the clusters reflect different levels of impairment rather than distinct groups. This contention is partially supported by correlational analyses on the DP group that show a small relationship between inversion indices for spacing discriminations in the Jane task (which, in theory, reflect configural processing; Mondloch et al., 2002; Rossion, 2008) and performance on the CFMT and CFPT (although note that the correlation is not significant after correction for multiple comparisons). One potential interpretation of these findings is that larger deficits in configural processing are (somewhat) associated with larger deficits in face recognition generally. However, the inversion indices for the CFPT and face matching (which should also reflect configural or holistic processing; Rossion, 2008) do not correlate with performance on either the CFMT or the famous faces task. Furthermore, the DP1 cluster did not show better performance than the DP2 cluster in any other upright face processing tasks – famous face recognition, face matching, composite face processing, or any condition in the Jane task.

This raises the question of whether the DP1 cluster shows a separate perceptual deficit compared to the DP2 cluster. One possibility is that the DP1 cluster shows deficits in featural perception. The DP1 group performed substantially worse than the DP2 cluster on inverted face trials in the CFPT and CFMT (although this finding was not replicated in the face matching task). As discussed above, inverting a face is likely to disrupt holistic and configural processing, while leaving featural processing somewhat intact (Maurer et al., 2002; Rossion, 2008); therefore, a deficit in inverted face perception may reflect a difficulty with feature processing. This is supported by the findings from the Jane task, which indicated that the DP1 cluster performed significantly worse at feature discriminations than the DP2 cluster. Further, supplementary analyses confirmed that the DP1 group (but not the DP2 group) were significantly impaired at inverted face perception when compared to control participants across multiple tasks (CFPT, face matching, feature discrimination).

Previous work supports the contention that featural processing can be impaired in some cases of DP – for example, Biotti and Cook (2016) found that a subset of DPs – those who showed poor performance on the CFPT – also performed poorly in a facial expression processing task which involved isolated face parts, suggestive of an early deficit in feature-shape encoding (see also Le Grand et al., 2006; Schmalzl et al., 2008; Yovel and Duchaine, 2006). Deficits in feature perception could also explain a somewhat reduced inversion effect as we observed in the larger DP cluster in the current study – several accounts of face processing explicitly note that feature shape is likely to be encoded as part of the configural or holistic face representation (McKone and Yovel, 2009), and spacing relationships are based on key points surrounding the features (see Piepers and Robbins, 2012 for an overview). Consequently, if it is the case that some individuals with DP form poor perceptual representations of individual features, it is conceivable that their representation of spatial relationships between those features is also disrupted. As such, a primary deficit in local feature processing resulting in a weakened – but not absent – capacity to process spatial information from faces could explain why these individuals still show significant inversion effects in some tasks, while also accounting for the fact that these inversion effects were smaller than controls in the task with the highest cognitive demands – specifically, the CFMT.

The presence of distinct patterns of performance reflecting deficits in holistic and featural processing is in line with models of face processing that posit separable "channels" for part-based processing and configural/holistic processing (Piepers and Robbins, 2012), both of which contribute to face recognition. However, further research (preferably with larger samples) is needed to confirm whether the DP1 cluster does in fact show a reliable deficit in part-based processing, and whether this is consistent across multiple face processing tasks. Given the current clustering findings are based on a widely-used screening measure, employed in DP research worldwide, it is possible for other researchers to replicate our clusters in their own samples, and examine featural processing in this group more systematically.

Our findings revealed some important differences between DP clusters. However, it also highlighted some similarities between clusters. Although the cluster analysis consistently discriminated between the DP subgroups on the basis of their inversion effects, there was no evidence that the groups differed in the strength of their composite face effect. These findings support recent research in the typical population which suggests that different measures of holistic processing may reflect different underpinning processes (Rezlescu et al., 2017).

Further supplementary analyses confirmed that there was no evidence of abnormal composite effects in either of the DP clusters. This is consistent with other recent research in DP (Biotti et al., 2017; Ulrich et al., 2017). The lack of group-level differences in the composite task is not due to a failure of the task to index holistic processing – note that on a whole-group level (i.e., ignoring the different DP clusters), a traditional composite effect was observed for accuracy. Nonetheless, it is possible that the use of alternative versions of the composite task (e.g., the 'interference' or 'complete' design, see Richler and Gauthier, 2014) could result in slightly different outcomes, as it is likely that it measures somewhat different processes to the traditional composite task used in this experiment (see Rossion, 2013 for an overview). It is not implausible

that individuals with DP may vary on those congruency-based measures, but not on the holistic processing measured by the traditional composite design. Future research into holistic processing deficits in DP should therefore consider using multiple measures which assess different potential "holistic processing" mechanisms.

