Self-Ratings of Face Recognition Ability are Influenced by Gender but not Prosopagnosia Severity

Ebony Murray and Sarah Bate
Department of Psychology, Bournemouth University, UK

Author Note
Ebony Murray, Department of Psychology, Bournemouth University, UK.
Sarah Bate, Department of Psychology, Bournemouth University, UK. SB is supported by a British Academy Mid-Career Fellowship.
Correspondence concerning this report should be addressed to Ebony Murray, Department of Psychology, Faculty of Science and Technology, Poole House, Bournemouth University, Fern Barrow, Poole, BH12 5BB.
Email: emurray@bournemouth.ac.uk
Abstract

Growing awareness of developmental prosopagnosia (DP) has resulted in large numbers of people self-referring for prosopagnosia screening. Objective assessment depends heavily on available resources; thus, some researchers use self-ratings of face recognition ability to reduce candidate lists. However, our own metacognitive awareness of our face recognition skills has been much debated, and there is mixed evidence on the reliability of self-report measures. Nevertheless, some behavioral trait questionnaires have proved more useful - although it remains unclear whether these instruments can tap prosopagnosia severity, or whether responses are influenced by participant gender (as in other developmental disorders). We investigated these issues in 47 DPs. No relationship was observed between questionnaire scores and prosopagnosia severity, but males were found to under-report prosopagnosia symptoms relative to females. Thus, we recommend caution in the interpretation of low scores on self-report questionnaires, and suggest that separate norms are developed for male and female participants.

Public Significance Statement

The present study suggests that, in individuals with developmental prosopagnosia (also referred to as "face-blindness"), there is no relationship between objective and subjective face recognition measures. However, males appear to under-report their symptoms relative to females, and we therefore recommend that, as a minimum, separate cut-offs for males and females on self-report measures need to be developed.

Keywords: Face recognition; prosopagnosia; meta-cognition; individual differences
Self-Ratings of Face Recognition Ability are Influenced by Gender but not Prosopagnosia Severity

The last 20 years has seen an abundance of interest in individuals with developmental prosopagnosia (DP), with increasing numbers of adults and children self-referring for clinical screening and/or research participation (e.g. Bate & Bennetts, 2015; Bennettts, Murray, Boyce & Bate, 2017; Bowles et al., 2009; Dalrymple & Palermo, 2015; Geskin & Behrmann, 2017). Larger sample sizes are clearly advantageous for progression of the field, but full objective screening of suspected cases is dependent on time and resources in both research and clinical settings. While the protocols used in DP diagnosis have been considered elsewhere (e.g. Bate & Tree, 2017; Bowles et al., 2009; Dalrymple & Palermo, 2015), perhaps a more contentious debate has evolved from suggestions that self-report instruments may be used to initially short-list candidates for the condition (e.g. Shah, Gaule, Sowden, Bird, & Cook, 2015). This debate has important implications for theories of metacognition – the awareness, understanding and self-assessment of one’s own cognitive processes (Flavell, 1979; Dunlosky & Metcalfe, 2009).

Generally, accurate metacognition is thought to have considerable value because it prompts an individual to seek appropriate learning opportunities (e.g. Finn, 2008; Metcalfe & Finn, 2008; Thiede, Anderson, & Therriault, 2003). In terms of prosopagnosia, if an individual believes that their face recognition skills are poor, they would have heightened awareness of the risk of identification errors in everyday life - including the potential for social or professional repercussions, and even personal safety (Yardley, McDermott, Pisarski, Duchaine & Nakayama, 2008; Dalrymple et al., 2014; Murray, Hills, Bennetts & Bate, 2018). Thus, they may choose to learn more about prosopagnosia and coping techniques, and even seek opportunities to participate in intervention programmes (for an overview, see Bate &
Self-ratings of prosopagnosia severity

Bennetts, 2015). From a researcher’s perspective, accurate self-awareness of face recognition skills would improve the efficiency of recruitment procedures. At present, in most cases, individuals self-refer themselves to researchers before being invited to participate in a full, objective screening session. Whilst these screening sessions can be both costly and time-consuming, self-report methods offer a rapid and cost-effective means of remotely gathering initial screening data from a large number of potential participants.

