Neuropsychologia 50 (2012) 1698–1707

Contents lists available at SciVerse ScienceDirect

Neuropsychologia

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

Crowd perception in prosopagnosia

Allison Yamanashi Leib[a,*], Amrita M. Puri[b], Jason Fischer[a], Shlomo Bentin[c], David Whitney[a], Lynn Robertson[a,d]

[a] University of California, Berkeley, United States
[b] Hendrix College, United States
[c] Hebrew University of Jerusalem, Israel
[d] Veterans Administration, Northern California Health Care System, United States

ARTICLE INFO

Article history:
Received 19 October 2011
Received in revised form 21 March 2012
Accepted 22 March 2012
Available online 11 April 2012

Keywords:
Prosopagnosia
Ensemble coding

ABSTRACT

Prosopagnosics, individuals who are impaired at recognizing single faces, often report increased difficulty when confronted with crowds. However, the discrimination of crowds has never been fully tested in the prosopagnosic population. Here we investigate whether developmental prosopagnosics can extract ensemble characteristics from groups of faces. DP and control participants viewed sets of faces varying in either identity or emotion, and were asked to estimate the average identity or emotion of each set. Face sets were displayed in two orientations (upright and inverted) to control for low-level visual features during ensemble encoding. Control participants made more accurate estimates of the mean identity and emotion when faces were upright than inverted. In all conditions, DPs performed equivalently to controls. This finding demonstrates that integration across different faces in a crowd is possible in the prosopagnosic population and appears to be intact despite their face recognition deficits. Results also demonstrate that ensemble representations are derived differently for upright and inverted faces, and the effects are not due to low-level visual information.

1. Introduction

Every day we interact with crowds of people. Whether it is on a city bus, in a classroom, or in a business meeting, we routinely view and extract important information from groups of faces, and do so rather rapidly. Indeed, recent studies have shown that people are adept at recognizing crowd characteristics, such as average gender, identity or emotion, even when crowds are viewed so briefly that information about any specific individual is not extracted (De Fockert & Wolfenstein, 2009; Haberman & Whitney, 2007, 2009). For example, as a passenger on a bus, we form a general impression of important characteristics of a crowd standing on the street corner, even if we are only able to view the crowd for a split-second as we ride by.

Given the frequency with which we interact with crowds, a deficit in perceiving crowd characteristics would likely pose a hindrance in a host of social situations. Anecdotal evidence suggests that individuals with prosopagnosia, a deficit in discriminating individual faces, feel overwhelmed in crowded situations, perhaps in part due to their inability recognize familiar faces in a crowd. For example, one prosopagnosic describes his experience walking into a reception hall, "*There are a lot of people there, perhaps as many as a hundred or so people. These are all people I am supposed to know, each with a supposedly unique face. My goal is to find just one specific individual. I can scan the room for hours in frustration*. . . (Asprin, 2011)." Another prosopagnosic expresses frustration saying, "*Faces in public are just all faces to me, I don't see them individually. This is especially [true] in crowded public areas. When I look into a crowd, most look very much alike to me* (BP, 2011)." Can prosopagnosics' discomfort with crowds be explained entirely by their deficits in perceiving single faces? Or could it reflect a more general impairment in integrating and extracting face-related information from a crowd? On the other hand, might prosopagnosics actually be better at ensemble coding because they do not perceive crowd members as distinct individuals?

The perceptual characteristics of developmental prosopagnosics[2] (DPs), individuals who have never fully developed the ability to recognize faces, have been increasingly studied during the last decade. However, almost all of the previous research used single faces to investigate processing in DPs. Although the study of individual face processing in DP added essential information aiding the understanding of problems related to individual face recognition, we know virtually nothing about how DPs extract information from groups of faces and whether it is normal or not.

* Corresponding author at: 4143 Tolman Hall #5050, Berkeley, CA 94702, United States. Tel.: +1 510 642 6266; fax: +1 510 642 5293.
E-mail addresses: ayleib@berkeley.edu, ayleib@gmail.com (A.Y. Leib).

[2] Developmental prosopagnosia is also referred to as congenital prosopagnosia.

0028-3932/$ – see front matter
http://dx.doi.org/10.1016/j.neuropsychologia.2012.03.026

When processing crowds, typical viewers initially discount individual faces in a group and instead formulate unitized percepts that accurately describe crowd characteristics (De Fockert & Wolfenstein, 2009; Haberman & Whitney, 2007, 2009). The ability to generate a gestalt percept of the crowd, independent of information derived from individual faces, can be viewed as a mechanism that compensates for the limited capacity of the visual system to process multiple items simultaneously. Redundant information across items in a scene is compressed into an average representation of the entire set, referred to as the "ensemble code" (Alvarez, 2011; Ariely, 2001; Chong & Treisman, 2003). This average representation provides a more precise description in comparison to individual evaluations of each member of the set because noise from one individual evaluation cancels out uncorrelated noise from another individual evaluation (Alvarez, 2011). As such, it has been shown that typical viewers can accurately extract both the mean emotional expression and mean identity of the crowd, although performance is at chance when they are asked to discriminate, identify, or localize individual members of a previously seen set (De Fockert & Wolfenstein, 2009; Haberman & Whitney, 2007, 2009).

Previous research suggests that DPs have trouble integrating individual face features into a gestalt (Behrmann, Avidan, Marotta, & Kimchi, 2005; de Gelder & Rouw, 2000; Lobmaier, Bölte, Mast, & Dobel, 2010), and may be generally impaired at identifying the global shape of a stimulus, showing such deficits for objects as well as faces (Avidan, Tanzer, & Behrmann, 2011; Behrmann & Avidan, 2005; Behrmann et al., 2005; Bentin, DeGutis, D'Esposito, & Robertson, 2007; Palermo et al., 2011). For alternative findings see: Le Grand et al. (2006), Duchaine, Yovel, & Nakayama (2007), Schmalzl, Palermo, Green, Brunsdon, and Coltheart (2008) and Lee, Duchaine, Wilson, and Nakayama (2010). Ensemble coding, like other holistic processing tasks, requires the integration of features across space (Alvarez, 2011) or time (Haberman, Harp, & Whitney, 2009). If DPs have difficulty with this type of integration in general, we may expect that they will have trouble forming a unitary percept of any attribute of a crowd, not just average identity. Alternatively, it is possible that the deficits DPs experience during individual face recognition tasks will be minimized via the process of ensemble coding. As mentioned previously, ensemble coding involves canceling out "noisy" individual evaluations, thereby achieving a more precise representation of the group as a whole. Although individual face evaluations are suboptimal in DP, the averaging process inherently reduces such imprecision. This leaves open the intriguing possibility that DPs, who are impaired at individual face identification, may be able to extract the mean identity of the crowd just as well as controls. If DPs do not experience interference by individual faces in the crowd, they could potentially be better than normal perceivers at extracting ensemble information.