One point that remains unclear from the current results is whether the differences between the two DP clusters are isolated to faces. We assessed object processing (including measures of holistic processing for objects) in two ways: house matching (and inversion effects), and the composite effect for dogs. The DP clusters did not significantly differ from each other on either of the holistic object processing measures, or on their performance in the house matching task. This provides some evidence that the differences between these clusters does not reflect a general holistic processing deficit (although note that there was little evidence of holistic object processing overall – neither the inversion effect for houses nor the composite effect for dogs was significant in any group). However, compared to controls, the DP1 cluster showed poorer performance for dogs in the composite task. Given the absence of a composite effect in the dog trials, this is unlikely to reflect any difficulty with the holistic processing of dogs in this group – instead, it may indicate a subtle deficit in part-based processing that extends beyond face parts.

In addition to assessing object processing, we also examined global and local shape processing using the Navon task. There were no differences between the DP clusters in any measure related to the Navon task. Further, while the DP group as a whole failed to show a global precedence effect – a somewhat unusual pattern of results – the supplementary analyses did not reveal significant differences between the DP clusters and controls on the global precedence measure, or on the global or local interference effects.

In sum, comparisons with controls in the composite task (dogs) offer some evidence that the DP1 cluster may show subtle deficits in non-face stimuli. We found no evidence of similar deficits in the DP2 cluster, nor were any differences apparent in the Navon task. Importantly, the differences between the clusters in dog matching performance were not significant, nor were any significant differences found in comparisons between groups for the houses; consequently, any conclusions about broader deficits in the DP1 cluster are tentative. It is possible that the lack of difference between clusters is simply an issue of power – while this study included a relatively large sample compared to previous work on DP, the sample may still be too small to detect subtle group differences.

The comparisons between the DP clusters and controls support the idea that subtle non-face object processing deficits may be present in some (but not all) individuals with DP. Other research has reached similar conclusions – in a recent study using a partially overlapping population, only 6/15 individuals with DP showed good evidence for intact object recognition, in the form of a classical dissociation between face and object matching abilities (Bate et al., 2019c) (see also Barton et al., 2019; Gerlach et al., 2022). The current study provides some indication that these broader deficits may arise from deficits in processing local features or parts of stimuli (including, but not limited to, faces). This supports the idea that there is at least some overlap between the featural or part-based processing processes applied to faces and objects, (McKone and Yovel, 2009; Piepers and Robbins, 2012). Crucially, the differing patterns of performance in the two DP clusters also suggests that these part-based processing deficits may be separable from holistic/configural processing deficits.

It is important to note that we are not claiming that these clusters explain all of the heterogeneity present in DP. Naturally, the results from a cluster analysis are dependent on the measures included in it. We chose to include measures derived from the CFPT, as this increases the generalisability and comparability of our results to other groups of DPs, and it offers a simple measure of holistic processing (inversion effects). However, it is important to acknowledge that the reliability for the CFPT is middling, which could add noise to the measurements and reduce the precision of the clustering process. Nonetheless, we also found similar clusters when three separate measures of face processing and inversion effects were averaged and subjected to the same cluster analysis, and when measures of face processing were derived from different tasks (e.g., a combination of the CFMT, CFPT, and matching task). This suggests that the clusters derived from the CFPT alone are relatively robust and reflect stable differences across tasks. However, it is possible that a separate analysis including a different battery of tasks could result in additional clusters, differentiating between cases based on the presence or absence of broader object recognition impairments, other perceptual processes (e.g., face adaptation or serial dependence), or even mnemonic vs perceptually based deficits. It is also possible that further research into these subtypes (or similarly derived groupings of DPs) will reveal more generalised perceptual deficits in one or both groups, if more targeted follow-up tests are used. However, the current results argue against a single explanation of DP which relies on domain-general perceptual processing deficits (e.g., Avidan et al., 2011; Geskin and Behrmann, 2018). Instead, research should take into account the complexity of face recognition and the heterogeneity of the DP population, and acknowledge that face processing deficits may reflect multiple, sometimes separable perceptual processes.