However, investigations into self-awareness have had varied outcomes, with single-item ratings of face recognition skills mostly producing only weak-to-moderate associations with objective measures of performance (e.g. Bobak, Pampoulou & Bate, 2016; Bowles et al., 2009; Palermo et al., 2017). Some researchers have attempted to develop lists of behavioural traits of DP, placing these items in either a checklist (Murray et al., 2018) or questionnaire (e.g. de Heering & Maurer, 2012; Kennerknecht, Ho & Wong, 2008; Palermo et al., 2017; Shah et al., 2015; Turano & Viggiano, 2016) format. While some questionnaires have continued to offer only low-to-moderate correlations with objective measures of face recognition ability (e.g. Palermo et al., 2017; Turano & Viggiano, 2016), the 20-item Prosopagnosia Index (PI20; Shah et al., 2015) has fared more successfully.

The PI20 is composed of a list of 20 behavioural traits that may be indicative of DP, and participants are required to rate the extent to which they agree or disagree with each item on a five-point scale, resulting in an overall score ranging from 20 to 100. The validity of the PI20 has been investigated in several follow-up studies using typical participants, repeatedly eliciting significant correlations using measures of familiar (Livingston & Shah, 2017; Ventura, Livingston & Shah, 2018) and unfamiliar (Gray, Bird & Cook, 2017; Livingston & Shah, 2017; Ventura et al., 2018) face recognition. Pertinently, data also indicate that candidates for DP tend to achieve higher PI20 scores than typical perceivers (Shah et al.,
Self-ratings of prosopagnosia severity 5

2015), indicating a potential role for the questionnaire in the initial stages of prosopagnosia screening.

What is less clear is whether self-report measures can be used to index the severity of face recognition impairments in people with prosopagnosia. Shah and colleagues (2015) suggested that PI20 scores in the ranges of 64-74, 75-84 and 85-100 are indicative of mild, moderate and severe DP respectively, although a specific analysis considering the relationship between prosopagnosia severity and PI20 scores was not reported. Later reanalysis of the same dataset by Livingston and Shah (2017) reported significant correlations in 18 DPs, using the Cambridge Face Memory Test (CFMT: Duchaine & Nakayama, 2006; \( r = -.62 \)) and a famous face recognition task (\( r = -.61 \)), although this has not yet been confirmed in alternative and larger samples. While these findings suggest the PI20 may be a useful means of determining prosopagnosia severity, replication of the finding in a larger sample is required, alongside the identification of appropriate cut-offs that might indicate different levels of impairment.

A novel question that has not yet been explored is whether participant gender influences the self-report of prosopagnosia symptoms. There is some precedent to suggest this may be the case. In typically developing samples, males tend to rate themselves more favourably than females on traits such as intelligence and attractiveness (e.g. Cooper, Kreig & Brownwell, 2018; Ehrlinger & Dunning, 2003; Sim, Saperia, Brown & Berneiri, 2015). Additionally, females with various developmental conditions appear to rate their symptoms as more severe than males (e.g. autism spectrum conditions (ASC): Lai et al., 2011; attention deficit hyperactivity disorder (ADHD): Vildalen, Brevik, Haavik & Lundervold, 2016). A very recent meta-analysis by Moseley, Hitchiner & Kirkby (2018) further supports the hypothesis that there may be gender differences in self-reported abilities. These authors found that females with autism (\( N = 136 \)) report more prevalent sensorimotor symptoms than males.
Self-ratings of prosopagnosia severity (N = 118). Taken the above evidence together, it may be plausible to hypothesise that females with prosopagnosia will self-report more severe face recognition difficulties than their male counterparts.

The current study addressed these issues in 47 adults with DP. All individuals completed the PI20, and we investigated whether prosopagnosia severity (as measured by performance on the CFMT: Duchaine & Nakayama, 2006) and participant gender influenced scores on the questionnaire.

**Method**

**Participants**

Our sample consisted of 50 adult DPs reported by the authors in a previous publication (Murray et al., 2018). Following existing protocols (e.g. Darylmple & Palermo, 2015) participants were objectively screened using the CFMT (Duchaine & Nakayama, 2006), Cambridge Face Perception Test (CFPT, Duchaine, Yovel & Nakayama, 2007), and a famous faces test that is suitable for adult UK participants (e.g. Bate, Adams, Bennetts & Line, 2017). Our diagnostic inclusion criteria followed the recommendation of Dalrymple and Palermo (2015): performance that differs from the mean score of age-matched controls by at least two standard deviations, at least on any two of the three tests. Full screening data are reported in the original publication.

An invitation to complete the PI20 was sent to these individuals, and 47 (16 male) replied. Their mean age was 53.0 years (range = 25-77 years; SD = 13.3). There was no difference in the average age of the two genders, t(45) = 1.818, p = .076, nor was there a difference in the CFMT scores of males (M = 36.38, SD = 4.38) versus females (M = 36.29, SD = 4.55), t(45) = .061, p = .951. Informed consent was obtained from all participants, and
they participated on a voluntary basis. The study was carried out in accordance with the institution’s Research Ethics Guidelines and was approved by the institution’s Research Ethics Committee.