The aim of this study was to explore whether DPs can successfully perceive ensemble characteristics of face sets, or "crowds." In order to distinguish between deficits specific to the perception of face identity and impairment in ensemble coding in general, we measured the ability to estimate not only the average identity of upright faces, but also the average emotional expression, an attribute for which DPs typically exhibit little impairment when performing judgments on individual faces (Bentin, Deouell & Soroker, 1999; Dobel, Bölte, Aicher, & Schweinberger, 2007; Duchaine, Parker, & Nakayama, 2003; Humphreys, Avidan, & Behrmann, 2007; Jones & Tranel, 2001; for an different view see Palermo et al., 2011). Accordingly, we limited our group of participants to those who reported no or very little impairment in emotional processing of faces. Furthermore, we included conditions in which the face sets were inverted to control for low-level visual effects during ensemble coding.

2. Methods

2.1. Participants

Four DP individuals (DP1, DP2, DP3, and DP4) participated in the experiment. Three of the 4 DPs were recruited from a volunteer pool of previously diagnosed prosopagnosics (results from original tests are reported in Section 2.3). One had not participated in previous studies and was newly screened. We asked the DP participants to describe their experiences with faces and whether they found it difficult to recognize individuals and/or the emotion of individual faces by vision alone. All of them reported experiencing severe difficulty in face recognition and reported that these deficits substantially interfered with daily functioning. For instance, DP1 (female, 43 years) noted she had difficulty watching movies because the characters appeared similar to her; DP2 (female, 30 years) mentioned having trouble finding her parents in an airport; and DP3 (female, 54 years) reported experiencing difficulty recognizing close friends out of context; DP4 (female, 59 years) noted that she was unable to distinguish between students in her classroom. By contrast, none of them reported problems recognizing emotions from faces, albeit DP1 described mild impairment in emotional processing. These questions were asked before any tests were administered.

We recruited twenty participants to serve as controls for this experiment. Five control participants were matched on age and gender to each of the prosopagnosics. Control group mean ages and standard deviations are as follows: DP1 control group: $M = 38.8$ SD = 3.03; DP2 control group: $M = 30$ SD = 3.32; DP3: control group $M = 53.4$ SD = 4.98; DP4 control group: $M = 59.8$ SD = 4.92. All control participants were recruited from the general population, first by a short telephone interview. All reported that they had no difficulty recognizing or identifying faces. Qualified participants were then asked to come into the lab for testing and underwent the same standardized and experimental tests as the DPs.

2.2. Standard face recognition tests

Before experimental testing, we administered two standardized face processing tests. In the Benton Face Recognition Test (BFRT, Benton & Van Allen, 1968) a greyscale target face is presented at the top in each display, and the participant is asked to select the face that is the same person from among 6 faces in different rotated orientations below it. In the Warrington Recognition Memory Test (WRMT, Warrington, 1984) 50 study faces (greyscale) with hair and clothing intact are presented sequentially. At test, two faces are presented. One is the same as a study face, and the participant is asked to choose which of the two faces was previously viewed. All faces are presented in the same orientation and with the same lighting conditions at both study and test. To control for general memory ability, 50 words are also presented sequentially for study. At test, a single sheet of paper is presented containing half study words and half new words. The participant verbally reports the words that were previously viewed. In addition, we presented the Berkeley Famous Faces Test, a locally developed test, in which participants were asked to identify 25 celebrity faces (e.g., Bill Clinton, Elvis Presley, etc.). The aim of this test was to assess face identification in DP relative to typically developed individuals with the same cultural background. Importantly, participants were not required to recall the name of the celebrity (although giving the correct name was clear evidence of recognition). For instance, if they said an American president for Bill Clinton, that was considered correct. After the test was concluded, we also controlled for participants' familiarity with celebrities by asking participants to confirm their exposure to each celebrity presented in the test. If the participant was unfamiliar with a particular celebrity, their response to this celebrity was excluded from the analysis. Table 1 shows participants' exposure levels. Despite these allowances, all DPs had great difficulty recognizing celebrity faces (all 60% or less accuracy) and all DPs were statistically worse than controls on at least one of the face measures (see below).

2.3. Results for standardized face recognition tests

Fig. 1a shows performance on the Berkeley Famous Faces Test. Each DP is shown as a single triangle in a given color with their controls shown in the same color as the DP to which they were matched. All DP's performed worse than their respective control participants on a t-test designed for single case studies with small n (all $p < .05$ Crawford, Garthwaite, & Howell, 2009). Fig. 1b shows performance on the BFRT. Two DPs (DP1 and DP2) scored significantly below the mean of their matched controls ($p < .05$ and the remaining two were trending in the same direction).[3]

Table 1 presents the scores on the Warrington Recognition Memory Test. Again, two of the DP participants (DP2 and DP4) performed significantly below controls and one was trending toward significance. Traditionally, Warrington scores are presented as the difference in performance between word and face recognition. Participants who score better on word recognition compared to face recognition are categorized as having a "face discrepancy" (Warrington, 1984). While a face

[3] Norms defined by Benton are as follows: *Normal* = 75–100%; *Borderline* = 72–74%; *Moderately Impaired* = 69–70%; *Severely Impaired* = <69%.

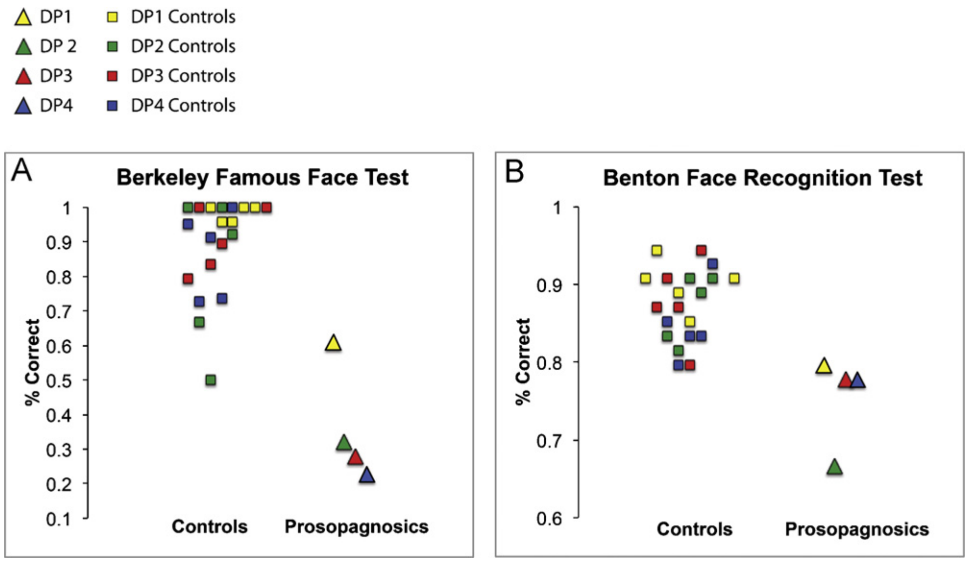

Fig. 1. (a) The left graph depicts individual DP and control performance on the Berkeley Famous Face tests. Each DP is shown as a single triangle in a given color with their matched controls shown in the same color as the DP to which they were matched. The DP distribution of performance is substantially below the control distribution of performance. (b) The right graph depicts DP performance compared to control performance on the Benton Facial Recognition Test. Once again, the DP distribution of performance is well below the control distribution of performance. Fifty-four is the total number of possible correct identifications. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of the article.)

discrepancy may represent an impairment in face recognition, it is also possible that a face discrepancy simply represents a proficient memory for words. Therefore, we have included the both the discrepancy score and the raw scores in Table 2.