In sum, the work presented here provides evidence that more than one perceptual deficit may underpin face recognition deficits in DP. Consequently, we suggest that behavioural heterogeneity in DP should be considered before drawing conclusions about universal patterns of deficits (or lack thereof) in this population. The idea that there may be different subtypes of face recognition deficits is not new, and has been explored using case series in both the acquired and developmental prosopagnosia literature (e.g., Barton, 2008; Gainotti and Marra, 2011; Le Grand et al., 2006; Schmalzl et al., 2008; Ulrich et al., 2017); however, this paper offers a new way to systematically examine this heterogeneity on a larger scale in the DP population. In this analysis, we focused on configural and holistic processing – one of the primary questions that has occupied the DP literature over the past two decades – but future work applying similar techniques may reveal similar heterogeneity in regards to object processing deficits, social perception abilities, and the relationship between face processing and other cognitive skills (e.g., Bate, Adams, Bennetts, et al., 2019; Corrow et al., 2019). Developing a more thorough understanding of the commonalities and differences between individuals with DP could open new avenues of research into the cognitive and neural underpinnings of face processing, the development of face processing abilities and deficits across the lifespan, and the development of targeted rehabilitation programmes to improve face recognition in individuals with DP.

Credit author statement

Rachel J Bennetts: Conceptualisation, Methodology, Formal analysis, Investigation, Data curation, Visualisation, Writing – original draft; **Nicola J Gregory**: Conceptualisation, Methodology, Investigation, Writing – review & editing; **Jeremy Tree**: Conceptualisation, Resources, Writing – review & editing; **Caroline Di Bernardi Luft**: Software, Formal analysis, Visualisation, Writing – review & editing; **Michael J Banissy**: Supervision, Funding acquisition, Resources, Writing – review & editing; **Ebony Murray**: Investigation, Data curation, Writing – review & editing; **Tegan Penton**: Investigation, Data curation, Writing – review & editing; **Sarah Bate**: Conceptualisation, Methodology, Investigation, Resources, Supervision, Project administration, Funding acquisition, Writing – review & editing.

Acknowledgements

MJB and TP were supported by an ESRC grant (ES/K00882X/1). SB is supported by a Leverhulme Trust Research Fellowship (RF-2020-105). We thank Dr Rachel Robbins and Dr Catherine Mondloch for providing the composite and Jane task stimuli used in the current study; Amanda

Adams for her assistance with data collection; and several anonymous reviewers for constructive comments on a previous version of the manuscript. We also thank all our participants with developmental prosopagnosia.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.neuropsychologia.2022.108332.

References

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

Baron-Cohen, S., Wheelwright, S., Skinner, R., Martin, J., Clubley, E., 2001. The autism-spectrum quotient (AQ): evidence from asperger syndrome/high-functioning autism, malesand females, scientists and mathematicians. J. Autism Dev. Disord. 31 (1), 5–17.

Barton, J.J.S., 2008. Structure and function in acquired prosopagnosia: lessons from a series of 10 patients with brain damage. J. Neuropsychol. 2 (1), 197–225. https://doi.org/10.1348/174866407X214172.

Barton, J.J.S., Albonico, A., Susilo, T., Duchaine, B., Corrow, S.L., 2019. Object recognition in acquired and developmental prosopagnosia. Cogn. Neuropsychol. 36 (1–2), 54–84. https://doi.org/10.1080/02643294.2019.1593821.

Barton, J.J.S., Cherkasova, M.V., Hefter, R., Cox, T.A., O'Connor, M., Manoach, D.S., 2004. Are patients with social developmental disorders prosopagnosic? Perceptual heterogeneity in the Asperger and socio-emotional processing disorders. Brain 127 (8), 1706–1716. https://doi.org/10.1093/brain/awh194.

Bate, S., Adams, A., Bennetts, R., Line, H., 2019a. Developmental prosopagnosia with concurrent topographical difficulties: a case report and virtual reality training programme. Neuropsychol. Rehabil. 29 (8), 1290–1312. https://doi.org/10.1080/09602011.2017.1409640.

