BRAIN
A JOURNAL OF NEUROLOGY

Holistic face training enhances face processing in developmental prosopagnosia

Joseph DeGutis,[1,2] Sarah Cohan[2] and Ken Nakayama[2]

1 Geriatric Research Education and Clinical Centre (GRECC), Boston Division VA Healthcare System, Jamaica Plain, MA, USA
2 Vision Sciences Laboratory, Department of Psychology, Harvard University, USA

Correspondence to: Joseph DeGutis,
Geriatric Research Education and Clinical Centre (GRECC),
Boston VA Healthcare System,
Boston, MA 02130
E-mail: degutis@wjh.harvard.edu

Prosopagnosia has largely been regarded as an untreatable disorder. However, recent case studies using cognitive training have shown that it is possible to enhance face recognition abilities in individuals with developmental prosopagnosia. Our goal was to determine if this approach could be effective in a larger population of developmental prosopagnosics. We trained 24 developmental prosopagnosics using a 3-week online face-training program targeting holistic face processing. Twelve subjects with developmental prosopagnosia were assessed before and after training, and the other 12 were assessed before and after a waiting period, they then performed the training, and were then assessed again. The assessments included measures of front-view face discrimination, face discrimination with view-point changes, measures of holistic face processing, and a 5-day diary to quantify potential real-world improvements. Compared with the waiting period, developmental prosopagnosics showed moderate but significant overall training-related improvements on measures of front-view face discrimination. Those who reached the more difficult levels of training ('better' trainees) showed the strongest improvements in front-view face discrimination and showed significantly increased holistic face processing to the point of being similar to that of unimpaired control subjects. Despite challenges in characterizing developmental prosopagnosics' everyday face recognition and potential biases in self-report, results also showed modest but consistent self-reported diary improvements. In summary, we demonstrate that by using cognitive training that targets holistic processing, it is possible to enhance face perception across a group of developmental prosopagnosics and further suggest that those who improved the most on the training task received the greatest benefits.

Introduction

Prosopagnosia is a deficit in the ability to perceive, learn and recognize faces, and can result from genetic and developmental causes as well as acute brain injury (Duchaine and Nakayama, 2006a). Recent studies suggest that prosopagnosics may be particularly deficient at building a holistic face representation sufficient for face identification (Bukach *et al.*, 2006; Ramon and Rossion, 2010; Avidan *et al.*, 2011; Palermo *et al.*, 2011; DeGutis *et al.*, 2012b). To identify individuals, prosopagnosics commonly rely on less reliable cues such as situational context, an individual's hair, voice and gait. Despite using these cues,

prosopagnosics are often still left with recognition deficits that may significantly impact their everyday social functioning. Long-term consequences could include a restricted social circle, more limited employment opportunities, and loss of self-confidence (Yardley *et al.*, 2008). Because of these potentially debilitating consequences and the high prevalence of prosopagnosia (up to 1 in 50 in the general population for developmental prosopagnosia; Kennerknecht *et al.*, 2006, 2008), developing effective rehabilitation programmes to enhance face recognition is an important endeavour.

Despite several attempts to improve face recognition in subjects with prosopagnosia, success has been limited. Approaches have varied, with some studies promoting 'normal' mechanisms of face processing (e.g. holistic processing, DeGutis *et al.*, 2007) and others fostering compensatory strategies (e.g. verbal strategies and feature-based processing, Brunsdon *et al.*, 2006; Schmalzl *et al.*, 2008). Additionally, some studies have focused on enhancing recognition of personally relevant faces (Mayer and Rossion, 2007) whereas others have focused on training more general skills that can be applied to any face (DeGutis *et al.*, 2007). The earliest reported prosopagnosia rehabilitation attempt was an acquired case (Beyn and Knyazeva, 1962) and since then, there have been approximately a dozen rehabilitation studies of acquired prosopagnosics (e.g. Ellis and Young, 1988; Polster and Rapcsak, 1996; Mayer and Rossion, 2007). Although there has been some evidence that acquired prosopagnosics can become better at recognizing personally familiar and famous faces (Mayer and Rossion, 2007), there is little evidence for training-related improvements in general face recognition abilities.

