The part-whole effect in super-recognisers and typical-range-ability controls

Elena Belanova, Josh P Davis, & Trevor Thompson

School of Human Sciences, Institute of Lifecourse Development, University of Greenwich, UK

Corresponding author: Dr Elena Belanova, Research Associate, School of Human Sciences, Institute of Lifecourse Development, University of Greenwich, London, SE10 9LS, elena.belanova@greenwich.ac.uk; +44(0)208 331 8859

Dr Josh P Davis, Reader in Applied Psychology, School of Human Sciences, Institute of Lifecourse Development, University of Greenwich, London, SE10 9LS, j.p.davis@greenwich.ac.uk; +44(0)208 331 8859

Dr Trevor Thompson, Associate Professor in Psychology, School of Human Sciences, Institute of Lifecourse Development, University of Greenwich, London, SE10 9LS, t.thompson@greenwich.ac.uk; +44(0)208 331 9632

Acknowledgements

The authors would like to thank Monika Durova, Gesmina Tsourrai, Diandra Bretfelean, Bethan Burnside, and Nikolay Petrov for help with data collection.

Abstract

Face recognition skills are distributed on a continuum, with developmental prosopagnosics and super-recognisers at the bottom and top ends, respectively. Holistic processing propensity is associated with face recognition ability and may be impaired in some developmental prosopagnosics and enhanced in some super-recognisers. Across two experiments we compared holistic processing of 75 super-recognisers and 89 typical-range ability controls using The *Part-Whole Effect* (PWE) paradigm. A subgroup of super-recognisers demonstrated enhanced PWEs in the nose region, suggesting they integrate the nose into the holistic face percept more effectively than controls. Focussed processing of the nose region, an optimal viewing position to extract the holistic properties of faces, has previously been associated with superior face recognition, and this may partly explain the superiority of some super-recognisers. However, a few super-recognisers generated extreme nose region performance patterns in an opposite direction across both experiments, suggesting their superiority is driven by alternative mechanisms. These results support proposals that super-recognition is associated with heterogeneous underlying processes.

Key words: Super-recognisers, Holistic Processing, Part-Whole Effect, Face Recognition, Individual Differences

Word count: 5950

Introduction

Super-Recognisers (SRs) and *Developmental Prosopagnosics* (DPs) inhabit the top and bottom ends respectively of a large quantitative spectrum of individual differences in human face recognition ability (e.g., Bobak, Bennetts, Parris, Jansari, & Bate, 2016; Bobak, Hancock, & Bate, 2016; Russell, Duchaine, & Nakayama, 2009; Tardif et al., 2018; for a review see Noyes, Phillips, & O'Toole, 2017). These differences are influenced by genetics (e.g. Shakeshaft & Plomin, 2015), and personality (e.g. Li, Tian, Fang, Xu, Li, & Liu, 2010), but are mainly face-specific in that they are not related to the recognition of non-face objects (e.g., Bobak, Bennetts, Parris, Jansari, & Bate, 2016), or other cognitive processes (e.g., McCaffery, Robertson, Young, Burton, 2018; Verhallen, Bosten, Goodbourn, Lawrance-Owen, Bargary, & Mollon, 2017; Wilmer et al., 2010; Yovel, Wilmer, & Duchaine, 2014).

The propensity to process faces, but not non-face objects, in a holistic, or whole-face manner (e.g., Diamond & Carey, 1986; Maurer, Le Grand, & Mondloch, 2002; Rossion, 2008; Tanaka & Farah, 1993) may partly be responsible. Human faces convey at least two levels of hierarchical identity information. Face detection, or distinguishing faces from other visual objects, is accomplished via first-order featural information such as that two eyes are placed above a nose and a mouth, respectively. The configuration and spatial relation between facial parts (e.g., the distance between eyes and nose, or nose to mouth) provides second-order information. The integration of these features and their configurations into a unified construct or Gestalt, is what we understand by holistic face processing, which has been viewed as the hallmark of effective face recognition (Maurer et al., 2002; Tanaka & Farah, 1993).

In typical participants (i.e. when SRs or DPs are not explicitly recruited), face recognition ability has been found to correlate with holistic processing propensity (e.g., DeGutis, Wilmer, Mercado, & Cohan, 2013; Richler, Cheung, & Gauthier, 2011; Wang, Li, Fang, Tian, & Liu, 2012; although see Konar, Bennett, & Sekuler, 2010; Horry, Cheong, & Brewer, 2014; Rossion, 2013). Similar effects are not found with non-face objects. For instance, a recent EEG study suggests that good face recognisers might rely more on holistic processing than bad face recognisers (Marzi et al., 2021). Impaired holistic processing also explains impaired face processing in some DPs (e.g., Behrmann & Avidan, 2005, DeGutis, Cohan, Mercado, Wilmer, & Nakayama, 2012; although see Le Grand et al., 2006), and enhanced holistic processing may explain superior face recognition in some SRs (e.g., Bobak, Bennetts et al., 2016; Russell et al., 2009).

The primary aim of the current research was to further investigate the importance of holistic processing in driving the face recognition ability of SRs. The mechanism driving SR's exceptional ability have important theoretical and practical implications. Their elevated skills provide a test bed to examine whether factors associated with face recognition ability in typical-range and prosopagnosic participants, are exhibited at the extreme top end of the spectrum. From an applied perspective, some police forces deploy SRs to operations drawing on their heightened ability to identify suspects (e.g., Davis, Lander, Evans, & Jansari, 2016; Davis, Treml, Forrest, & Jansari, 2018; Robertson, Noyes, Dowsett, Jenkins, & Burton, 2016). Understanding which cognitive processes underlie SRs' skills may assist in the design of reliable recruitment selection tools.

Given that holistic processing could partly explain individual differences in face recognition ability (e.g., DeGutis et al., 2013; Richler et al., 2011; Wang et al., 2012; but see Konar et al., 2010; Horry et al., 2014), enhanced holistic processing might be predicted to be a characteristic of SRs. Indeed, eye tracking studies show that SRs display greater gaze fixations around the nose area (Bobak, Parris, Gregory, Bennetts, & Bate, 2017), associated with a central focus to extract holistic information more effectively (Hsiao & Cottrell, 2008;

Peterson & Eckstein, 2012). However, the limited published research to which small numbers of SRs were recruited, found that only a proportion of SRs display enhanced holistic processing (e.g., Belanova, Davis, & Thompson, 2018; Bobak, Bennetts, et al., 2016; Hendel, Starrfelt, & Gerlach, 2018; Russell et al., 2009). This suggests that rather than only holistic processing, SR's face processing superiority may be driven by heterogeneous underlying processes (e.g., Bate et al., 2018; Noyes et al., 2017).

That being said, these studies employed the *Inversion Effect* (Yin, 1969) and the *Composite Face Effect* (Young, Hellawell, & Hay, 1987) to measure holistic processing in SRs, and both paradigms have recently been criticised for their poor ability to measure individual differences in holistic processing (e.g., DeGutis et al., 2012; DeGutis et al., 2013; Klargaard, Starrfelt, & Gerlach, 2018; Rezlescu, Susilo, Wilmer, & Caramazza, 2017). For instance, the Inversion Effect in which upright faces are easier to process than upside down faces, an effect less pronounced with objects, is typically explained as inversion disrupting holistic face processing (e.g., Rossion, 2008; Valentine, 1988; Yin 1969). However, holistic processing may also drive inverted face processing (Curby et al., 2013; Richler et al., 2011), while individuals with holistic processing impairments do not always display Inversion Effects (e.g., Klargaard, Starrfelt, & Gerlach, 2018). It may not be surprising therefore that the Inversion Effect is enhanced in some, but not all SRs (Belanova et al., 2018; Bobak, Bennetts, et al., 2016; Hendel et al., 2018; Russell et al., 2009).

Unlike the Inversion Effect paradigm, the Composite Face Test (Young et al., 1987) directly disrupts the holistic percept of face stimuli in order to demonstrate the critical role of holistic processing in face perception. With the Composite Face Test participants match (as 'same' or 'different') pairs of top face-halves while ignoring the bottom face-halves. When two face-halves are aligned, holistic processing binds the features and their configurations, so that they appear as a unified whole. However, this effect is also limited in its ability to measure individual differences in holistic processing. For instance, DPs show normal Composite Face Effects (e.g. Le Grand et al. 2006), while Rezlescu, Susilo, Wilmer, and Caramazza (2017) identified a paradox which limits use with individuals with superior face processing skills. Since the Composite Face Effect is reflected in the number of matching errors participants make during aligned versus misaligned trials, participants with better face processing skills will make fewer errors, and thereby display a reduced composite effect, which does not necessarily reflect their reduced reliance on holistic processing. Given these limitations, the present study examined SR's holistic processing using a third holistic processing measure, the *Part-Whole Effect.*

The Part-Whole Effect (PWE) (Tanaka & Farah, 1993; for a review see Tanaka & Simonyi, 2016), illustrates that individual facial features (eyes, nose, mouth) are more easily recognised in the context of the entire face than when presented individually (e.g., DeGutis et al., 2012; 2013; Wang et al., 2012) (see Figure 1). In a standard Part-Whole Test, participants learn a set of target faces, and their memory of individual features (eyes, nose, and mouth) is tested in a two-alternative forced choice paradigm. In the whole condition, participants view these features in a context of the entire face, which helps them identify the target feature easily. In the part condition, however, participants view these features in isolation, making this task more difficult. The Part-Whole Test measures one's ability to integrate parts into a holistic percept while allowing to differentiate holistic processing propensity across different face regions (eyes, nose, mouth), making it a more comprehensive examination of individual differences in holistic processing (DeGutis et al., 2012; 2013).