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

Bate, S., Bennetts, R.J., Tree, J.J., Adams, A., Murray, E., 2019c. The domain-specificity of face matching impairments in 40 cases of developmental prosopagnosia. Cognition 192, 104031. https://doi.org/10.1016/j.cognition.2019.104031.

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

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

Bennetts, R.J., Murray, E., Boyce, T., Bate, S., 2017. Prevalence of face recognition deficits in middle childhood. Q. J. Exp. Psychol. 70 (2), 234–258. https://doi.org/10.1080/17470218.2016.1167924.

Bentin, S., DeGutis, J.M., D'Esposito, M., Robertson, L.C., 2007. Too many trees to see the forest: performance, event-related potential, and functional magnetic resonance imaging manifestations of integrative congenital prosopagnosia. J. Cognit. Neurosci. 19 (1), 132–146. https://doi.org/10.1162/jocn.2007.19.1.132.

Biotti, F., Cook, R., 2016. Impaired perception of facial emotion in developmental prosopagnosia. Cortex 81, 126–136. https://doi.org/10.1016/j.cortex.2016.04.008.

Biotti, F., Gray, K.L.H., Cook, R., 2018. Is developmental prosopagnosia best characterised as an apperceptive or mnemonic condition? Neuropsychologia 124, 285–298. https://doi.org/10.1016/j.neuropsychologia.2018.11.014.

Biotti, F., Wu, E., Yang, H., Jiahui, G., Duchaine, B., Cook, R., 2017. Normal composite face effects in developmental prosopagnosia. Cortex 95, 63–76. https://doi.org/10.1016/j.cortex.2017.07.018.

Bobak, A.K., Bennetts, R.J., Parris, B.A., Jansari, A., Bate, S., 2016. An in-depth cognitive examination of individuals with superior face recognition skills. Cortex 82, 48–62. https://doi.org/10.1016/j.cortex.2016.05.003.

Corrow, S.L., Dalrymple, K.A., Barton, J.J., 2016. Prosopagnosia: current perspectives. Eye Brain 8, 165–175. https://doi.org/10.2147/EB.S92838.

Corrow, S.L., Stubbs, J.L., Schlaug, G., Buss, S., Paquette, S., Duchaine, B., Barton, J.J.S., 2019. Perception of musical pitch in developmental prosopagnosia. Neuropsychologia 124, 87–97. https://doi.org/10.1016/j.neuropsychologia.2018.12.022.

Dalrymple, K.A., Garrido, L., Duchaine, B., 2014. Dissociation between face perception and face memory in adults, but not children, with developmental prosopagnosia. Developmental Cognitive Neuroscience 10, 10–20. https://doi.org/10.1016/j.dcn.2014.07.003.

Dalrymple, K.A., Palermo, R., 2016. Guidelines for studying developmental prosopagnosia in adults and children: guidelines for studying DP in adults and children. Wiley Interdisciplinary Reviews: Cognit. Sci. 7 (1), 73–87. https://doi.org/10.1002/wcs.1374.

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

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

Duchaine, B., Germine, L., Nakayama, K., 2007a. Family resemblance: ten family members with prosopagnosia and within-class object agnosia. Cogn. Neuropsychol. 24 (4), 419–430. https://doi.org/10.1080/02643290701380491.

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

Duchaine, B., Yovel, G., Nakayama, K., 2007b. No global processing deficit in the Navon task in 14 developmental prosopagnosics. Soc. Cognit. Affect Neurosci. 2 (2), 104–113. https://doi.org/10.1093/scan/nsm003.

Folstein, M.F., Folstein, S.E., McHugh, P.R., 1975. Mini-mental state": a practical method for grading the cognitive state of patients for the clinician. J. Psychiatr. Res. 12 (3), 189–198.

Gainotti, G., Marra, C., 2011. Differential contribution of right and left temporo-occipital and anterior temporal lesions to face recognition disorders. Front. Hum. Neurosci. 5 https://doi.org/10.3389/fnhum.2011.00055.

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

Gerlach, C., Barton, J.J.S., Albonico, A., Malaspina, M., Starrfelt, R., 2022. Contrasting domain-general and domain-specific accounts in cognitive neuropsychology: an outline of a new approach with developmental prosopagnosia as a case. In: Behavior Research Methods. https://doi.org/10.3758/s13428-021-01774-4.