Both Benton and Warrington tests have been criticized as measures for assessing prosopagnosia because results might be affected by the presence of non-face cues during study (Duchaine & Nakayama, 2004, 2006), we nonetheless presented these data to allow comparisons with many previous studies of DP.

During their original testing sessions, 3 prosopagnosics completed other face recognition tests, such as the Cambridge Face Memory Test (CFMT) and the Cambridge Face Perception Test (CFPT). To maximize consistency, we required the 4th prosopagnosic to complete these tasks as well. Both the CFMT & the CFPT require participants to match a target face to test faces positioned in different orientations and with varying levels of noise added. Importantly, both of these tests exclude non-face details, and therefore more accurately assess face recognition than either the Warrington or Benton tests (Bowles et al., 2009; Duchaine & Nakayama, 2006). Internal reliability, as measured by Chronbach's alpha is: CMFT, $\alpha = .89$; CMPT, $\alpha = .74$ (Bowles et al., 2009). The CFPT measures the ability to *perceive* faces (the target face and test faces are presented simultaneously), while the CFMT measures the ability to *remember* faces (target and test faces are presented sequentially with delay in between). All of our prosopagnosics scored 2 standard deviations below the mean, or worse, confirming that these participants are impaired at both perceiving and remembering faces. Table 3 includes the scores for prosopagnosics along with control means and standard deviations, as reported by (Bowles et al., 2009; Duchaine & Nakayama, 2006) Please note that a higher score in the CFPT indicates poor performance, whereas a lower score in the CFMT indicates poor performance.

Table 1
The scores of prosopagnosic participants and their matched controls on the Berkeley Famous Face Test. Exposure level to the 25 celebrities is recorded in the right column. If a participant was not exposed to a celebrity, this trial was deleted from the test.

Subject	Percent correct	Exposure out of 25
DP1	.61	23
Matched Controls		
	1	23
	1	23
	1	23
	.96	25
	.96	24
DP2	.32	25
Matched Controls		
	1	24
	.67	15
	.5	15
	1	23
	.92	25
DP3	.28	25
Matched Controls		
	.80	24
	1	25
	.83	24
	.89	19
	1	25
DP4	.22	22
Matched Controls		
	.95	21
	.73	22
	.91	23
	.74	19
	1	25

Table 2
The scores of prosopagnosic participants and their matched controls on the Warrington Recognition Memory Test. The upper table represents participants' raw scores out of 50 trials. The lower table represents participants' scores on the face memory test relative to participants' scores on the word memory test (the traditional scoring method of the Warring Recognition Memory Test).

DP1 score	DP2 score	DP3 score	DP4 score
37	29	43	38
Matched Controls	Matched Controls	Matched Controls	Matched Controls
39	45	37	50
40	34	44	50
35	39	46	42
45	44	42	46
44	42	44	41
Control Average	Control Average	Control Average	Control Average
40.6	40.8	42.6	45.8

DP1 score	DP2 score	DP3 score	DP4 score
13 FD	20 FD	7 FD	11 FD
Matched Controls	Matched Controls	Matched Controls	Matched Controls
11 FD	3 FD	13 FD	1 WD
9 FD	14 FD	5 FD	1 WD
13 FD	6 FD	4 FD	8 FD
5 FD	5 FD	7 FD	2 FD
6 FD	7 FD	3 FD	2 FD
Control Average	Control Average	Control Average	Control Average
8.8 FD	7 FD	6.4 FD	2 FD

FD: face discrepancy all scores are out of 50; WD: word discrepancy.

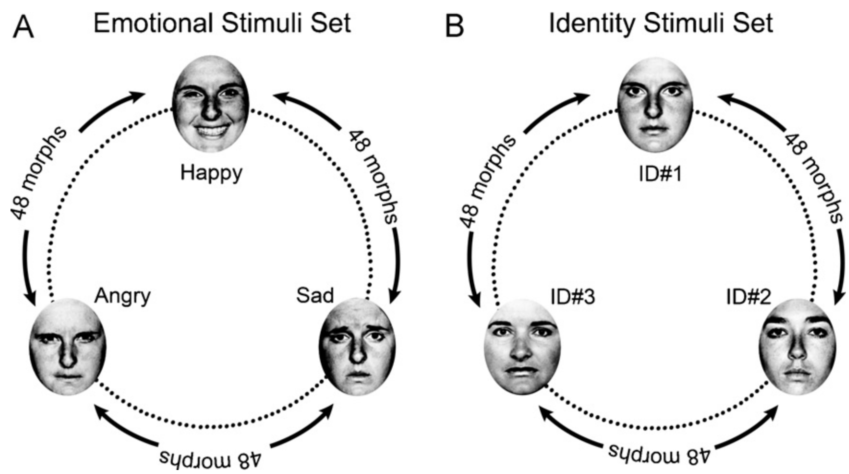

Fig. 2. (a) A schematic representation of the "Emotions" stimulus set. We morphed between 3 original faces (happy, sad, and angry) to generate 147 faces, each with the same identity but displaying a slightly different emotional expression. (b) The "Identities" stimulus set also consisted of 147 faces generated from 3 original faces; in this set each face had a different identity but shared the same neutral expression.

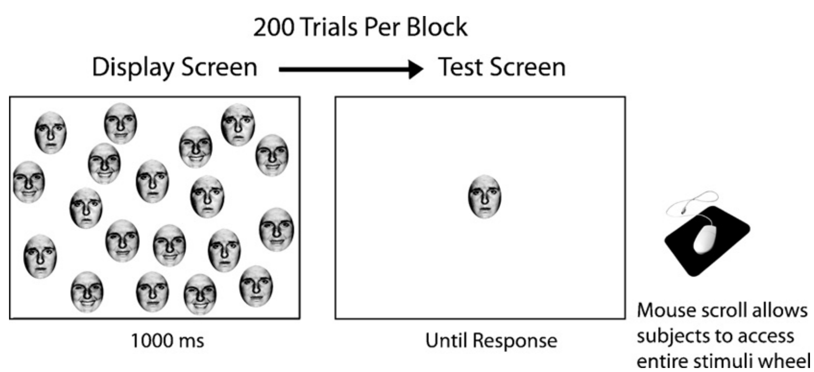

Fig. 3. Participants viewed the display for 1000 ms. Using a scrolling mouse, participants chose the face that best represented their estimate of the mean identity/emotion of the display.

2.4. *Stimuli*

Our ensemble coding experiment incorporated two face sets. One set was comprised of 147 photographs of the same face displaying diverse emotional expressions, while the other set was comprised of 147 faces with neutral emotional expressions but diverse identities. For both tested dimensions 18 faces were selected from the respective set for each study display (described below) and the sets also served as a continuum from which participants' selected their estimated mean judgment on each trial. Both stimulus sets consisted of morphed greyscale faces from the Ekman gallery (Ekman & Friesen, 1976). Each face subtended 2.86×3.53 degrees of the visual angle. The morphing procedure for the "Emotions" set proceeded as follows: First, 3 faces (same identity) with happy, sad, or angry emotional expressions were selected from the Ekman gallery. Next, the faces were linearly morphed to produce 48 morphs between each pair of basic emotions (i.e., 48 morphs between happy and sad, 48 morphs between sad and angry, and 48 morphs between angry and happy). Note that the set forms a circular continuum where there is no beginning or ending face. Morphs were created using Morph 2.5 (Gryphon Software, San Diego, CA). Fig. 2a provides a schematic of the Emotion stimulus set. The same morphing technique was used to create the "Identities" set, except using 3 faces with different *identities* (and neutral emotions) from the Ekman gallery (*see* Fig. 2b). All stimuli were viewed on a monitor with resolution of 1280×1240 and 75 Hz refresh rate.