DeGutis et al. (2012) compared DP's PWE in eyes, nose and mouth regions, and found key differences in holistic processing between DPs and controls. While the nose region generated close to floor performance and similar PWE in DPs and controls, the PWE in the eyes, the most salient face region for identity discrimination (e.g., Bentin, Golland, Flevaris, Robertson, & Moscovitch, 2006) was observed in controls but absent in DPs. On the other hand, both DPs and controls demonstrated a significant PWE in the mouth region, which may explain why some DPs do not differ from controls on the Inversion Effect and Composite Face Effect (e.g., Klargaard et al., 2018; Le Grand et al., 2006), as their holistic processing impairment may be constrained to the eyes region only. DeGutis et al. (2012, see also DeGutis et al., 2013) also used a regression-based method (to isolate parts-based processing from holistic processing and demonstrated that not all DPs display a significant PWE in the mouth region, and more importantly, that greater PWE in the mouth region was associated with greater face recognition (measured by the Cambridge Face Memory Test) in DPs.

No published studies have explored the PWE with SRs. If DPs and SRs occupy the opposite ends of the same face processing ability spectrum, then SRs might be expected to show enhanced PWE in the region of the eyes compared to controls. On the other hand, Bobak et al. (2017) demonstrated that four out of eight SRs spend more time fixating the nose area, which has been associated with a central focus to extract holistic information more effectively (Hsiao & Cottrell, 2008; Peterson & Eckstein, 2012). Therefore, it is possible that SRs' enhanced attention to this area may contribute to their face recognition superiority (Bobak et al., 2017). Thus, SRs might also be expected to demonstrate enhanced PWE in the nose region.

The present study

This paper describes two experiments investigating holistic processing in SRs and typical-ability controls using the PWE drawing on the regression-based technique described by DeGutis et al. (2012). Experiment 1 examined the PWE with an online sample, Experiment 2, in the laboratory. Previous research (e.g., DeGutis et al., 2012; 2013) has found correlations between the PWE and face recognition ability as measured by the Cambridge Face Memory Test (CFMT) (Duchaine & Nakayama, 2006). Similar effects were predicted for the current research across the entire participant sample. More specifically, it was hypothesised that compared to typical-ability controls, SRs would demonstrate stronger PWE effects in the eyes and nose, but not the mouth region.

Experiment 1

Research investigating super-recognition tends to attract participants with far better face recognition abilities than is standard in the population (e.g., Davis, Bretfelean, Belanova, & Thompson, 2020; Satchell, Davis, Julle-Danière, Tupper, & Marshman, 2019). A similar recruitment bias was expected here, and to reduce its effects following precedent, a three-component strategy was employed in both experiments.

With the inclusion of all participants, the first component examined correlations between face recognition and the regressed PWE in the eyes, nose, and mouth regions. The second component compared performances of SRs and controls scoring within the 'typical ability' range. For this, consistent with previous research, minimum criteria for SR was a score on the extended version of the CFMT (CFMT+; Russell *et al*., 2009) expected to be achieved by about 2% of the population (i.e. 2 SD (≥ 95 out of 102) above the mean scores (M = 70.7) achieved by a representative UK sample (Bobak, Pampoulov et al., 2016)). As many participants scored slightly below this threshold, a CFMT+ score between 58–83 was designated as typical-range ability control criteria (i.e., within 1 SD of Bobak, Pampoulov et al.'s (2016) control mean).

Some researchers (e.g., Bate et al., 2018; Robertson, Black, Chamberlain, Megreya, & Davis, 2020) propose that using two face processing tests for SR group classification is more reliable than using a single test such as the CFMT+. Indeed, some CMFT+ scorers meeting SR criteria perform poorly on alternative face processing tests (e.g., Belanova et al., 2018; Davis et al., 2020). Therefore, a second test to verify SR and Control group ability was employed. In Experiment 1, this was the *Glasgow Face Matching Test* (GFMT: Burton et al., 2010). The GFMT is a measure of simultaneous face matching, and has been used as a second SR and typical-ability control verifying test previously (Correll, Ma., & Davis, 2021; Davis, Bretfelean, Belanova, & Thompson, 2020; Satchell et al., 2019). Despite one being a memory test, the other a perception test, scores on the CFMT+ and the GFMT typically have a medium correlation of about $r = 0.50$, indicative of 25% of the variance driving individual differences in face recognition ability (e.g., Verhallen et al., 2017). This leaves 75% unaccounted for. Nevertheless, the GFMT suffers ceiling effects (e.g., see Davis et al., 2016), and therefore a different ability-verifying test, a *Sequential Face Matching Test (SFMT),* replaced the GFMT in Experiment 2.

Nevertheless, excluding additional participants who may fractionally fall beneath two-test verification thresholds, inevitably reduces statistical power. For this reason, we ran two sets of between-group analyses. The first analyses used CFMT+ scores alone to define SRs and controls. Outcomes are reported in the main Results section. Additional between-group analyses, in which the second tests described above were used to verify SR and Control group ability are reported in the supplementary materials. Despite the reduced power, outcomes mostly matched the main analyses. Any differences are evaluated in the General Discussion.

The third component compared CFMT+ defined SR's individual PWE performances with the mean performances of typical-ability controls using an exploratory comparison of z-scores similar to the modified t-tests for single cases described by Crawford and Howell (1998). This component was key, given that previous research has suggested heterogenous mechanisms underly super-recognition (e.g., Noyes, Hill, & O'Toole, 2018). Individual SR propensity to rely on holistic processing of the configurations associated with the different face-parts might be obscured by between-group analyses. Typically, significant outlier performances on face recognition tests are arbitrarily classified as such if they fall beyond 2 SD above or below the control mean. However, residual effects generated by the PWE were expected to be relatively small. Therefore, so as to not miss borderline cases, a more liberal cut off of 1.5 SD was applied here (for a discussion of this point see Bate, Bennetts et al., 2018; Bate, Frowd et al., 2018).

Design

This research received ethical approval from the University of Greenwich. A correlational design examined the relationship between face recognition scores and performances on the PWE in eyes, nose, and mouth regions evaluated using regression-based measures. Independent-measures design compared a sub-set of SRs and typical-ability controls on all outcomes. Finally, individual analyses compared each SRs' residuals ('part' trials regressed from 'whole' trials, reflecting holistic processing propensity) to the control group mean performances.

Materials

Cambridge Face Memory Test (extended) (CFMT+) (Russell *et al*., 2009): The test commonly used to define SR ability comprises 102 trials; and is the extended version of the original 72-trial CFMT (Duchaine & Nakayama, 2006). In the learning stage, participants are required to memorise six faces from their internal features. In the recognition stage, participants are required to recognise the six target faces from different viewpoints in different lighting conditions in a three-alternative forced-choice paradigm. With each block

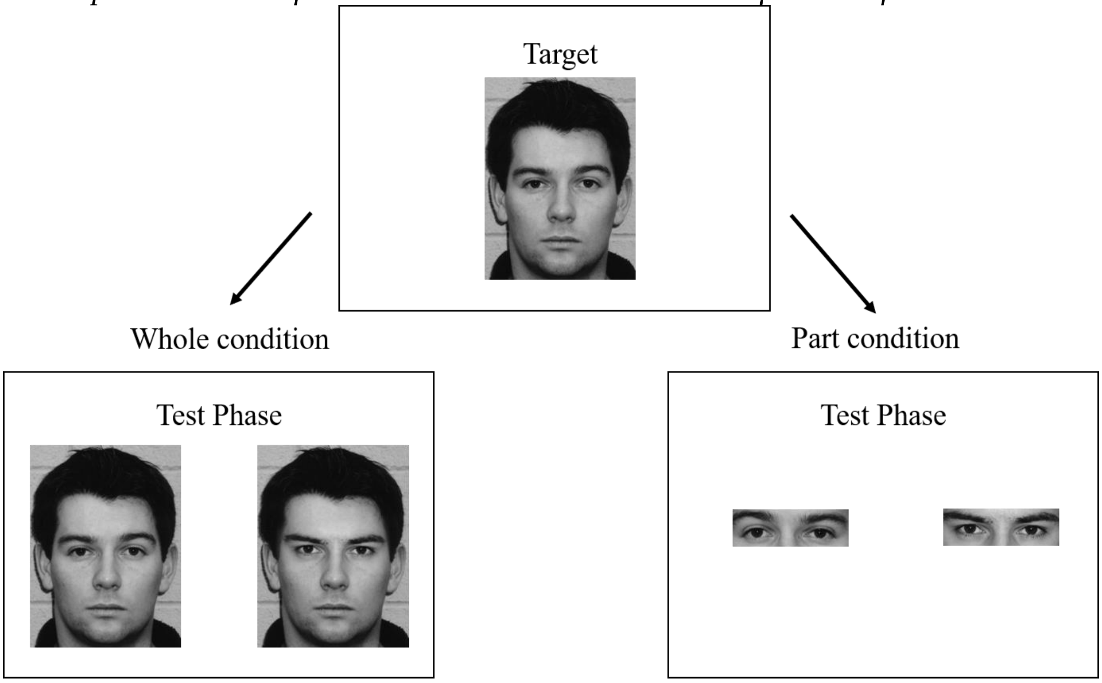
(4 in total) viewing conditions become more difficult: visual noise; facial expressions, and regularly repeating distractor faces increase difficulty.

Glasgow Face Matching Test (GFMT; Burton et al., 2010): was used as the ability-verifying test and is described in the Supplementary Materials (SM_01).

Part-Whole Test: The White-Caucasian adult face stimuli used for this test were extracted from the database of The Park Aging Mind Laboratory at The University of Texas at Dallas (Minear & Park, 2004). Adobe Photoshop was used to create 24 new target faces (50% female) by substituting original internal features (eyes, nose, mouth) with features from different images. The visual angle of the face presentation was 3.8° x 6.7°.

Figure 1

An example of the Part-Whole Face Test (eyes) trial used in the current research. In the Learning Phase, participants view the whole face for 1 second. The two-alternative forced choice Test Phase follows immediately, asking participants to identify the face (whole condition) or the individual feature (part condition – in this case the eyes) that corresponds to the target image seen earlier. Note - the nose or mouth replace the eyes when appropriate. Participants are not informed in advance whether a whole face or a feature will be displayed.