Gerlach, C., Klargaard, S.K., Petersen, A., Starrfelt, R., 2017. Delayed processing of global shape information in developmental prosopagnosia. PLoS One 12 (12), e0189253. https://doi.org/10.1371/journal.pone.0189253.

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

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

Holdnack, H.A., 2001. Wechsler Test of Adult Reading: WTAR. The Psychological Corporation, San Antonio, TX.

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

Humphreys, G., Riddoch, M.J., 1993. Birmingham Object Recognition Battery. Lawrence Erlbaum Associates, Hove, UK.

Jain, A.K., Murty, M.N., Flynn, P.J., 1999. Data clustering: a review. ACM Comput. Surv. 31 (3), 264–323.

JASP Team. (2019). JASP (0.10.0.0) [Computer software]..

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). Am. J. Med. Genet. 140A (15), 1617–1622. https://doi.org/10.1002/ajmg.a.31343.

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

Lange, R.T., Iverson, G.L., Senior, G.J., Chelune, G.J., 2002. A primer on cluster analysis applications to cognitive rehabilitation research. J. Cognit. Rehabil. 20 (1), 16–33.

Larner, A.J., 2012. Mini-Mental Parkinson (MMP) as a dementia screening test: comparison with the mini-mental state examination (MMSE). Curr. Aging Sci. 5 (2), 136–139.

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

Lewandowski, K.E., Sperry, S.H., Cohen, B.M., Öngür, D., 2014. Cognitive variability in psychotic disorders: a cross-diagnostic cluster analysis. Psychol. Med. 44 (15), 3239–3248. https://doi.org/10.1017/S0033291714000774.

Liu, T.T., Behrmann, M., 2014. Impaired holistic processing of left-right composite faces in congenital prosopagnosia. Front. Hum. Neurosci. 8 https://doi.org/10.3389/fnhum.2014.00750.

Macmillan, N.A., Creelman, C.D., 2005. Detection Theory: A User's Guide. Lawrence Erlbaum Associates Publishers.

Maurer, D., Le Grand, R., Mondloch, C.J., 2002. The many faces of configural processing. Trends Cognit. Sci. 6 (6), 255–260. https://doi.org/10.1016/S1364-6613(02)01903-4.

McKone, E., Robbins, R., 2007. The evidence rejects the expertise hypothesis: reply to Gauthier & Bukach. Cognition 103 (2), 331–336. https://doi.org/10.1016/j.cognition.2006.05.014.

McKone, E., Yovel, G., 2009. Why does picture-plane inversion sometimes dissociate perception of features and spacing in faces, and sometimes not? Toward a new theory of holistic processing. Psychonomic Bull. Rev. 16 (5), 778–797. https://doi.org/10.3758/PBR.16.5.778.

Mondloch, C.J., Le Grand, R., Maurer, D., 2002. Configural face processing develops more slowly than featural face processing. Perception 31 (5), 553–566. https://doi.org/10.1068/p3339.

Murphy, J., Gray, K.L.H., Cook, R., 2017. The composite face illusion. Psychonomic Bull. Rev. 24 (2), 245–261. https://doi.org/10.3758/s13423-016-1131-5.

Murray, E., Hills, P.J., Bennetts, R.J., Bate, S., 2018. Identifying hallmark symptoms of developmental prosopagnosia for non-experts. Sci. Rep. 8 (1) https://doi.org/10.1038/s41598-018-20089-7.

Navon, D., 1977. Forest before trees: the precedence of global features in visual perception. Cognit. Psychol. 9 (3), 353–383. https://doi.org/10.1016/0010-0285(77)90012-3.

Pacheco, A., Reis, A., Araújo, S., Inácio, F., Petersson, K.M., Faísca, L., 2014. Dyslexia heterogeneity: cognitive profiling of Portuguese children with dyslexia. Read. Writ. 27 (9), 1529–1545. https://doi.org/10.1007/s11145-014-9504-5.

Palermo, R., Willis, M.L., Rivolta, D., McKone, E., Wilson, C.E., Calder, A.J., 2011. Impaired holistic coding of facial expression and facial identity in congenital prosopagnosia. Neuropsychologia 49 (5), 1226–1235. https://doi.org/10.1016/j.neuropsychologia.2011.02.021.