**Procedure**

Participants were able to complete the PI20 in a variety of ways, according to their own preference. Invitations were sent via email, and 16 individuals returned an electronic copy of their responses in the same manner. Twenty-five participants completed the questionnaire using an online survey platform, and six returned a paper copy via the post or during a visit to the university. A one-way ANOVA showed that the method of completion did not significantly affect PI20 scores ($F(2,44) = .939, p = .399$).

**Results**

A multiple regression was performed to predict PI20 scores from CFMT performance and gender (see Table 1). Because only a subset ($N = 31$) of the DPs had completed the Famous Faces test, we did not enter these data into the analysis in order to retain a higher sample size (no significant correlation was observed between PI20 and famous faces scores in these individuals: $r = .012, p = .949$). The complete model was significant, $F(2,44) = 3.235, p = .049$, explaining 8.9% of the variance (Adj. $R^2 = .089$). CFMT scores were not a significant predictor of PI20 scores ($\beta = -.005, t = .036, p = .972$), and there was no correlation between the two ($r = -.008, p = .956$). However, gender was a significant predictor of PI20 scores ($\beta = .358, t = 2.543, p = .015$). PI20 scores were in the range of 71–91 for male participants, and 67–99 for females (see Figure 1). There was no significant correlation between CFMT and PI20 scores for males ($r = .246, p = .358$) nor for females ($r = -.111, p = .553$), and there was
no difference between these correlations as measured by a Fisher’s r-to-z transformation ($z = 1.08, p = .14$). However, an independent-samples $t$-test indicated that females ($M = 83.26, SD = 6.76$) returned higher PI20 scores than males ($M = 78.13, SD = 5.89$), $t(45) = 2.572, p = .013, d = .81$ (see Figure 1).

< Insert Table 1 and Figure 1 >

Discussion

This investigation found that (a) self-report of prosopagnosia symptoms does not index disorder severity, and (b) that gender influences self-ratings of prosopagnosia symptoms. These findings have important implications for the use of self-report measures in research and clinical practice.

First, while the PI20 offers a rapid and cost-effective method to identify potential DP candidates (the PI20 takes only a few minutes to complete in most cases), objective testing is clearly required to reveal the extent of a person’s face recognition difficulties. These findings differ from previous work (Livingston & Shah, 2017), where the correlation between CFMT and PI20 scores in DPs was more substantial ($N = 18, r = -.62$). It is possible that this discrepancy results from a restriction of range in CFMT and/or PI20 scores in the current sample. Indeed, Livingston and Shah (2017) used more liberal inclusion protocols for DP than the more typically used approach used here (see Burns, Tree & Weidemann, 2014; Duchaine & Nakayama, 2006; Liu, Corrow, Pancaroglu, Duchaine & Barton, 2015; Parketny, Towler & Eimer, 2015; Starrfelt, Klargaard, Petersen & Gerlach, 2018). As such, some of the DPs used by Shah and colleagues differed from controls by, at most, one standard deviation. When the more widely-used protocols for DP diagnosis are applied to their data (i.e. performance that is at least two SDs below the control mean on at least two dominant tests of
face recognition: Dalrymple & Palermo, 2015; measured against norms that are calculated from a sufficiently large age-matched control sample: Bowles et al., 2009), at least six of the individuals would not be classified as DP. When these data points are removed, the PI20 correlations no longer reach significance. Thus, it may be that the previously reported correlations were driven by individuals who would be better categorised as “typical perceivers” – and, interestingly, validity studies investigating the relationship between the PI20 and the CFMT have consistently reported moderate correlations in the typical population (e.g. Gray et al., 2017; Ventura et al., 2018). The PI20 may therefore be sufficiently sensitive to tap face recognition skills across the typical spectrum, but not to gauge severity in people with prosopagnosia.