As shown in Figure 1, participants viewed a target face and then were asked to sequentially match each target face to two probe images, whereby only one of the probe images matched the target image. The PWE was administered under two conditions, *whole* and *part*. The whole condition comprised 72 trials and involved the whole target face being matched to two probe whole faces. The probe images were virtually identical but differed in one feature only (eyes, nose, or mouth). The part condition comprised 72 trials and involved the whole target face being matched to two probe features (e.g., two sets of eyes, two noses, or two mouths). To ensure that participants are not informed in advance whether a whole face or a feature will be displayed during the Test phase, the two conditions (whole and part) were administered in a random manner within the same blocks.

Each trial (144 in total) began with a target image (1000msec) in the centre of the screen, followed by a two-alternative image design in which two faces or two sets of one feature matched the target image. Participants selected the image that corresponded to the target image by clicking the cursor on the image of their choice. Participants were unaware of which condition (whole or part), or feature (eyes, nose, mouth) of each upcoming trial would be tested. The probe images remained on the screen until a response was made.

To calculate the PWE for correlational and individual analyses, part trials were regressed from the whole trials (see DeGutis et al., 2013).

Participants

A sample of 278 participants who had previously completed the CFMT+ (M = 87.94, SD = 8.89) and the GFMT (M = 37.57, SD = 2.32) in unpublished research (mean age = 39.9, SD = 12.5, male = 83, female = 195, White ethnicity = 249 (89.56%)) finished the study. As with previous research (e.g. Davis et al., 2020; Robertson et al., 2020), SRs were selected based on CFMT+ scores (95+ out of 102), 2 SD above the mean of Bobak, Pampoulov et al.’s (2016) representative sample (M = 70.7; SD = 12.3). Because a high proportion of the sample scored slightly below this threshold, following precedent (e.g. Correll et al., 2020; Davis et al., 2020), participants scoring within 1 SD of that mean (58-83 out of 102) were selected as typical-ability controls. Participants scoring 84-94 or 57 and below were excluded from between-groups and individual level analyses. Fifty-five SRs and 64 controls were identified based on these CFMT+ criteria. Note that despite efforts to select a typical-ability sample, controls still significantly outperformed Bobak, Pampoulov et al.’s (2016) mean on the CFMT+, $t(63) = 7.06$, $p < .001$, Cohen’s $d = 0.57$. Table 1 displays SR and control group demographic information and mean CFMT+ and GFMT performances. Independent-measures t-tests demonstrated significant differences with strong effect sizes, whereby SRs outperformed controls on both tests.

Table 1
Criteria for SR and control groups, and results for t-tests comparing their CFMT+ outcomes and demographics in Experiment 1

	SRs (n = 55)	Controls (n = 64)				
	Group inclusion criteria					
CFMT+ (out of 102)	95-102	58-83				
	M (SD)	M (SD)	t	df	p	d
CFMT+	97.53 (1.97)	76.28 (6.33)	25.46	76.88	<.001	4.53
GFMT	38.24 (2.70)	36.44 (2.53)	3.75	117	<.001	0.69
Age	36.33 (10.70)	44.70 (13.93)	-3.70	115.61	<.001	0.67
			χ^2		p	V
White Ethnicity	48 (87.3%)	58 (90.6%)	0.34	1	>.2	0.05
Gender	Male = 14 Female = 41	Male = 18 Female = 46	0.11	1	>.2	0.03

Procedure

After providing informed consent and demographic information, participants were administered the Part-Whole Test. Participants were debriefed at the end.

Results

Reliability and Correlation analyses: Table 2 displays reliability (Guttman's λ2) for the Part-Whole Test conditions and the PWE residuals, as well as their correlations with the CFMT+. Descriptively, the strongest correlation between the CFMT+ and the residuals was that of the total PWE, the lowest the eyes, although all correlations were small, positive and significant.

Table 2
PWE Reliability and correlations with CFMT+ (Experiment 1)

	Reliability λ2	CFMT+ Correlation
Total Part-Whole Test		
Whole trials	.80	**.45***
Part trials	.70	**.36***
PWE Residuals	.53	**.26***
Eyes		
Whole	.71	**.34***
Part	.55	**.31***
PWE Residuals	.50	**.15***
Nose		
Whole	.70	**.33***
Part	.48	**.24***
PWE Residuals	.50	**.23***
Mouth		
Whole	.65	**.33***
Part	.52	**.26***
PWE Residuals	.48	**.22***

**All correlations are significant ($p < .05$).*

Between SR and control group analyses: Three 2 (condition: Whole, Part) x 2 (group: SRs; Controls) ANOVAs also compared SR's PWE to that of controls (See Table 3).

Table 3

ANOVAs examining PWE in SRs (n = 55) and Controls (n = 64) in Experiment 1

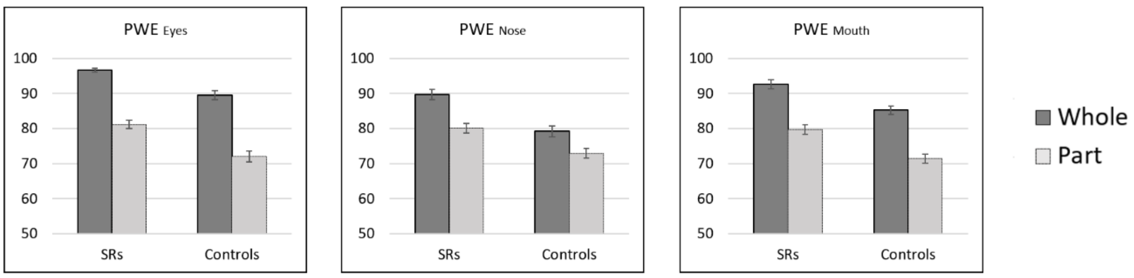
df (1, 117)	PWE $_{Eyes}$ *F*	η^2	PWE $_{Nose}$ *F*	η^2	PWE $_{Mouth}$ *F*	η^2
Main effects						
Group	28.75 *	.197	27.39 *	.190	25.72 *	.180
Condition	322.30 *	.734	47.65 *	.289	155.49 *	.571
Interaction	1.09	.009	2.15	.018	<1	.002

$p < .050$ *

With strong effect sizes, all three ANOVAs revealed significant main effects of group, with superior performances recorded for SRs than controls; and significant main effects of condition, with superior performances recorded in the whole than in the part conditions. Of critical importance, however, are the two-way interactions, as significant effects would reflect group differences in the strength of the PWE between SRs and controls. While the impact of the PWE in SRs appears slightly greater in all regions than controls (see raw effect sizes reported in Figure 2), none of the two-way interactions were significant. Associated with very small effect sizes, these group-level results imply no reliable differences in the strength of the PWE between SRs and controls.

Figure 2

PWE in SRs (n = 55) and controls (n = 64) in Eyes, Nose, and Mouth regions. Effect sizes (Cohen's d) are displayed for SRs and Controls under each plot. Error bars = standard errors of the mean.



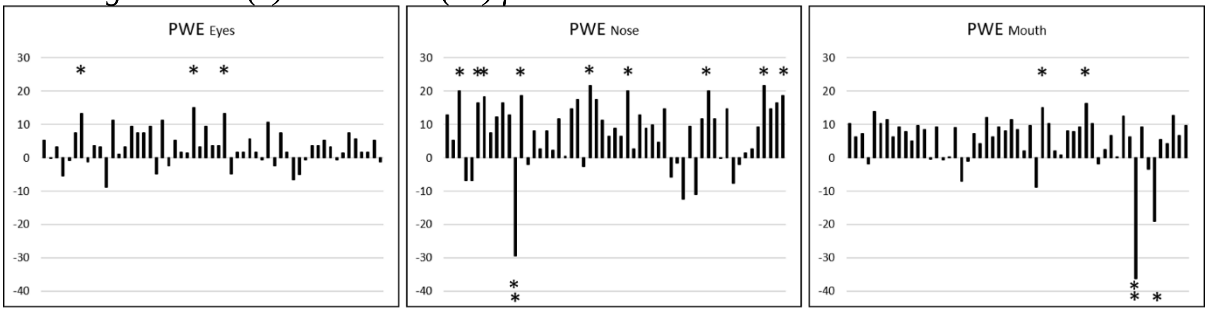
SRs $d = 1.90$; Controls $d = 1.64$ SRs $d = .82$; Controls $d = .48$ SRs $d = 1.22$; Controls $d = 1.11$

Additional planned paired comparisons of whole *vs* part trials for each group separately, were designed to generate an interpretable effect size measure (Cohen's d). SRs and controls displayed a significant PWE in all face regions ($p < .001$), with slightly stronger effect sizes observed for SRs (see Figure 2).

Individual level analyses: Individual residual scores for PWE for the Eyes, Nose and Mouth regions were computed for SRs as the deviation from the control group regression line. The SR residual scores that deviated from the control mean ($n = 64$) by at least 1.5 SD were treated as extreme scores. As with CFMT+ scores for SRs described in the participants section, significant outlier performances on face recognition tests are often arbitrarily classified as such if more than 2 SD above or below the control mean. However, residual effects generated by the PWE will always be relatively small (see General Discussion). Therefore, so as to not miss borderline cases, a more liberal criterion for judging extreme residual scorers was applied to the PWE (for a discussion of this point see Bate, Bennetts et al., 2018; Bate, Frowd et al., 2018). PWE residuals that were 1.5 SD above or below control means are marked (*) in Figure 3 and were treated as severe outliers here. Full results of individual analyses including the upper and lower bound confidence intervals (95%) of the estimated proportion of the population expected to fall below each SR (Crawford, Garthwaite, & Porter, 2010) are reported in the Supplementary Materials (SM_02). As can be seen from Figure 3, most SRs demonstrated residual effects above the control mean in all three part-face conditions, although extreme positive scores in the PWE nose comparison ($n = 9$), outnumbered those in the PWE eye ($n = 3$), and PWE mouth ($n = 2$) conditions, consistent with expectations. Nevertheless, one SR in the PWE nose, and two in the PWE mouth conditions demonstrated extreme results in the opposite unexpected negative direction.

Figure 3

Individual residuals reflecting PWE in the Eyes, Nose, and Mouth regions in SRs. Scores deviating >1.5 SD (*) and >2 SD (**) from the control mean are marked.