Piepers, D.W., Robbins, R.A., 2012. A review and clarification of the terms "holistic," "configural," and "relational" in the face perception literature. Front. Psychol. 3 https://doi.org/10.3389/fpsyg.2012.00559.

Prior, M., Eisenmajer, R., Leekam, S., Wing, L., Gould, J., Ong, B., Dowe, D., 1998. Are there subgroups within the autistic spectrum? A cluster analysis of a group of children with autistic spectrum disorders. J. Child Psychol. Psychiatry Allied Discip. 39 (6), 893–902.

Rezlescu, C., Pitcher, D., Duchaine, B., 2012. Acquired prosopagnosia with spared within-class object recognition but impaired recognition of degraded basic-level objects. Cogn. Neuropsychol. 29 (4), 325–347. https://doi.org/10.1080/02643294.2012.749223.

Rezlescu, C., Susilo, T., Wilmer, J.B., Caramazza, A., 2017. The inversion, part-whole, and composite effects reflect distinct perceptual mechanisms with varied relationships to face recognition. J. Exp. Psychol. Hum. Percept. Perform. 43 (12), 1961–1973. https://doi.org/10.1037/xhp0000400.

Richler, J.J., Gauthier, I., 2013. When intuition fails to align with data: a reply to Rossion (2013). Vis. Cognit. 21 (2), 254–276. https://doi.org/10.1080/13506285.2013.796035.

Richler, J.J., Gauthier, I., 2014. A meta-analysis and review of holistic face processing. Psychol. Bull. 140 (5), 1281–1302. https://doi.org/10.1037/a0037004.

Richler, J.J., Palmeri, T.J., Gauthier, I., 2012. Meanings, mechanisms, and measures of holistic processing. Front. Psychol. 3 https://doi.org/10.3389/fpsyg.2012.00553.

Robbins, R., McKone, E., 2007. No face-like processing for objects-of-expertise in three behavioural tasks. Cognition 103 (1), 34–79. https://doi.org/10.1016/j.cognition.2006.02.008.

Rossion, B., 2008. Picture-plane inversion leads to qualitative changes of face perception. Acta Psychol. 128 (2), 274–289. https://doi.org/10.1016/j.actpsy.2008.02.003.

Rossion, B., 2013. The composite face illusion: a whole window into our understanding of holistic face perception. Vis. Cognit. 21 (2), 139–253. https://doi.org/10.1080/13506285.2013.772929.

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

Schmalzl, L., Palermo, R., Coltheart, M., 2008. Cognitive heterogeneity in genetically based prosopagnosia: a family study. J. Neuropsychol. 2 (1), 99–117. https://doi.org/10.1348/174866407X256554.

Susilo, T., McKone, E., Dennett, H., Darke, H., Palermo, R., Hall, A., Pidcock, M., Dawel, A., Jeffery, L., Wilson, C.E., Rhodes, G., 2010. Face recognition impairments despite normal holistic processing and face space coding: evidence from a case of developmental prosopagnosia. Cogn. Neuropsychol. 27 (8), 636–664. https://doi.org/10.1080/02643294.2011.613372.

Tanaka, J.W., Gordon, I., 2011. Features, configuration, and holistic face processing. In: Rhodes, G., Calder, A.J., Johnson, M., Haxby, J.V. (Eds.), Oxford Handbook of Face Perception. Oxford University Press, pp. 177–194. https://doi.org/10.1093/oxfordhb/9780199559053.013.0010.

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

Valentine, T., 1988. Upside-down faces: a review of the effect of inversion upon face recognition. Br. J. Psychol. 79 (4), 471–491. https://doi.org/10.1111/j.2044-8295.1988.tb02747.x.

Weigelt, S., Koldewyn, K., Kanwisher, N., 2012. Face identity recognition in autism spectrum disorders: a review of behavioral studies. Neurosci. Biobehav. Rev. 36 (3), 1060–1084. https://doi.org/10.1016/j.neubiorev.2011.12.008.

Yin, R.K., 1969. Looking at upside-down faces. J. Exp. Psychol. 81 (1), 141–145. https://doi.org/10.1037/h0027474.

Yovel, G., Duchaine, B., 2006. Specialized face perception mechanisms extract both part and spacing information: evidence from developmental prosopagnosia. J. Cognit. Neurosci. 18 (4), 580–593. https://doi.org/10.1162/jocn.2006.18.4.580.