While this explanation may reflect the wider range of scores that are available on both the PI20 and CFMT for unimpaired perceivers compared to DPs, it may be that the latter have more limited insight into their face recognition difficulties. This hypothesis is supported by qualitative data collected from DP participants (Murray et al., 2018). Unlike those with the acquired form of the condition (e.g. Barton, 2008; Bate et al., 2015), people with DP have no point of comparison nor experience an abrupt loss of their face recognition skills: many individuals tested in our laboratory did not become aware of their difficulties until mid- or even late-adulthood. Further, many DPs navigate daily life fairly well using compensatory strategies, identifying people via voice, gait or general appearance and manner (Fine, 2008; Murray et al., 2018). This may falsely suggest that they are able to recognise others in the same manner as most others in the general population. Face recognition difficulties have also been reported to be highly heritable (e.g. Duchaine, Germine & Nakayama, 2007; Wilmer et al., 2010). In such cases, DPs may compare their abilities to family members who are also poor at recognizing faces and, consequently, may not become aware of their difficulties for a long period of time. These factors potentially explain why some individuals who meet
Self-ratings of prosopagnosia severity

objective criteria for DP return typical scores on self-report measures (e.g. Bowles et al., 2009; Gray et al., 2017; Palermo et al., 2017). Although this did not occur in the current study (all 47 DPs scored above the recommended cut-off on the PI20), it is important to note that they had all received feedback on their face recognition skills (i.e. they were aware that they met the ‘diagnostic’ criteria for DP) prior to completing the questionnaire (for a discussion see Gray et al., 2017; Palermo et al., 2017). Thus, it remains to be seen whether a large sample of “naïve” DPs elicit a significant relationship between objective and subjective measures of prosopagnosia severity.

Second, our finding of gender differences in self-ratings of face recognition skills has important implications for the use of self-report instruments in both clinical and research settings. The findings also fit with those reported for other developmental disorders, where gender effects have also been described for the self-report of ASC and ADHD symptoms (Lai et al., 2011; Moseley et al., 2018; Vildalen et al., 2016). The direction and underpinnings of the effect reported here are unclear: while women over-reported prosopagnosia symptoms compared to men, they also showed no relationship between PI20 and CFMT scores. A higher (albeit non-significant) relationship was observed in men, suggesting they may have more accurate insights into their face recognition skills. While all comparisons between males and females were non-significant here, it is possible that this results from a lack of power, and larger sample sizes are required in future work. Indeed, while a higher correlation is reported for males than females, the sample contained fewer males. In contrast, Shah and colleagues’ sample had more males than females (N = 12 and N = 6 respectively) and so it may be that this partly drove the higher effect size observed in their findings. It should also be investigated whether gender effects are driven by certain behavioural traits (i.e. certain items in the PI20), and across all self-report instruments (e.g. those developed by de Heering & Maurer, 2012; Kennerknecht, et al., 2008; Palermo et al., 2017; Turano & Viggiano, 2016).
Indeed, if some items have greater sensitivity to the traits of male or female DPs, the existing instrument may require some expansion. Until these issues are resolved, different cut-offs should be applied for male and female participants. Further, it remains unknown whether the effect is restricted to those with prosopagnosia, or whether an analogous gender effect can be found across the entire face recognition spectrum (e.g. see metacognitive studies of “super recognition” by Bate et al., 2018; Bobak et al., 2016).

In sum, this investigation finds no support for the notion that self-report measures can index prosopagnosia severity, but reports the important novel finding that the self-report of prosopagnosia traits is influenced by participant gender. Should self-report instruments such as the PI20 become widely adopted by researchers and/or clinicians, future research needs to probe the underpinnings of this gender effect, and, as a minimum, develop separate cut-offs (or even instruments) for use with males and females.

Acknowledgements
SB is supported by a British Academy Mid-Career Fellowship (MD17004).

Disclosure of Interest
The authors report no conflict of interest.
References


Self-ratings of prosopagnosia severity


Self-ratings of prosopagnosia severity


Ventura, P., Livingston, L., & Shah, P. (2018). Adults Have Moderate-To-Good Insight into their Face Recognition Ability: Further Validation of the 20-Item Prosopagnosia


Table 1. Summary of the multiple regression analysis predicting PI20 scores from CFMT performance and gender.

<table>
<thead>
<tr>
<th>Variable</th>
<th>$t$</th>
<th>$b$</th>
<th>$B$</th>
<th>$F$</th>
<th>$R^2$</th>
<th>Adj. $R^2$</th>
<th>$p$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Overall Model</td>
<td>3.235</td>
<td>.128</td>
<td>.089</td>
<td>.049</td>
<td>.927</td>
<td>.015</td>
<td></td>
</tr>
<tr>
<td>CFMT Scores</td>
<td>-.036</td>
<td>-.008</td>
<td>-.005</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gender</td>
<td>2.543</td>
<td>5.132</td>
<td>.358</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
**Figure 1.** Females returned significantly higher PI20 scores than males.