Discussion

Experiment 1 was designed to investigate the relationship between face recognition scores and PWE in participants with superior face recognition ability. CFMT+ scores positively correlated with the PWE in all face regions, suggesting that consistent with previous research, PWE outcomes are associated with individual differences in face recognition ability (DeGutis et al., 2013). CFMT+ defined SRs demonstrated stronger effect sizes for PWE in all face regions, although between-group comparisons revealed no statistically significant differences.

Using liberal criteria for outlier detection, individual analyses, on the other hand, revealed that a small subgroup of SRs (16.4%) generated a substantially greater PWE $_{\text{Nose}}$, as they were at least 1.5 SD above the control mean. These findings are in line with Bobak et al.'s (2017) observation that four out of eight SRs fixated the nose area to a greater extent than controls. The nose area could be the most convenient fixation point for a more effective holistic extraction of information (e.g., Hsiao & Cottrell, 2008; Peterson & Eckstein, 2012), and some SRs appear to integrate the nose region into the holistic percept more effectively than controls.

However, as found in previous research evaluating holistic processing in SRs (e.g. Bobak, Bennetts et al., 2016), some SRs displayed a heterogenous pattern of results, with only 25.5% SRs showing an enhanced PWE across all face regions (i.e. above the mean performance of controls), while 3 SRs (5.5%) showed an extreme effect in the opposite direction. Not all SRs' exceptional face processing skills can therefore be attributed to their PWE, meaning other factors must contribute or explain their superiority.

Given the heterogeneity of SR's responses as indicated by the different and sometimes opposite patterns observed in the individual analyses in Figure 3, it is not surprising that these effects were not captured by the between-group comparisons of CMFT+ defined SRs and controls (Table 2). This further emphasises the importance of including individual analyses in examination of SR's performance (Noyes et al., 2018) as their superiority in face processing appears to rely on different strategies or mechanisms.

As this was the first investigation of the PWE in SRs, Experiment 2 was designed to examine the PWE in a new participant sample in an attempt to replicate the results and to clarify whether the heterogeneity of the results in Experiment 1 reflects SRs' different strategies or, if other unknown factors contributed to these heterogenous responses and lack of group differences. For instance, the lack of control over conditions when experiments are conducted via the internet could partly contribute to the heterogenous responses observed in Experiment 1. It is possible that online participants in Experiment 1 were distracted, or perhaps not as motivated as participants who sign up to laboratory-based studies. Experiment 2 therefore replicated Experiment 1 by examining the PWE in a new sample of SRs invited to our lab.

Experiment 2

Experiment 2 was designed to replicate Experiment 1 but in the more controlled setting of a laboratory. A different ability-verifying test, a *Sequential Face Matching Test (SFMT),* replaced the GFMT, which may suffer ceiling effects (e.g., see Davis et al., 2016) to support the CFMT+ defined group classifications described in the supplementary data. The hypotheses matched those for Experiment 1.

Method

Design and Materials

Experiment 2 closely replicated Experiment 1's design except that tests were administered in the laboratory, and the SFMT replaced the GFMT as a second face recognition ability verifying test.

Sequential Face Matching Test (SFMT). This test was used to verify face recognition ability in Experiment 2 and is described in the Supplementary Materials (SM_03).

Part-Whole Test (PWE): The lab-based version of the PWE was identical to the online version described in Experiment 1. However, the test was presented in PsychoPy2 (Pierce, 2007) and participants responses were made on the keyboard ('1' for the image on the left, and '0' for the image on the right).

Participants

All participants provided informed consent as well as permission to access face recognition test scores from previous unpublished research conducted by our team. Forty-nine white-Caucasian participants (CFMT+ M = 85.39, SD = 11.97, age M = 34.23, SD = 10.39, males = 18, females = 30)[1] contributed to Experiment 2.

[1] Age (n = 2) and gender (n = 1) information were missing for some participants

Table 4
Criteria for SR and control groups, and results for t-tests comparing their CFMT+, Sequential Face Matching Test (SFMT) outcomes, and demographics in Experiment 2

	SRs (n = 20)	Controls (n = 25)				
	SRs and controls group inclusion criteria					
CFMT+ *(out of 102)*	95 - 102	59 - 83				
	M (SD)	*M* (SD)	*t*	*df*	*p*	*d*
CFMT+	97.25 (2.12)	74.52 (5.52)	17.39	32.32	<.001	5.43
SFMT	79.05 (9.56)	71.93 (7.25)	2.56	36	.015	0.84
Age	35.80 (1.68)	30.91 (11.39)	1.68	38.35	.101	0.51
			χ^2		*p*	*V*
Gender	Male = 11 Female = 9	Male = 6 Female = 18	4.14	1	.042	.307

[A] *Note that SFMT data was not available for some control participants due to experimenter error.*

The same criteria as in Experiment 1 was used to select SRs and controls. Overall, 20 SRs and 25 controls were identified based on these CFMT+ criteria. Controls once again significantly outperformed Bobak, Pampoulov et al.'s (2016) mean on the CFMT+, $t(24) = 3.46$, $p = .002$, Cohen's $d = 0.40$. Table 4 displays SR and control group demographic information and mean CFMT+ and Sequential Face Matching Test performances. Independent-measures t-tests demonstrated significant differences with strong effect sizes, whereby SRs outperformed controls on both tests.

Procedure

After providing informed consent, participants completed the SFMT and the PWE in a laboratory. The experiment took approximately 20 minutes, after which participants were debriefed.

Results

Reliability and Correlation analyses: Table 5 displays reliability (Guttman's λ2) of the Part-Whole Test conditions and PWE residuals as well as their correlations with the CFMT+. CFMT+ scores showed a significant moderate positive relationship with holistic processing in the eyes and nose regions, as indicated by PWE_{Eyes} and PWE_{Nose} residuals. However, no significant relationship was found with the mouth region, which may be a consequence of reduced power compared to Experiment 1.

Table 5
Reliability and correlations between CFMT+ and PWE in Experiment 1 (n = 49)

	Reliability λ2	CFMT+ Correlation
Total Part-Whole Test		
Whole trials	.83	**.62***
Part trials	.73	**.47***
PWE Residuals	.66	**.44***
Eyes		
Whole	.80	**.42***
Part	.57	.27_
PWE Residuals	.78	**.33***
Nose		
Whole	.70	**.48***
Part	.48	.22_
PWE Residuals	.69	**.46***
Mouth		
Whole	.65	**.50***
Part	.57	**.49***
PWE Residuals	.36	.25_

**Correlations are significant at p <.05 level*

Between-Group analyses:

Three 2 (condition) x 2 (group: SRs; controls) ANOVAs also compared SR's PWE in each part condition to that of controls (see Table 6). With strong effect sizes, all three ANOVAs revealed significant main effects of group, with SRs displaying stronger performances than controls; and significant main effects of condition, with superior performances recorded in the whole than in the part conditions. As with Experiment 1, the critical two-way interactions were not significant, implying no group differences in the strength of PWE within each face region between SRs and controls.

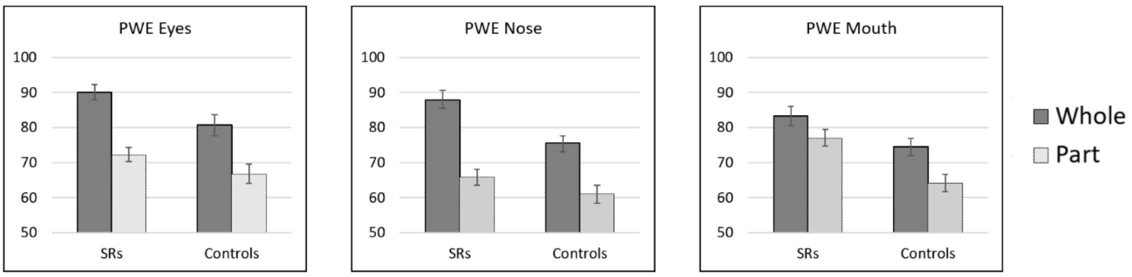
Table 6.
ANOVAs examining PWE in SRs (n = 20) and Controls (n = 25) in Experiment 2

df	PWE $_{Eyes:}$		PWE $_{Nose:}$		PWE $_{Mouth:}$	
(1, 43)	*F*	η^2	*F*	η^2	*F*	η^2
Main Effects						
Group	5.21 *	.108	10.83 *	.201	13.07 *	.233
Condition	77.52 *	.643	64.26 *	.599	16.63 *	.279
Interaction	1.17	.026	2.62	.057	1.01	.023

$p < .050$ *

As with Experiment 1, planned paired comparisons of whole *vs* part trials for each group separately, were designed to generate an interpretable effect size measure (d). These showed that controls demonstrated a significant PWE in all face regions. SRs on the other hand displayed a significant PWE in the eyes and nose regions only, both with greater effect sizes than controls. SRs did not show a significant PWE in the mouth region (see Figure 4).

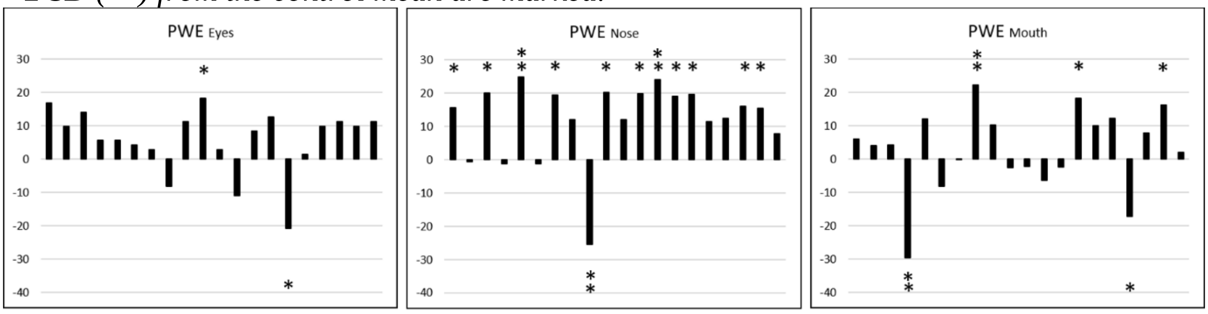
Figure 4.
*PWE in SRs and controls in Eyes, Nose and Mouth regions. Error bars = standard error of the mean. Significant within group differences are marked * and effect sizes are indicated below each figure.*



*SRs $d = 1.64$; *Controls $d = 1.09$ *SRs $d = 1.58$; *Controls $d = .91$ SRs $d = .43$; *Controls $d = .82$

Individual level analyses: Individual residual scores for PWE for the Eyes, Nose and Mouth regions were computed for SRs ($n = 20$) as the deviation from the control regression line. SR's scores were compared against the control mean. As with Experiment 1, PWE residuals 1.5 SD above or below control means were deemed as severe outliers relative to typical performance and are marked (*) in Figure 5. Full results of individual analyses are reported in the Supplementary Materials (SM_04). As can be seen from Figure 5, most SRs demonstrated residual effects above the control mean in all three conditions, although the numbers of extreme positive effects in the PWE nose comparison (*n* = 11), outnumbered those in the eye (*n* = 2), and mouth (*n* = 3) conditions, consistent with expectations and Experiment 1. Nevertheless, one SR in the PWE nose, PWE eye, and two in the PWE mouth conditions demonstrated extreme results in the opposite unexpected negative direction.