Table 3
The scores of prosopagnosic participants in the Cambridge Face Perception Test and the Cambridge Face Recognition Test. The average scores of controls, as reported by Bowles et al., 2009 and Duchaine and Nakayama are also reported.

	CFPT		CFMT	
	DP scores	Average control scores	DP scores	Average control scores
DP1	58	34 (12.143)[a]	40	57.92 (7.91)
DP2	56	32 (12.143)[a]	39	57.92 (7.91)
DP3	62	37 (12.143)[a]	40	57.92 (7.91)
DP4	74	44 (12.143)[a]	32	50.7 (8.4785)[a]

[a] These norms reflect age-matched controls from Bowles, McKone, Dawel, Duchaine, and colleagues. They suggest that CMFT scores should be compared against age-matched norms if the participant is over 50 years old at time of testing, and the CFPT should always be compared against age-matched norms (Bowles et al., 2009).
Non-starred scores represent the typical means and standard deviations for the CMFT reported by Duchaine and Nakayama (2006).

2.5. *Design, task and procedure*

The tested dimension (emotion or identity) and the orientation (upright or inverted) were blocked with 200 trials in each of 4 blocks. On each trial participants viewed a study display of 18 faces randomly jittered within a 4×5 grid. Each study display was presented for 1000 ms during which the participants were instructed to form an impression of the "average" emotion (or identity) of the faces presented in the study display.[4] Immediately (one screen-refresh interval) after offset of the study display, a test screen with one face presented at fixation appeared. This test-face was selected at random from the complete set of 147 faces (of identities or emotions). Scrolling the mouse, the participant could change this face, screening through the remaining faces in the set in order to select the face that best represented their estimate of the mean emotional expression or identity on that trial. The choice was reported by mouse click. An ITI of 200 ms separated the response from the next trial (Fig. 3).

Half of the study displays in each block contained face sets that were heterogeneous with respect to emotion (or identity) and the other half contained homogeneous face sets. In the heterogeneous condition, the faces included in each

[4] Due to experimenter error, DP2 viewed the display for 3000 ms during the Emotion Upright condition. DP2 viewed all other conditions for the standard duration (1000 ms).

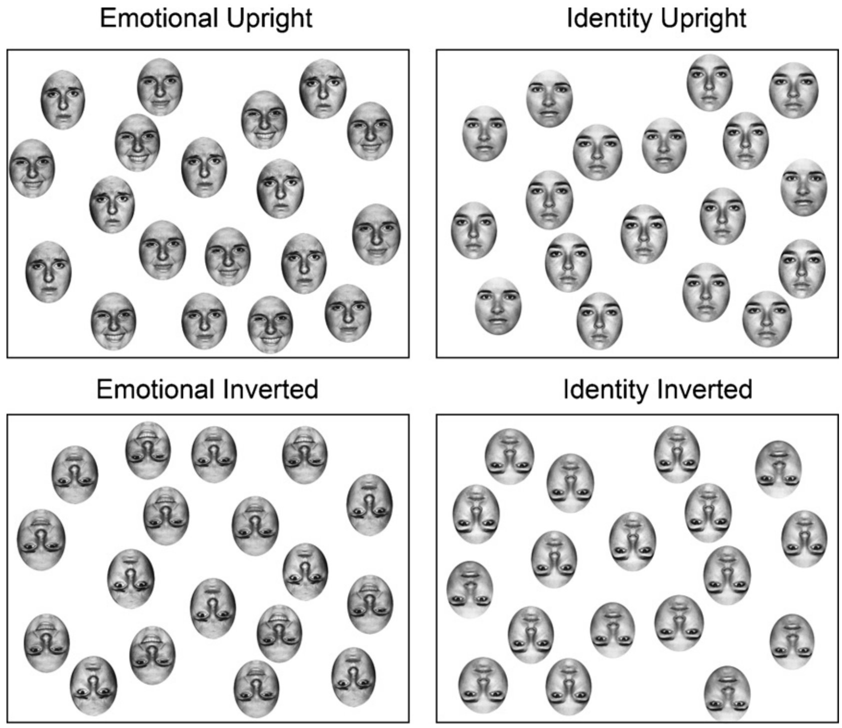

Fig. 4. Examples in four blocked conditions: (A) emotional upright, (B) emotional inverted, (C) identity upright and (D) identity inverted.

study display represented 6 different emotions (or identities), each repeated 3 times. The mean of each study display was chosen randomly on each trial, and the faces comprising the display were 5, 15, and 25 steps away from the mean in either direction within each set. Notably, in the heterogeneous condition, the face representing the actual mean was never included in the display. In the homogeneous condition, all faces in the display were identical. There were 4 counterbalanced blocks: (1) Emotion Upright [EU]: Upright faces varying in emotion (same identities), (2) Emotion Inverted [EI]: Inverted faces varying in emotion (same identities), (3) Identity Upright [IU]: Upright faces varying in identity (neutral emotions), (4) Identity Inverted [II]: Inverted faces varying in identity (neutral emotions). Each prosopagnosic completed the blocks in the same order as their matched controls. Fig. 4 provides an illustration of heterogeneous display and Fig. 5 provides an illustration of a homogeneous display.

2.6. Analysis

First, we defined how far away the participant's responses were from the true mean of the study display and assigned a numerical value on each trial reflecting the difference between the actual and reported means. Recall that the response set contained 147 faces on a circular continuum (of either emotional expression or identity). By analyzing the rectified standard deviation (SD) of the error distribution – the collection of difference scores between the set mean and the participant's response – the degree of tuning to the mean could be estimated for each participant. The SDs were also compared with simulated data to rule out the possibility that participants only sampled 1 face in each display and made their mean judgment on the basis of that face (see below). This analysis was conducted separately for DP and control participants.

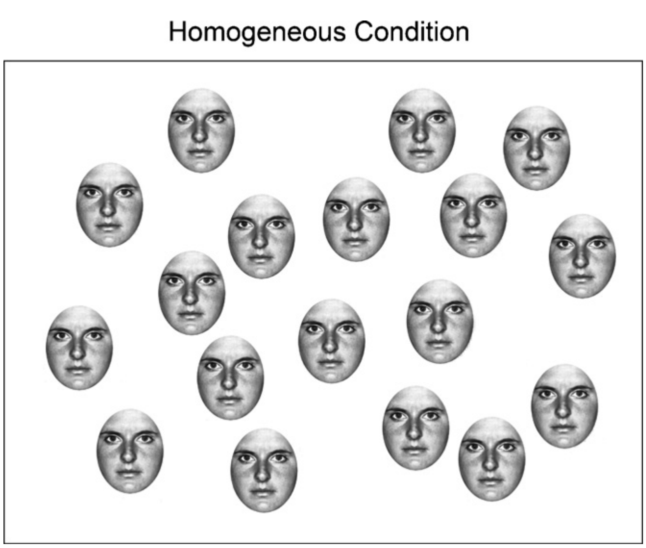

Fig. 5. In the homogeneous condition, participants viewed a set of 18 identical faces.