Figure 5.
Individual residuals reflecting PWE in SRs and controls. Scores deviating >1.5 SD (*) and >2 SD (**) from the control mean are marked.



Discussion

Consistent with Experiment 1, Experiment 2 showed a significant small-to-moderate positive relationship between CFMT+ scores and the PWE in the eyes and nose regions, but not the mouth region. While the reduced power potentially resulted in the non-significant correlation between CFMT+ and PWE $_{\text{Mouth}}$, the coefficient strengths were virtually identical across the two experiments, suggesting that the PWE in all face regions is associated with face recognition ability (see also DeGutis et al., 2013).

Critically, in line with predictions and correlational findings, effect sizes demonstrated that compared to controls, SRs generated stronger PWE in the eyes and nose regions, although between-group comparison were again non-significant. Individual analyses revealed again that the nose region seems to be processed more effectively by a subgroup of SRs (55%). Few other SRs showed enhanced PWE in other face regions, while 4 SRs showed an extremely reduced PWE (eyes, *n* = 1; nose, *n* = 1; mouth, *n* = 2). While uncontrolled factors could contribute to the participants' performances (e.g., distractions, fatigue, or motivation), the results nevertheless highlight that heterogenous factors may underlie super-recognition.

General Discussion

The present study reports two experiments investigating the relationship between face recognition ability and holistic processing as assessed by the Part-Whole Effect (PWE). The results suggest a subgroup of SRs may rely on holistic processing as assessed by the PWE to a greater extent than controls, although this superiority may be limited to a specific face region. As a group, SRs displayed greater effect sizes for PWE $_{\text{Eyes and Nose}}$ (Experiments 1 and 2) and PWE $_{\text{Mouth}}$ (Experiment 1 only), suggesting that they indeed benefit from seeing the whole face when identifying individual features. However, these differences were mainly not significant with between-groups comparisons. Instead, individual analyses on SRs in which scores are compared to the control group mean provides greater understanding of these effects (Noyes et al., 2018).

Exploratory individual analyses across both experiments (Figures 3 and 5) showed that most SRs exceeded control means, while 30 out of 75 SRs (40.0%) displayed an enhanced PWE, with scores exceeding the mean by at least 1.5 SD. Using a more liberal criterion of 1.5 SD (as opposed to the commonly employed 2 SD) to detect severe outliers may be more informative, especially when trying to detect extreme performance on such bespoke tasks as Part-Whole Test, where varying stimuli may not all generate an equal effect. For example, some face parts may be more distinctive than others, while some whole faces may interact differently with their face parts. Increasing the number of trials could potentially strengthen effect sizes and offset this natural variability, but this could increase participant fatigue and decrease their concentration. Accordingly, the cut off of 1.5 SD has been used by other studies investigating individual SR scores (e.g., Bate, Bennetts et al., 2018; Bate, Frowd et al., 2018; Jenkins et al., 2021), so as not to miss informative extreme scores. Indeed, the more stringent cut off of 2 SD has resulted in very few significant effects in the current study (see Figures 3 and 5).

Importantly, consistent with our hypothesis, the vast majority of the extreme individual effects observed in both experiments, were restricted to the nose region. Therefore, in line with previous research suggesting that some SRs spend more time focussing on the nose region (Bobak et al., 2017), SRs in this study appeared to process this face region more effectively. Previous research has highlighted that the nose area is the optimal viewing position in face processing (Hsiao & Cottrell, 2008; Peterson & Eckstein, 2012), potentially because it allows for a more efficient spread of attention across the face. Indeed, Bobak et al. (2017) found a positive association between fixations on the nose and face recognition ability but could not conclude that this increased spatial attention to the nose region also stood for enhanced holistic processing. Our study therefore adds to this earlier research and demonstrates that SRs indeed benefit from a more enhanced holistic processing of the nose region, which could contribute to their exceptional face recognition skills.

It is also noteworthy that a small number of SRs (9.3%) show strong opposite patterns, which could potentially account for the lack of significant group differences reported in both experiments. The reduced holistic processing in these SRs could feasibly suggest that their superior face processing skills are the result of enhanced part-based processing, though this is yet to be directly tested.

On the other hand, approximately half of the SRs generated a PWE in the control range. As such, although holistic processing may be a strong element in the superior skills of some SRs, other factors may explain the superiority of the remainder. While this heterogeneity may not be entirely due to the different holistic processing mechanisms employed by participants (as other unrelated factors, such as fatigue, may also be involved), the results suggest that the superiority of not all SRs is based on holistic processing.

Consistent with previous research (e.g., DeGutis et al., 2013), the PWE correlated with face recognition scores in both experiments, though some of these correlations were weak. Newly developed holistic tests that have been designed to better capture individual differences in holistic processing may provide a solution (e.g., *Vanderbilt Holistic Processing Test*, Richler et al., 2014). On the other hand, researchers also propose that holistic processing should not be expected to strongly correlate with face recognition ability as other factors (e.g. parts-based processing) contribute to effective face processing (e.g., Degutis et al., 2013).

Finally, it should be noted that using stricter criteria (i.e., using two tests) to classify participants as SRs and controls generated almost identical between-group findings as when using only the CFMT+ for group classification, with one exception (see SM_01). The most prominent pattern of results across both experiments was that a significant number of CFMT+ only defined SRs generated a greater PWE in the nose area. Interestingly, while these individual case observations were not detected in the CFMT+ defined group analyses (Experiments 1 and 2), they were observed in the CFMT+ and GFMT defined group analyses (Experiment 1). Indeed, CFMT+ and GFMT defined SRs generated a significantly greater PWE $_{Nose}$ compared to controls. On the other hand, no between-group differences in PWE were observed in Experiment 2, regardless of group classification. It is possible that the GFMT (Experiment 1) was a more reliable test to verify ability than SFMT (Experiment 2), but this was not directly tested here.

In conclusion, this was the first study to examine SRs on one of the traditional holistic processing measures, the Part-Whole Effect, proposing that for some SRs exceptional face recognition skills may be associated with more effective processing of the nose area, whereby they integrate the nose into the holistic face percept more effectively than controls. However, given the heterogenous nature of the individual results, it appears other factors drive exceptional recognition ability in SRs with typical holistic processing.

References

Bate, S., Bennetts, R., Hasshim, N., Portch, E., Murray, E. Burns, E., Dudfield, G. (2018). The limits of super recognition: An other-ethnicity effect in individuals with extraordinary face recognition skills. *Journal of Experimental Psychology: Human Perception and Performance, 45*(3), 363-377. doi:10.1037/xhp0000607

Bate, S., Frowd, C., Bennetts, R., Hasshim, N., Murray, E., Bobak, A. K., Wills, H., & Richards, S. (2018). Applied screening tests for the detection of superior face recognition. *Cognitive Research: Principles and Implications, 3*(22), 1-19. doi:10.1186/s41235-018-0116-5

Behrmann, M., & Avidan, G. (2005). Congenital prosopagnosia: Face-blind from birth. *Trends in Cognitive Sciences, 9*(4), 180-187. doi:10.1016/j.tics.2005.02.011

Belanova, E., Davis, J. P., & Thompson, T. (2018). Cognitive and neural markers of super-recognisers' face processing superiority and enhanced cross-age effect. *Cortex, 108,* 92-111. doi:10.1016/j.cortex.2018.07.008

Bentin, S., Golland, Y., Flevaris, A., Robertson, L. C., & Moscovitch, M. (2006). Processing the trees and the forest during initial stages of face perception: electrophysiological evidence. *Journal of Cognitive Neuroscience, 18*(8), 1406-1421. doi: 10.1162/jocn.2006.18.8.1406

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. doi:10.1016/j.cortex.2016.05.003

Bobak, A. K., Hancock, P. J., & Bate, S. (2016). Super-recognisers in action: Evidence from face-matching and face memory tasks. *Applied Cognitive Psychology, 30*(1), 81-91. doi:10.1002/acp.3170

Bobak, A. K., Pampoulov, P., & Bate, S. (2016). Detecting superior face recognition skills in a large sample of young British adults. *Frontiers in Psychology, 7*(1378), 1-11. doi:10.3389/fpsyg.2016.01378

Bobak, A. K., Parris, B. A., Gregory, N. J., Bennetts, R. J., & Bate, S. (2017). Eye-movement strategies in developmental prosopagnosia and "super" face recognition. *The Quarterly Journal of Experimental Psychology, 70*(2), 201-217. doi:10.1080/17470218.2016.1161059

Burton, A. M., White, D., & McNeill, A. (2010). The Glasgow face matching test. *Behavior Research Methods, 42,* 286–291. doi:10.3758/BRM.42.1.286.

Correll, J., Ma., D. S., & Davis, J. P. (2021). Perceptual tuning through contact? Contact interacts with perceptual (not memory-based) face-processing ability to predict cross-race recognition. *Journal of Experimental Social Psychology, 92,* 104058, https://doi.org/10.1016/j.jesp.2020.104058

Crawford, J. R., Garthwaite, P. H., & Porter, S. (2010). Point and interval estimates of effect sizes for the case-controls design in neuropsychology: Rationale, methods, implementations, and proposed reporting standards. *Cognitive Neuropsychology*, *27*, 245-260. doi:10.1080/02643294.2010.513967

Crawford, J.R., & Howell, D.C. (1998). Comparing an individual's test score against norms derived from small samples. *The Clinical Neuropsychologist, 12,* 482-486. DOI: 10.1076/clin.12.4.482.7241

Davis, J. P., Bretfelean, D., Belanova, E., & Thompson, T. (2020). Super-recognisers: face recognition performance after variable delay intervals. *Applied Cognitive Psychology*, DOI:10.1002/acp.3712

Davis, J. P., Lander, K., Evans, R., & Jansari, A. (2016). Investigating predictors of superior face recognition ability in police super-recognisers. *Applied Cognitive Psychology, 30*(6), 827-840. doi:10.1002/acp.3260

Davis, J. P., Treml. F., Forrest, C., & Jansari, A (2018). Identification from CCTV: Assessing police super-recognisers ability to spot faces in a crowd and susceptibility to change blindness. *Applied Cognitive Psychology, 32*(3), 337-353. doi:10.1002/acp.3405

DeGutis, J., Cohan, S., Mercado, R. J., Wilmer J., & Nakayama, K. (2012). Holistic processing of the mouth but not the eyes in developmental prosopagnosia. *Cognitive Neuropsychology*, *29*(5-6), 419-446. doi: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. doi:10.1016/j.cognition.2012.09.004

Diamond, R., & Carey, S. (1986). Why faces are and are not special: an effect of expertise. *Journal of Experimental Psychology: General*, *115*(2), 107-117. doi:10.1037/0096-3445.115.2.107

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 subjects. *Neuropsychologia, 44*(4), 576-585. doi:10.1016/j.neuropsychologia.2005.07.001

Hendel, R. K., Starrfelt, R., Gerlach, C. (2018). The good, the bad, and the average: Characterizing the relationship between face and object processing across the face recognition spectrum. *Neuropsychologia*, *124*, 274-284. doi:10.1016/j.neuropsychologia.2018.11.016

Horry, R., Cheong, W., & Brewer, N. (2014) The Other-Race Effect in perception and recognition: insights from the complete composite task. *Journal of Experimental Psychology: Human Perception and Performance, 41*(2), 508-524. doi:10.1037/xhp0000042

Jenkins, R. E., Tsermentseli, S., Monks, C. P., Robertson, D. J., Stevenage, S. V., Symons, A. E., & Davis, J. P. (2021). Are super-face-recognisers also super-voice-recognisers? Evidence from cross-modal identification tasks. *Applied Cognitive Psychology, 35*, 590-605. DOI: 10.1002/acp.3813

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. doi:10.1016/j.neuropsychologia.2018.03.026.

Konar, Y., Bennett, P. J., & Sekuler, A. B.(2010). Holistic processing is not correlated with face-identification accuracy. *Psychological Science, 21*(1), 38-43. doi:10.1177/0956797609356508

Le Grand, R., Cooper, P. A., Mondloch, C. J., Lewis, T. L., Sagiv, N., de Gelder, B., & Maurer, D. (2006). What aspects of face processing are impaired in developmental prosopagnosia? *Brain and Cognition, 6,* 139-158. doi:10.1016/j.bandc.2005.11.005

Li, J., Tian, M., Fang, H., Xu, M., Li, H., & Liu, J. (2010) Extroversion predicts individual differences in face recognition. *Communication & Integrative Biology*, *3*, 295-298.

doi:10.4161/cib.3.4.12093

Marzi, T., Gronchi, G., Turano, M.T., Giovannelli, F., Giganti, F., Rebai, M., & Viggiano, M.P. (2021). Mapping the Featural and Holistic Face Processing of Bad and Good Face Recognizers. *Behavioural Sciences, 11*, 1-14. https://doi.org/10.3390/bs11050075

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

McCaffery, J. M., Robertson, D. J., Young, A. W., & Burton, A. M. (2018). Individual differences in face identity processing. *Cognitive research: principles and implications*, *3*(1), 21. doi:10.1186/s41235-018-0112-9

Minear, M., & Park, D. C. (2004). A lifespan database of adult facial stimuli. *Behavior Research Methods, Instruments, & Computers, 36*, 630-633. doi:10.3758/BF03206543

Noyes, E., Hill, M.Q., & O'Toole, A.J. (2018). Face recognition ability does not predict person identification performance: using individual data in the interpretation of group results. *Cognitive Research: Principles and Implications, 3*, 1 – 13. https://doi.org/10.1186/s41235-018-0117-4

Noyes, E., Phillips, P. J., & O'Toole, A. J. (2017). *What is a super-recogniser?* In M. Bindemann & A. M. Megreya (eds.) Face processing: Systems, Disorders, and Cultural Differences. New York, NY: Nova.

Rezlescu, C., Susilo, T., Wilmer, J. B., & Caramazza, A. (2017). The inversion, part-whole, and composite effects reflect distinct perceptual mechanisms with Varied Relationships to Face Recognition. *Journal of Experimental Psychology: Human Perception and Performance*, *43*, 1961-1973. doi:10.1037/xhp0000400

Richler, J. J., Cheung, O. S., & Gauthier, I. (2011). Holistic Processing Predicts Face Recognition. *Psychological Science, 22*(4), 464–471. doi:10.1177/0956797611401753

Richler, J. J., Floyd, R. J., & Gauthier, I. (2014). The Vanderbilt Holistic Face Processing Test: A short and reliable measure of holistic face processing**.** *Journal of Vision, 14*(11), 10-10. doi:10.1167/14.11.10

Robertson, D J., Black, J., Chamberlain, B., Megreya, A., & Davis, J. P. (2020). Super-recognisers show an advantage for other race face identification. *Applied Cognitive Psychology*.

Robertson, D. J., Noyes, E., Dowsett, A. J., Jenkins, R., Burton, A. M. (2016). Face recognition by Metropolitan Police super-recognisers. *PLoS One, 11*(2), 1-8. doi:10.1371/journal.pone.0150036

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

Rossion, B. (2013). The composite face illusion: a whole window into our understanding of holistic face perception. *Visual Cognition, 21*, 139-253. doi:10.1080/13506285.2013.772929

Russell, R., Chatterjee, G., & Nakayama, K. (2012). Developmental prosopagnosia and superrecognition: No special role for surface reflectance processing. *Neuropsychologia, 50*(2), 334-340. doi: 10.1016/j.neuropsychologia.2011.12.004

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

Satchell, P. L., Davis, J., Julle-Danière, E., Tupper, N., & Marshman, P. (2019). Recognising faces but not traits: Accurate personality judgment from faces is unrelated to superior face memory. *Journal of Research in Personality, 79,* 49-58. doi:10.1016/j.jrp.2019.02.002.

Shakeshaft, N. G., & Plomin, R. (2015). Genetic Specificity of face recognition. *Proceedings of the National Academy of Sciences, 112*(41), 1-6. doi:10.1073/pnas.1421881112

Tanaka, J. W., & Farah, M. J. (1993). Parts and wholes in face recognition. *The Quarterly Journal of Experimental Psychology Section A, 46*(2), 225-245. doi:10.1080/17470218.2016.1146780

Tanaka, J. W., & Simonyi, D. (2016). The "parts and the wholes" of face recognition: a review of the literature. *The Quarterly Journal of Experimental Psychology, 69,* 1876-1889.

Tardif, J., Duchesne, X. M., Cohan, S., Royer, J., Blais, C., Fiset, D., … Gosselin, F. (2018). Use of face information varies systematically from developmental prosopagnosics to super-recognizers. *Psychological Science, 30,* 300-308. doi: 10.1177/0956797618811338

Verhallen, R. J., Bosten, J. M., Goodbourn, P. T., Lawrance-Owen, A. J., Bargary, G., & Mollon, J. D. (2017). General and specific factors in the processing of faces. *Vision Research, 141*, 217–227. doi:10.1016/j.visres.2016.12.014

Wang, R., Li, J., Fang, H., Tian, M., & Liu, J. (2012). Individual differences in holistic processing predict face recognition ability. *Psychological Science, 23*(2), 169–177. doi:10.1177/0956797611420575

Wilmer, J. B., Germine, L., Chabris, C. F., Chatterjee, G., Williams, M., Loken, E., … Duchaine, B. C. (2010). Human face recognition ability is specific and highly heritable. *Proceedings of the National Academy of Sciences, 107*(11), 5238-5241. doi:10.1073/pnas.0913053107

Yin, R. K. (1969). Looking at upside-down faces. *Journal of Experimental Psychology, 81*, 141-145. doi:10.1037/h0027474

Young, A. W., Hellawell, D., & Hay, D. C. (1987). Configurational information in face perception. *Perception 16*(6), 747-759. doi:10.1068/p160747

Yovel, G., Wilmer, J. B., & Duchaine, B. (2014). What can individual differences reveal about face processing? *Frontiers in Human Neuroscience, 8*(August), 1-9. doi:10.3389/fnhum.2014.00562

Supplementary Materials (SM_01)

Between-group analyses with robust two-test group classification (i.e. two tests) in Experiment 1

Glasgow Face Matching Test (GFMT; Burton et al., 2010). This test was employed as a second ability-verifying test in Experiment 1. The test comprises 40 trials and displays simultaneously presented pairs of white-Caucasian male and female facial images. Participants are required to respond whether the pairs are the same (50%) or different (50%).

Note that when two face processing tests (CFMT+ and GFMT) were used to classify participants into groups, numbers dropped from 55 SRs and 64 controls (as reported in the main paper) to 19 SRs and 51 controls. GFMT defined SRs scored 40/40 (M = 40.0, SD = 0) while controls scored < 39/40 (M = 35.67, SD = 2.24). The Results section reports analyses on the CFMT+ defined SRs and controls, while analyses on the CFMT+/GFMT defined groups are reported here. The differences in the analyses are commented on in the General Discussion. Tables S1 and S2 demonstrate between-group data and analyses comparing PWE in SRs and controls when using two tests to verify participants ability (CFMT+ and GFMT) in Experiment 1.