3. Results

Figs. 6 and 7 show the rectified standard deviation of the error distributions for both controls and DPs during the heterogeneous and homogeneous conditions. The pattern clearly indicates that DP's performance falls well within the distribution of control performance during the heterogeneous condition and suggests that DPs can successfully perform ensemble coding on crowds of faces. Small sample t-tests further confirm that prosopagnosics' performance is similar to control performance. There were no significant differences between prosopagnosic and control performance, except in one case, where the prosopagnosic performed better than her matched control sample (DP1 Emotion Inverted SD = 17.62). Matched control mean SD = 19.90 see Table 4 below.

We further tested whether individual face recognition abilities were correlated with ensemble coding performance by conducting a non-parametric, bi-variate correlational analysis between the performance on the Berkeley Famous Face Test and performance on the ensemble coding tests. We examined controls and prosopagnosics as one group because this combined data set best exhibits variance of performance on the Berkeley Famous Face Test. These scores were not correlated with performance on the ensemble coding tasks during any condition (Emotion-Upright: $r(24) = -.056$, $p > .79$; Emotion-Inverted: $r(24) = .173$, $p > .42$; Identity-Upright: $r(24) = -.150$, $p > 48$; Identity-Inverted: $r(24) = .085$, $p > 69$). Thus, the wide range of performance in individual face discrimination tasks was not correlated with ensemble coding scores, further

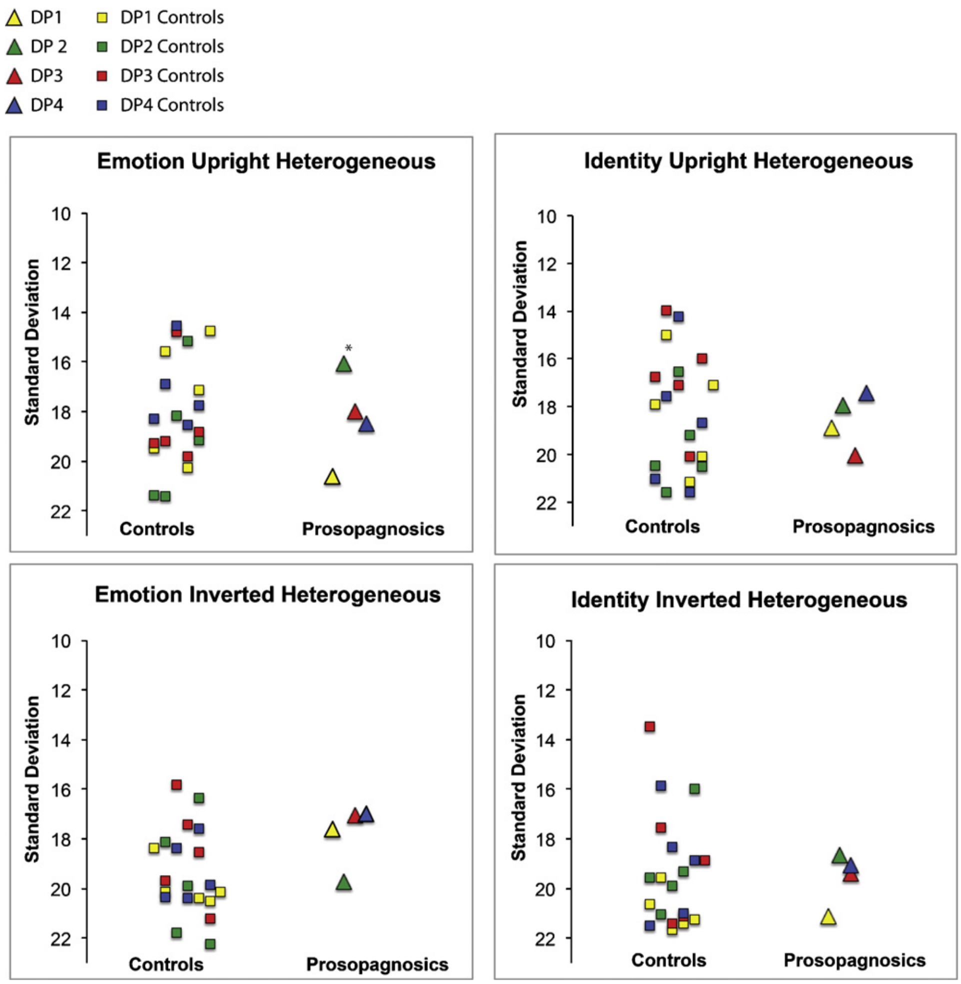

Fig. 6. The rectified standard deviation of the error distribution for the individual participants during the heterogeneous condition. Again, each DP is shown as a single triangle in a given color with their controls shown in the same color as the DP to which they were matched. In contrast to their performance on standardized face tests, DP performance in the experimental tasks is scattered across the range of control performance, indicating that DPs successfully ensemble code crowds of faces. *In this condition, this participant was exposed to the display for 3000 ms. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of the article.)

Table 4
The scores from small sample t-tests for each prosopagnosic compared to their matched control group within each condition of the ensemble coding task. There are no significant differences between controls and prosopagnosics except in one case, where the prosopagnosic performed better than her matched control group.

	DP1 vs. Controls	DP2 vs. Controls	DP3 vs. Controls	DP4 vs. Controls
Emotion Upright	1.46(4), $p = .22$	−1.26(4), $p = .28$	−.20(4), $p = .85$	.88(4), $p = .43$
Emotion Inverted	−2.90(4), $p = .04$*	.01(4), $p = .99$	−.79(4), $p = .47$	−2.00(4), $p = .11$
Identity Upright	.28(4), $p = .79$	−.98(4), $p = .38$	1.61(4), $p = .18$	−.44(4), $p = .68$
Identity Inverted	.31(4), $p = .77$	−.27(4), $p = .79$	.31(4), $p = .77$	−.01(4), $p = .99$

* The prosopagnosic performed better than their matched control group.

dissociating individual face recognition ability from performance on the ensemble coding task. This pattern is not unique to the Berkeley Famous Face Test. All standard face tests were correlated with each other using parametric measures and no test was correlated with the ensemble coding task (see Appendix Table A1). If the ensemble coding task were unreliable, this would contribute to a lack of correlation. To minimize this potential, we conducted a Chronbach's alpha test, which confirmed that the ensemble coding task was reliable, $\alpha = .775$.[5]

One potential account for these results is that both controls and DPs were overwhelmed by the difficulty of the task. To compensate for task difficulty, both groups may potentially choose a face at random from the display rather than engaging in ensemble coding to extract a summary of the crowd. If this scenario were true, the participants would not be processing the group of faces as a whole, but only extracting the features from one face. To ensure that both groups were extracting mean representations and thereby engaging in some manner of integration across the faces in the display, we designed a model that simulated performance based on picking *only* one face in the display. For the simulation, a face was selected at random from the heterogeneous display. Next, noise was added based on each individual's standard deviation in the homogeneous condition. Finally, the program chose a face within the noise distribution and the corresponding value was subtracted from the

[5] We did not run a correlation with the controls alone because performance was at ceiling on the Berkeley Famous Face Test and close to ceiling on the Benton Face Recognition Test.

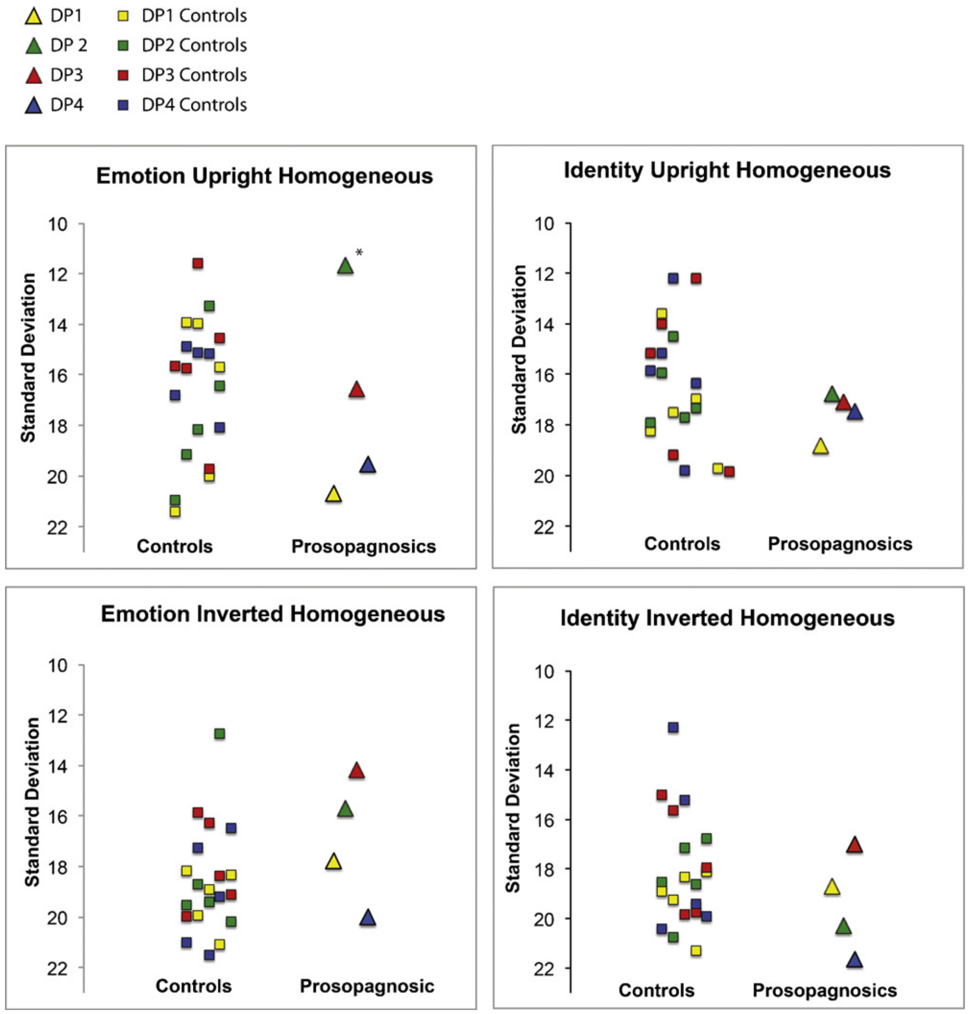

Fig. 7. The rectified standard deviation of the error distribution for the individual participants during the homogeneous condition. *In this condition, this participant was exposed to the display for 3000 ms.

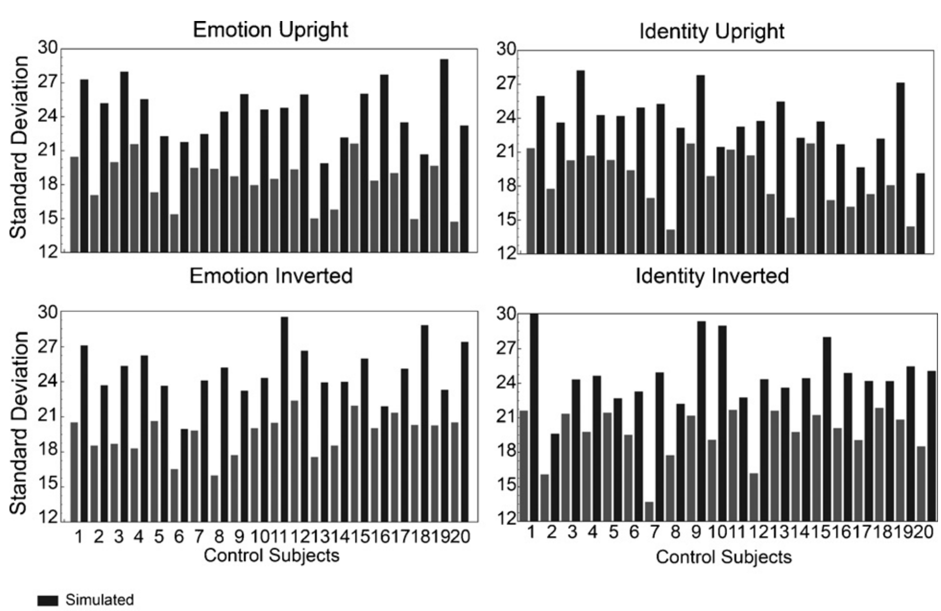

Fig. 8. Controls true performance (grey) compared to the simulated performance (black) for individual participants. Controls performed significantly better than the simulation predicts if judgments were based on picking one face at random from the display. This comparison confirms that controls participants engaged in some sort of ensemble coding.

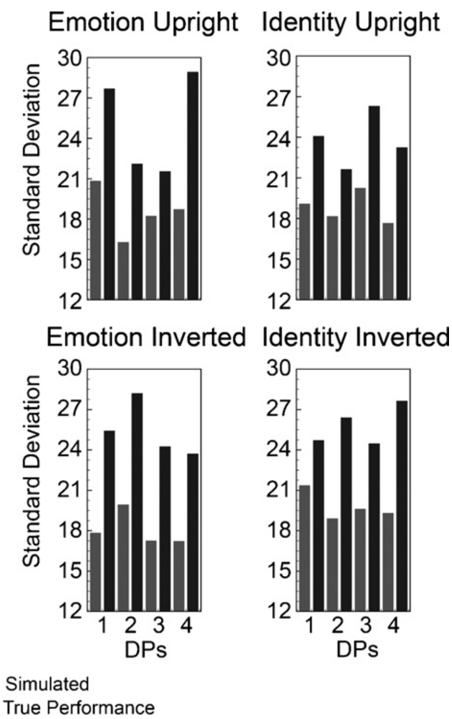

Fig. 9. Prosopagnosics true performance (grey) compared to the simulated performance (black) for individual participants. Controls performed significantly better than the simulation predicts if judgments were based on picking one face at random from the display. This comparison confirms that controls participants engaged in some sort of ensemble coding.

mean of the display. By repeating this Monte Carlo procedure, we obtained a simulated error distribution and the associated standard deviation.