Table S1.

SRs and controls performance (%) on Whole and Part conditions

	SRs (n = 19) M (SD)		Controls (n = 51) M (SD)	
Eyes				
Whole	97.15	(3.42)	89.87	(10.04)
Part	80.48	(7.86)	71.90	(12.85)
Nose				
Whole	95.61	(4.91)	77.86	(12.29)
Part	81.14	(10.70)	71.73	(11.00)
Mouth				
Whole	96.05	(4.27)	84.72	(9.88)
Part	84.43	(8.98)	70.51	(10.20)

Table S2

ANOVAs examining PWE *in SRs and Controls using two tests to verify ability (CFMT+ and GFMT in Experiment 1*

df	PWE $_{Eyes}$		PWE $_{Nose}$		PWE $_{Mouth:}$	
(1, 68)	*F*	η^2	*F*	η^2	*F*	η^2
Main effects						
Group	10.53 *	.134	32.39 *	.323	38.01 *	.359
Condition	167.26 *	.711	37.15 *	.353	79.12 *	.538
Interaction	<1	.003	6.10 *	.082	<1	.012
			t	*d*		
SRs (1, 18)			5.80 *	1.46		
Controls (1, 50)			3.33 *	0.47		

Supplementary materials (SM_02)

Individual analyses for Experiments 1

Tables S3 – S5 report individual analyses for each SR in Experiment 1 (*n* = 55). These results are derived from the classification of SR and control groups based on the CFMT+ only. The tables show the SR's residual scores for each face region (Eyes, Nose, and Mouth), which are the deviations from the control group regression line. The table shows SR's deviation scores against the control mean. The tables report z scores for each SR residual score, as well as the upper and lower bound confidence intervals (95%) of the estimated proportion of the population expected to fall below each SR (Crawford, Garthwaite, & Porter, 2010). For example, in Table S3, 71.77% of population CI_{95}[62.41, 80.14] is expected to generate a residual score below the one generated by SR 1 with a residual score of 5.19.

Table S3.

Individual analyses for PWE eyes in Experiment 1 (control M = 0.23, SD = 8.50)

				PWE eyes		
SRs	CFMT+	residual	z	% population	CI lower	CI upper
1	102	5.19	0.58	71.77	62.41	80.14
2	101	-0.22	-0.05	47.91	38.29	57.63
3	101	3.28	0.36	63.85	54.17	72.92
4	101	-5.39	-0.66	25.71	17.63	34.89
5	101	-0.89	-0.13	44.82	35.30	54.57
6	101	7.44	0.85	79.84	71.21	87.12
7	100	13.19	1.53	93.23	87.71	97.02
8	100	-1.22	-0.17	43.31	33.84	53.06
9	100	3.61	0.40	65.28	55.63	74.24
10	100	3.28	0.36	63.85	54.17	72.92
11	99	-8.89	-1.07	14.55	8.40	22.30
12	99	11.28	1.30	89.91	83.23	94.85
13	99	1.03	0.09	53.71	43.97	63.28
14	99	3.28	0.36	63.85	54.17	72.92
15	99	9.36	1.07	85.47	77.73	91.62
16	98	7.44	0.85	79.84	71.21	87.12
17	98	7.44	0.85	79.84	71.21	87.12
18	98	9.36	1.07	85.47	77.73	91.62
19	98	-4.72	-0.58	28.27	19.89	37.63
20	98	11.28	1.30	89.91	83.23	94.85
21	98	-2.47	-0.32	37.68	28.51	47.40
22	98	5.19	0.58	71.77	62.41	80.14
23	98	1.69	0.17	56.74	46.98	66.20
24	98	1.36	0.13	55.23	45.48	64.74
25	98	15.11	1.75	95.64	91.25	98.38
26	98	3.28	0.36	63.85	54.17	72.92
27	98	9.36	1.07	85.47	77.73	91.62
28	97	3.61	0.40	65.28	55.63	74.24
29	97	3.61	0.40	65.28	55.63	74.24
30	97	13.19	1.53	93.23	87.71	97.02
31	97	-4.72	-0.58	28.27	19.89	37.63
32	97	1.69	0.17	56.74	46.98	66.20
33	97	1.69	0.17	56.74	46.98	66.20
34	97	5.53	0.62	73.08	63.81	81.31
35	97	1.69	0.17	56.74	46.98	66.20
36	97	-0.56	-0.09	46.34	36.76	56.08
37	96	10.61	1.22	88.49	81.43	93.86
38	96	-2.47	-0.32	37.68	28.51	47.40
39	96	7.44	0.85	79.84	71.21	87.12
40	96	1.69	0.17	56.74	46.98	66.20
41	96	-6.64	-0.81	21.28	13.82	30.04
42	96	-5.06	-0.62	26.95	18.72	36.23
43	96	-0.56	-0.09	46.34	36.76	56.08
44	96	3.61	0.40	65.28	55.63	74.24
45	95	3.61	0.40	65.28	55.63	74.24
46	95	5.19	0.58	71.77	62.41	80.14
47	95	3.28	0.36	63.85	54.17	72.92
48	95	-0.56	-0.09	46.34	36.76	56.08
49	95	1.36	0.13	55.23	45.48	64.74
50	95	7.44	0.85	79.84	71.21	87.12
51	95	5.53	0.62	73.08	63.81	81.31
52	95	1.69	0.17	56.74	46.98	66.20
53	95	1.69	0.17	56.74	46.98	66.20
54	95	5.19	0.58	71.77	62.41	80.14
55	95	-1.22	-0.17	43.31	33.84	53.06

Table S4.

Individual analyses for PWE nose in Experiment 1 (control M = -0.003, SD = 11.52)

SRs	CFMT+	residual	z	PWE nose % population	CI lower	CI upper
1	102	12.90	1.12	86.47	78.93	92.37
2	101	5.22	0.45	67.28	57.70	76.09
3	101	19.92	1.73	95.45	90.96	98.28
4	101	-6.84	-0.59	27.90	19.56	37.24
5	101	-6.84	-0.59	27.90	19.56	37.24
6	101	16.41	1.43	91.88	85.84	96.17
7	100	18.16	1.58	93.86	88.60	97.39
8	100	7.42	0.64	73.76	64.53	81.90
9	100	12.24	1.06	85.22	77.42	91.42
10	100	16.41	1.43	91.88	85.84	96.17
11	99	12.90	1.12	86.47	78.93	92.37
12	99	-29.43	-2.55	0.69	0.11	2.05
13	99	18.60	1.62	94.30	89.23	97.64
14	99	-2.01	-0.17	43.17	33.71	52.92
15	99	8.07	0.70	75.53	66.45	83.45
16	98	2.59	0.23	58.80	49.05	68.17
17	98	8.07	0.70	75.53	66.45	83.45
18	98	2.15	0.19	57.32	47.57	66.76
19	98	11.58	1.01	83.89	75.85	90.39
20	98	0.40	0.04	51.38	41.67	61.02
21	98	14.65	1.27	89.42	82.61	94.51
22	98	17.50	1.52	93.17	87.61	96.98
23	98	-2.67	-0.23	40.95	31.60	50.71
24	98	21.67	1.88	96.67	92.91	98.89
25	98	17.50	1.52	93.17	87.61	96.98
26	98	11.14	0.97	82.96	74.77	89.65
27	98	6.32	0.55	70.60	61.18	79.10
28	97	8.73	0.76	77.26	68.35	84.94
29	97	6.32	0.55	70.60	61.18	79.10
30	97	19.92	1.73	95.45	90.96	98.28
31	97	2.59	0.23	58.80	49.05	68.17
32	97	12.90	1.12	86.47	78.93	92.37
33	97	8.73	0.76	77.26	68.35	84.94
34	97	9.83	0.85	79.99	71.38	87.24
35	97	4.57	0.40	65.25	55.60	74.22
36	97	14.65	1.27	89.42	82.61	94.51
37	96	-5.74	-0.50	31.15	22.48	40.66
38	96	-1.58	-0.14	44.62	35.10	54.37
39	96	-12.32	-1.07	14.64	8.47	22.40
40	96	9.39	0.82	78.92	70.18	86.35
41	96	-11.01	-0.96	17.34	10.58	25.57
42	96	11.58	1.01	83.89	75.85	90.39
43	96	19.92	1.73	95.45	90.96	98.28
44	96	11.58	1.01	83.89	75.85	90.39
45	95	-0.26	-0.02	49.12	39.46	58.82
46	95	14.65	1.27	89.42	82.61	94.51
47	95	-7.50	-0.65	26.04	17.92	35.25
48	95	-2.01	-0.17	43.17	33.71	52.92
49	95	1.49	0.13	55.10	45.35	64.62
50	95	2.59	0.23	58.80	49.05	68.17
51	95	9.17	0.80	78.38	69.58	85.89
52	95	21.67	1.88	96.67	92.91	98.89
53	95	14.65	1.27	89.42	82.61	94.51
54	95	16.41	1.43	91.88	85.84	96.17
55	95	18.60	1.62	94.30	89.23	97.64

Table S5.
Individual analyses for PWE mouth in Experiment 1 (control M = 0.02, SD = 9.68)