We compared the simulated standard deviation against participants' true performance for both control and DP groups independently. Paired t-tests revealed significant differences between simulated and non-simulated results. The simulated standard deviations for the controls were significantly higher, compared to the participants' true performance (Emotion-Upright: $t(19) = 15.82$, $p < .0001$; Emotion-Inverted: $t(19) = 11.51$, $p < .0001$; Identity-Upright: $t(19) = 10.41$, $p < .0001$; Identity-Inverted: $t(19) = 8.11$, $p < .0001$). Similarly, the DPs simulated standard deviations were also significantly higher compared to each DP's true performance (Emotion-Upright: $t(3) = 4.60$, $p < .02$; Emotion-Inverted: $t(3) = 19.22$, $p < .0002$; Identity-Upright $t(3) = 8.97$, $p < .01$; Identity-Inverted: $t(3) = 5.21$, $p < .01$). The smaller variance in the real data compared to the simulation indicates that performance genuinely reflected ensemble coding and was not an artifact of task difficulty. Fig. 8 shows simulated vs. true performance for controls, while Fig. 9 shows simulated vs. true performance for DPs.

To ensure that our methods were sensitive enough to detect differences across conditions, we conducted a 3 way ANOVA on the control data. The factors were Dimension (emotion, identity), Orientation (upright, inverted), and Study Display Condition (heterogeneous, homogenous). The ANOVA revealed that the SD was significantly larger with inverted compared to upright faces [$F(1,19) = 15.08$ $p < .001$, $\eta p2 = .443$]. The better performance with upright than inverted faces is consistent with reliance on holistic processing in the upright case to achieve ensemble coding and replicates earlier results (Haberman & Whitney, 2009). The Orientation × Dimension interaction was not significant. As expected, the ANOVA also revealed better performance for the homogeneous compared to the heterogeneous condition [$F(1,19) = 57.57$, $p < .001$, $\eta p2 = .752$]. This is not surprising because the homogeneous task should be easier (requires only matching) than the experimental task (requires ensemble coding). All other main effects and interactions did not reach significant levels.

4. Discussion

In the current study, we tested whether DPs' impairments in single face recognition affects their percept of crowd characteristics. Although processing single faces is an important element of social interaction, crowd perception is also an integral aspect of daily experience, both for evaluating the probability that certain individuals might be in a crowd, and for ascertaining the emotional tenor. Indeed, much of daily life is filled with crowd analysis at some level. We engage with crowds during school and work, intermingle with crowds while shopping and attending sporting events, and we view crowds regularly while watching the news, advertisements, and movies.

Based on their difficulty recognizing individual faces and their own reports, one might expect that DPs should be impaired at crowd recognition as well. Many DPs express frustration when navigating crowds; this could be because DPs exhibit deficits in individual face processing which could lead to challenges finding faces in crowds. There are at least two sorts of information that are available while screening crowds. First, there is information about individual identities (e.g., could my child be on the playground?) and second there is information about the crowd as a whole (e.g., are all the children happy?). Whereas DPs have increased difficulty detecting individual faces in crowds, the present results clearly indicate that they have access to the summary statistical information about crowds of faces (expression and identity). Surprisingly, although DPs cannot identify individual faces explicitly, they can rapidly extract the mean identity of the crowd itself, as evidenced by the fact that their accuracy and precision in all conditions was comparable to controls. Our results suggest a counterintuitive phenomenon: Although one of our DPs complained that she struggles to identify her family members in an airport and another finds it difficult to find her own child in a group of children on a playground, DPs are potentially able to incorporate the very same faces they have trouble identifying into an ensemble statistic that represents the crowd as a whole.

While at first glance, it may seem counterintuitive that DPs can perceive the "identity of the crowd", our results reinforce previous findings showing that ensemble coding mitigates the imprecise perception of individual items. The fact that DPs performed no worse in the identity conditions compared to the emotion conditions suggests that ensemble coding can serve as a compensatory mechanism under uniquely impoverished conditions. Typical perceivers rapidly assess the "gist" of a scene, when insufficient time or attention is available to process each item (Alvarez & Oliva, 2009; Haberman & Whitney, 2011). We extend these findings by showing that individuals with developmental prosopagnosia, who by definition exhibit imprecise individual face evaluations, may similarly achieve a veridical ensemble code. How does ensemble coding mitigate imprecise perception? Like any averaging process, ensemble coding is more precise when greater numbers of items are being pooled (assuming that the noise is uncorrelated) and less precise when smaller numbers of items are being ensemble coded (Alvarez, 2011). Thus, it is possible that prosopagnosics benefit from the large number of presented faces.

Consistent with previous research, our results suggest that ensemble coding processes are distinct from individuation processes. Indeed many studies show that successful ensemble coding occurs without face individuation (De Fockert & Wolfenstein, 2009; Haberman & Whitney, 2007, 2011). For example, a face that changes expression might go unnoticed, but the average expression in the crowd can nevertheless be reported with precision (Haberman & Whitney, 2011). Similarly, prosopagnosics show reduced sensitivity to individual faces, but they are capable of forming an ensemble percept of the crowd. The lack of correlation between individual face recognition tests and ensemble coding tasks reinforces the idea that individual face recognition is distinct from the ensemble coding process. While there was strong consistency between individual face recognition tasks, there was no correlation between the sensitivity to individual faces and the sensitivity to crowd expression/identity. These results suggest that the fidelity of the crowd percept is not strictly dependent on the fidelity of the individual face percept. Thus, prosopagnosics may be face blind, but they are crowd-aware.

A common explanation for prosopagnosia is that local features are prioritized and interfere with the analysis necessary to identify the face as a whole (Behrmann & Avidan, 2005). Past experiments on developmental prosopagnosia explored whether local biases interfere with the gestalt perception of a single face; however, no experiment (that we are aware of) investigated whether local preferences interfere with DPs ability to integrate information from *many* faces into a unitary representation. Results from our within-group simulations indicate that our DP participants perceive group characteristics based on the integration of features across at least 2 faces (conservatively). Moreover, similarities between control and DP performance suggests that prosopagnosics' pattern of integration is well within the range of typical perceivers. Thus, whatever precedence may be given to local features by DPs within the crowd, it is clearly not at the expense of the general percept of the crowd as a whole.

It is an essential part of the human experience to evaluate the crowd surrounding us. Prosopagnosics typically describe their percept of a crowd as bewildering. One states, "*I see faces that are human*. . .*but they all look more or less the same*" (Sellers, 2006 as cited in Bakalar, 2006). While anecdotal reports abound, until now, the study of explicit impairments with crowd processing in DP remained empirically untested. Specifically, research in the ensemble coding field focused on normal perceivers and it was unclear whether people with face recognition deficits were capable of calculating the average "face" of the crowd or even the average of multiple items in a display. We show that, like typical perceivers, DPs can rapidly extract a unitary representation of the crowd and that deficits in individual face discrimination do not diminish the accuracy of the ensemble code.