SRs	CFMT+	residual	z	*PWE mouth* % population	CI lower	CI upper
1	102	10.30	1.06	85.20	77.40	91.41
2	101	6.13	0.63	73.33	64.08	81.53
3	101	7.30	0.75	77.09	68.15	84.79
4	101	-1.70	-0.18	43.03	33.58	52.79
5	101	13.80	1.42	91.86	85.81	96.16
6	101	10.30	1.06	85.20	77.40	91.41
7	100	11.47	1.18	87.75	80.51	93.32
8	100	6.13	0.63	73.33	64.08	81.53
9	100	9.13	0.94	82.30	74.01	89.13
10	100	7.80	0.80	78.59	69.81	86.07
11	99	4.97	0.51	69.32	59.83	77.94
12	99	9.63	0.99	83.58	75.50	90.15
13	99	8.47	0.87	80.52	71.97	87.67
14	99	-0.37	-0.04	48.41	38.77	58.12
15	99	9.13	0.94	82.30	74.01	89.13
16	98	-0.53	-0.06	47.76	38.14	57.48
17	98	0.13	0.01	50.45	40.76	60.11
18	98	8.97	0.93	81.88	73.52	88.79
19	98	-7.03	-0.73	23.63	15.82	32.63
20	98	-1.03	-0.11	45.73	36.18	55.48
21	98	7.30	0.75	77.09	68.15	84.79
22	98	4.30	0.44	66.88	57.29	75.72
23	98	11.97	1.24	88.74	81.74	94.03
24	98	6.13	0.63	73.33	64.08	81.53
25	98	9.13	0.94	82.30	74.01	89.13
26	98	7.97	0.82	79.09	70.37	86.49
27	98	11.47	1.18	87.75	80.51	93.32
28	97	8.47	0.87	80.52	71.97	87.67
29	97	1.97	0.20	57.89	48.13	67.30
30	97	9.63	0.99	83.58	75.50	90.15
31	97	-8.87	-0.92	18.28	11.34	26.66
32	97	14.97	1.54	93.48	88.05	97.16
33	97	10.30	1.06	85.20	77.40	91.41
34	97	1.97	0.20	57.89	48.13	67.30
35	97	0.80	0.08	53.17	43.44	62.77
36	97	7.97	0.82	79.09	70.37	86.49
37	96	7.80	0.80	78.59	69.81	86.07
38	96	9.13	0.94	82.30	74.01	89.13
39	96	16.13	1.66	94.82	90.00	97.94
40	96	10.3	1.06	85.20	77.40	91.41
41	96	-1.70	-0.18	43.03	33.58	52.79
42	96	2.47	0.25	59.87	50.13	69.18
43	96	6.63	0.68	74.97	65.85	82.97
44	96	0.13	0.01	50.45	40.76	60.11
45	95	12.47	1.29	89.67	82.93	94.69
46	95	6.13	0.63	73.33	64.08	81.53
47	95	-36.2	-3.74	0.02	<0.01	0.12
48	95	9.13	0.94	82.30	74.01	89.13
49	95	-3.37	-0.35	36.47	27.38	46.16
50	95	-19.03	-1.97	2.76	0.85	6.13
51	95	5.47	0.56	71.08	61.68	79.53
52	95	4.30	0.44	66.88	57.29	75.72
53	95	12.63	1.30	89.96	83.29	94.88
54	95	6.63	0.68	74.97	65.85	82.97
55	95	9.63	0.99	83.58	75.50	90.15

Supplementary Materials (SM_03)

Between-group analyses with robust two-test group classification (i.e. two tests) in Experiment 2

Sequential Face Matching Test (SFMT). This test was used to verify face recognition ability in Experiment 2. The face stimuli were obtained from the Park Aging Mind database at The University of Texas at Dallas (Minear & Park, 2004) with Adobe Photoshop used to remove external features (i.e. hair). The test was administered using PsycoPy2 (Pierce, 2007). Each trial (n = 48) began with a central fixation cross (500msec), followed by a target face (500msec), an inter-stimulus interval (500msec), and a probe face (500msec). The probe was followed by another fixation cross which stayed on screen until a response was made. The participants used the keyboard to respond whether face pairs were the same ('S') or different ('D'). The visual angle of target image presentation was 4.9° by 5.7°. The probe image size was reduced by 24% compared to target image, in order to encourage judgements based on identity, and not on image variation.

When two tests were used for group classification, the cut-off point on the SFMT was M = 75.5, which was the midpoint between CFMT+ defined SR (M = 79.1, SD = 9.6) and control means (M = 71.9, SD = 7.3). With the stricter selection criteria, participant numbers dropped from 20 SRs and 25 controls (reported in the main paper) to 15 SRs (M = 83.46, SD = 4.35) and 15 controls (M = 68.75, SD = 5.80). Analyses for CFMT+ defined groups are reported in the Results section, while group analyses for CMFT+ and SFMT defined groups are reported here. The comparison of analyses outcomes is commented on in the General Discussion. Tables S6 and S7 report between-group data and analyses comparing PWE in SRs and controls when using two tests to verify participants ability (CFMT+ and SFMT) in Experiment 2.

Table S6.

SRs and controls performance (%) on Whole and Part conditions

	SRs (n = 15)		Controls (n = 15)	
	M (SD)		M (SD)	
Eyes				
Whole	88.06	(10.55)	80.28	(16.40)
Part	72.50	(7.68)	66.94	(12.55)
Nose				
Whole	96.94	(12.78)	76.39	(10.64)
Part	65.17	(9.68)	62.78	(15.39)
Mouth				
Whole	82.50	(13.47)	73.61	(12.76)
Part	76.94	(12.19)	63.61	(10.14)

Table S7.

ANOVAs examining PWE in SRs and Controls in Experiment 2 using two tests to verify ability

df	PWE $_{Eyes}$		PWE $_{Nose}$		PWE $_{Mouth:}$	
(1, 80)	*F*	η^2	*F*	η^2	*F*	η^2
Main Effects						
Group	3.02	.097	3.01	.097	9.48 *	.253
Condition	40.17 *	.589	39.39 *	.584	8.84 *	.240
Interaction	<1	.008	2.50	.082	<1	.025

$p < .050$ *

Individual analyses for Experiments 2

Tables S8 – S10 report individual analyses for each SR in Experiment 2 (*n* = 20). These results are derived from the classification of SR and control groups based on the CFMT+ only. The tables show the SR's residual scores for each face region (Eyes, Nose, and Mouth), which are the deviations from the control group regression line. The tables report z scores for each SR residual score, as well as the upper and lower bound confidence intervals (95%) of the estimated proportion of the population expected to fall below each SR. For example, in Table S8, 90.07% of population CI_{95}[78.68, 97.04] is expected to generate a residual score below the one generated by SR 1 with a residual score of 16.82.

Table S8.

Individual analyses for PWE eyes in Experiment 2 (control M = 0.13, SD = 11.98)

				PWE eyes		
SRs	CFMT+	residual	z	% population	CI lower	CI upper
1	101	16.82	1.35	90.07	78.68	97.04
2	101	9.82	0.81	78.23	63.65	89.54
3	100	13.99	1.16	86.61	73.9	95.14
4	99	5.66	0.46	67.26	51.77	80.79
5	99	5.7	0.47	67.37	51.88	80.89
6	99	4.24	0.34	63.03	47.45	77.14
7	99	2.82	0.23	58.62	43.06	73.21
8	98	-8.22	-0.70	25.04	12.93	40.02
9	98	11.24	0.93	81.39	67.36	91.8
10	97	18.2	1.51	92.39	82.21	98.12
11	96	2.86	0.23	58.74	43.19	73.33
12	96	-11.01	-0.93	18.55	8.16	32.57
13	96	8.41	0.69	74.78	59.79	86.93
14	96	12.62	1.04	84.16	70.75	93.64
15	95	-20.72	-1.74	5.04	0.91	13.44
16	95	1.49	0.11	54.39	38.94	69.34
17	95	9.82	0.81	78.23	63.65	89.54
18	95	11.28	0.93	81.47	67.46	91.85
19	95	9.82	0.81	78.23	63.65	89.54
20	95	11.24	0.93	81.39	67.36	91.8

Table S9.

Individual analyses for PWE nose in Experiment 2 (control M = 0.01, SD = 10.31)

SRs	CFMT+	residual	z	*PWE nose* % population	CI lower	CI upper
1	101	15.61	1.51	92.46	82.31	98.15
2	101	-0.55	-0.05	47.91	32.79	63.26
3	100	19.95	1.93	96.50	89.56	99.53
4	99	-1.22	-0.12	45.40	30.46	60.86
5	99	24.78	2.40	98.65	94.69	99.93
6	99	-1.22	-0.12	45.40	30.46	60.86
7	99	19.45	1.89	96.16	88.86	99.44
8	98	11.95	1.16	86.64	73.94	95.16
9	98	-25.39	-2.46	1.18	0.06	4.84
10	97	20.28	1.97	96.71	89.99	99.58
11	96	11.95	1.16	86.64	73.94	95.16
12	96	19.78	1.92	96.39	89.32	99.50
13	96	24.11	2.34	98.45	94.13	99.90
14	96	19.11	1.85	95.91	88.34	99.37
15	95	19.61	1.90	96.27	89.09	99.47
16	95	11.45	1.11	85.64	72.63	94.56
17	95	12.45	1.21	87.59	75.21	95.71
18	95	15.95	1.55	92.88	82.98	98.32
19	95	15.45	1.50	92.25	81.99	98.06
20	95	7.78	0.75	76.66	61.87	88.37

Table S10.

Individual analyses for PWE mouth in Experiment 2 (control M =0.18, SD =10.79)

SRs	CFMT+	residual	z	*PWE mouth* % population	CI lower	CI upper
1	101	5.98	0.54	69.85	54.48	82.97
2	101	3.98	0.35	63.36	47.78	77.43
3	100	4.15	0.37	63.93	48.36	77.93
4	99	-29.69	-2.77	0.60	0.01	2.93
5	99	11.98	1.09	85.29	72.18	94.35
6	99	-8.19	-0.78	22.71	11.16	37.42
7	99	-0.19	-0.03	48.67	33.50	63.99
8	98	22.15	2.04	97.13	90.91	99.67
9	98	10.15	0.92	81.30	67.25	91.74
10	97	-2.52	-0.25	40.41	25.91	55.98
11	96	-2.19	-0.22	41.56	26.95	57.12
12	96	-6.35	-0.61	27.92	15.21	43.15
13	96	-2.35	-0.23	41.01	26.45	56.57
14	96	18.31	1.68	94.38	85.52	98.89
15	95	9.98	0.91	80.90	66.77	91.46
16	95	12.15	1.11	85.63	72.61	94.56
17	95	-17.19	-1.61	6.38	1.39	15.78
18	95	7.81	0.71	75.26	60.32	87.30
19	95	16.31	1.50	92.22	81.93	98.04
20	95	1.98	0.17	56.43	40.92	71.22