Acknowledgements

This research was supported by the Veterans Administration, National Institutes of Health Grant MH-0164458 (to L.C.R. & SB) and NSF 074689 (DW). Lynn C. Robertson has a Senior Research Career Scientist award from the Veterans Administration and is affiliated with the VA Clinical Sciences Research Service, Department of Veterans Affairs Medical Center, Martinez, CA. We would like to thank Joe Degutis for providing previous test results and Jason Haberman for providing Matlab code. Also, we would like to express appreciation to research assistants Jennifer Dainer, Nicola Bernard, and Ashli Butler for helping to test participants. Finally, we are grateful to Saeideh Ghahghaei for helpful discussion.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.neuropsychologia.2012.03.026.

References

Alvarez, G. A. (2011). Representing multiple objects as an ensemble enhances visual cognition. *Trends in Cognitive Sciences*, *15*(3), 122–131.

Alvarez, G. A., & Oliva, A. (2009). Spatial ensemble statistics are efficient codes that can be represented with reduced attention. *Proceedings of the National Academy of Sciences*, *106*(18), 7345–7350.

Ariely, D. (2001). Seeing sets: Representation by statistical properties. *Psychological Science*, *12*(2), 157–162.

Asprin, G. (2011, January 22). *Face off: Living with prosopagnosia* [A friend among strangers: A room of friends, or an isolation chamber?]. Retrieved from: <http://www.psychologytoday.com/blog/face/201101/friend-among-strangers>.

Avidan, G., Tanzer, M., & Behrmann, M. (2011). Impaired holistic processing in congenital prosopagnosia. *Neuropsychologia*, *49*(9), 2541–2552.

Bakalar, N. (2006, February 18). Just another face in the crowd, Indistinguishable Even if It's Your Own. *The New York Times*, Retrieved from: <http://www.nytimes.com/2006/07/18/health/psychology/18face.html>.

Behrmann, M., & Avidan, G. (2005). Congenital prosopagnosia: Face-blind from birth. *Trends in Cognitive Sciences*, *9*(4), 180–187.

Behrmann, M., Avidan, G., Marotta, J. J., & Kimchi, R. (2005). Detailed exploration of face-related processing in congenital prosopagnosia: 1. Behavioral findings. *Journal of Cognitive Neuroscience*, *17*(7), 1130–1149.

Bentin, S., Deouell, L. Y., & Soroker, N. (1999). Selective visual streaming in face recognition: Evidence from developmental prosopagnosia. *Neuroreport*, *10*(4), 823–827.

Bentin, S., DeGutis, J. M., D'Esposito, M., & Robertson, L. C. (2007). Too many trees to see the forest: Performance, event-related potential, and functional magnetic resonance imaging manifestations of integrative congenital prosopagnosia. *Journal of Cognitive Neuroscience*, *19*, 132–146.

Benton, A. L., & Van Allen, M. W. (1968). Impairment in facial recognition in patients with cerebral disease. *Cortex*, *4*, 344–358.

Bowles, D., McKone, E., Dawel, A., Duchaine, B., Schmalzl, L., Palermo, R., et al. (2009). Diagnosing prosopagnosia: Effects of aging and participant-stimulus ethnic match on the Cambridge Face Memory Test and Cambridge Face Perception Test. *Cognitive Neuropsychology*, *26*, 423–455.

Chong, S. C., & Treisman, A. (2003). Representation of statistical properties. *Vision Research*, *43*(4), 393–404.

Crawford, J. R., Garthwaite, P. H., & Howell, D. C. (2009). On comparing a single case with a control sample: An alternative perspective. *Neuropsychologia*, *47*(13), 2690–2695.

De Fockert, J., & Wolfenstein, C. (2009). Rapid extraction of mean identity from sets of faces. *Quarterly Journal of Experimental Psychology*, *62*(9), 1716–1722.

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

Dobel, C., Bölte, J., Aicher, M., & Schweinberger, S. R. (2007). Prosopagnosia without apparent cause: Overview and diagnosis of six cases. *Cortex*, *43*(6), 718–733.

Duchaine, B. C., Parker, H., & Nakayama, K. (2003). Normal recognition of emotion in a prosopagnosic. *Perception*, *32*, 827–838.

Duchaine, B. C., & Nakayama, K. (2004). Developmental prosopagnosia and the Benton Facial Recognition Test. *Neurology*, *62*(7), 1219–1220.

Duchaine, B., & Nakayama, K. (2006). The Cambridge Face Memory Test: Results for neurologically intact individuals and an investigation of its validity using inverted face stimuli and prosopagnosic participants. *Neuropsychologia*, *44*(4), 576–585.

Duchaine, B., Yovel, G., & Nakayama, K. (2007). No global processing deficit in the Navon task in 14 developmental prosopagnosics. *Social Cognitive and Affective Neuroscience*, *2*(2), 104–113.

Ekman, P., & Friesen, W. V. (1976). Measuring facial movement. *Environmental Psychology and Nonverbal Behavior*, *1*(1), 56–75.

Haberman, J., & Whitney, D. (2007). Rapid extraction of mean emotion and gender from sets of faces. *Current Biology*, *17*(17), R751–R753.

Haberman, J., & Whitney, D. (2009). Seeing the mean: Ensemble coding for sets of faces. *Journal of Experimental Psychology, Human Perception and Performance*, *35*(3), 718–734.

Haberman, J., Harp, T., & Whitney, D. (2009). Averaging facial expression over time. *Journal of Vision*, *9*(11), 1–13.

Haberman, J., & Whitney, D. (2011). Efficient summary statistical representation when change localization fails. *Psychonomic Bulletin & Review*, *18*, 855–859.

Humphreys, K., Avidan, G., & Behrmann, M. (2007). A detailed investigation of facial expression processing in congenital prosopagnosia as compared to acquired prosopagnosia. *Experimental Brain Research*, *176*(2), 356–373.

Jones, R. D., & Tranel, D. (2001). Severe Developmental Prosopagnosia in a Child With Superior Intellect. *Journal of Clinical and Experimental Neuropsychology*, *23*(3), 265–273.

Lee, Y., Duchaine, B., Wilson, H. R., & Nakayama, K. (2010). Three cases of developmental prosopagnosia from one family: Detailed neuropsychological and psychophysical investigation of face processing. *Cortex*, *46.8*(46), 949–964.

Le Grand, R., Cooper, P. A., Mondloch, C. J., Lewis, T. L., Sagiv, N., de Gelder, B., et al. (2006). What aspects of face processing are impaired in developmental prosopagnosia? *Brain and Cognition*, *61*(2), 139–158.

Lobmaier, J., Bölte, J., Mast, F., & Dobel, C. (2010). Configural and featural processing in humans with congenital prosopagnosia. *Advances in Cognitive Psychology*, *6*(-1), 23–34.

Palermo, R., Willis, M. L., Rivolta, D., McKone, E., Wilson, C. E., & Calder, A. J. (2011). Impaired holistic coding of facial expression and facial identity in congenital prosopagnosia. *Neuropsychologia*, *49*(5), 1226–1235.

Schmalzl, L., Palermo, R., Green, M., Brunsdon, R., & Coltheart, M. (2008). Training of familiar face recognition and visual scan paths for faces in a child with congenital prosopagnosia. *Cognitive Neuropsychology*, *25*(5), 704–729.

Warrington, E. K. (1984). *Recognition memory test*. Windsor, England: NFER-Nelson.