The more recent developmental prosopagnosia rehabilitation literature has shown slightly more promising results (Brunsdon *et al.*, 2006; DeGutis *et al.*, 2007; Schmalzl *et al.*, 2008), possibly due to developmental prosopagnosics having a more intact face processing infrastructure than acquired prosopagnosics. Brunsdon *et al.* (2006) published the first reported attempt to rehabilitate a developmental prosopagnosic and demonstrated improvements in an 8-year-old participant (AL) using compensatory 'feature naming' training. In particular, AL was taught to perceive, discuss and remember five experimenter-defined distinctive facial characteristics of 17 personally familiar faces. The first two characteristics were always age and gender (which AL could likely recognize) (Chatterjee and Nakayama, 2012; DeGutis *et al.*, 2012a) and the other three characteristics were distinctive facial features such as 'long thin face', 'wide nostrils', 'high curved eyebrows', 'wrinkles around the eyes', and 'freckles'. After 14 practice sessions over 1 month, AL showed improved recognition of the trained target faces that generalized to novel views of the target faces and versions of the faces without hair, as well as anecdotal real-life improvements of recognizing these faces. Using the same training procedure with a 4-year-old developmental prosopagnosic, Schmalzl *et al.* (2008) showed improvements in recognizing target faces and importantly, a more normal pattern of eye movement scan paths that generalized to normal scanning of untrained faces. Together, these results suggest that by training compensatory mechanisms in children with developmental prosopagnosia, it is possible to enhance recognition of trained faces and that this may lead to more normal patterns of face processing

(but see Dalrymple *et al.*, 2012 where similar training in a 10-year-old was unsuccessful). Despite these positive results, this compensatory approach is slow to learn, challenging for the participant to implement, and requires intensive involvement from a therapist or caregiver.

In contrast to this compensatory approach, DeGutis and colleagues (2007) demonstrated that training targeting holistic face processing was successful at improving general face recognition abilities in an adult developmental prosopagnosic, Subject MZ. Based on the idea that prosopagnosics can apply some limited holistic processing to faces (Barton *et al.*, 2003; Bukach *et al.*, 2006; DeGutis *et al.*, 2012b), this training program aimed to expand prosopagnosics' spatial integration abilities by requiring speeded judgements that rely on combining eyebrow/eye and mouth/nose spacing information (Fig. 1). Though some researchers may suggest that this training particularly targets second order configural processing (processing the subtle relations amongst features) (Maurer *et al.*, 2002), we believe that second order configural processing is simply one aspect of a more general holistic face processing mechanism (Rossion, 2008). Considering this, we suggest that this training programme enhances more general holistic face processing mechanisms rather than just second order configural processing. We had MZ perform this training task for several months and afterwards she showed improvements on tests of general face recognition and also experienced daily life improvements. After training, she additionally demonstrated a more normal pattern of event-related potential selectivity, showing a greater N170 (an occipito-temporal potential normally selective to faces and thought to reflect holistic face processing; Jacques and Rossion, 2009) to faces than objects, and enhanced functional MRI connectivity within right hemisphere face-selective regions during face viewing. This suggests that it is possible to enhance face recognition in an adult with developmental prosopagnosia using a remedial approach and that this approach can potentially enhance signatures of normal face processing. A particular advantage of this approach is that the skill learned can be applied to new faces and does not require explicit use of strategies.

Although the above demonstrations are proof of the principle that enhancing face recognition in developmental prosopagnosics is possible, it is still uncertain whether face recognition can be enhanced in a wider, more general population of developmental prosopagnosics. It could be that these few successful cases exist amongst failed attempts, but due to publication bias such failed attempts have gone unreported. It could also be that certain subgroups of developmental prosopagnosics respond to rehabilitation where others do not. To determine if improved face recognition in the general developmental prosopagnosic population or a certain subgroup of developmental prosopagnosics is possible, it is imperative to train a larger, more representative sample. Besides the sample size issue, another outstanding question is the degree of generalization of the training. Previous studies either concentrated on assessments that were similar to the training or included tests with potentially substantial practice effects (e.g. repeating the same memory test). In addition, previous studies did not include a broad array of tasks and stimuli, and/or did not formally assess everyday face recognition. Assessing developmental prosopagnosics on a more comprehensive battery that includes theoretically
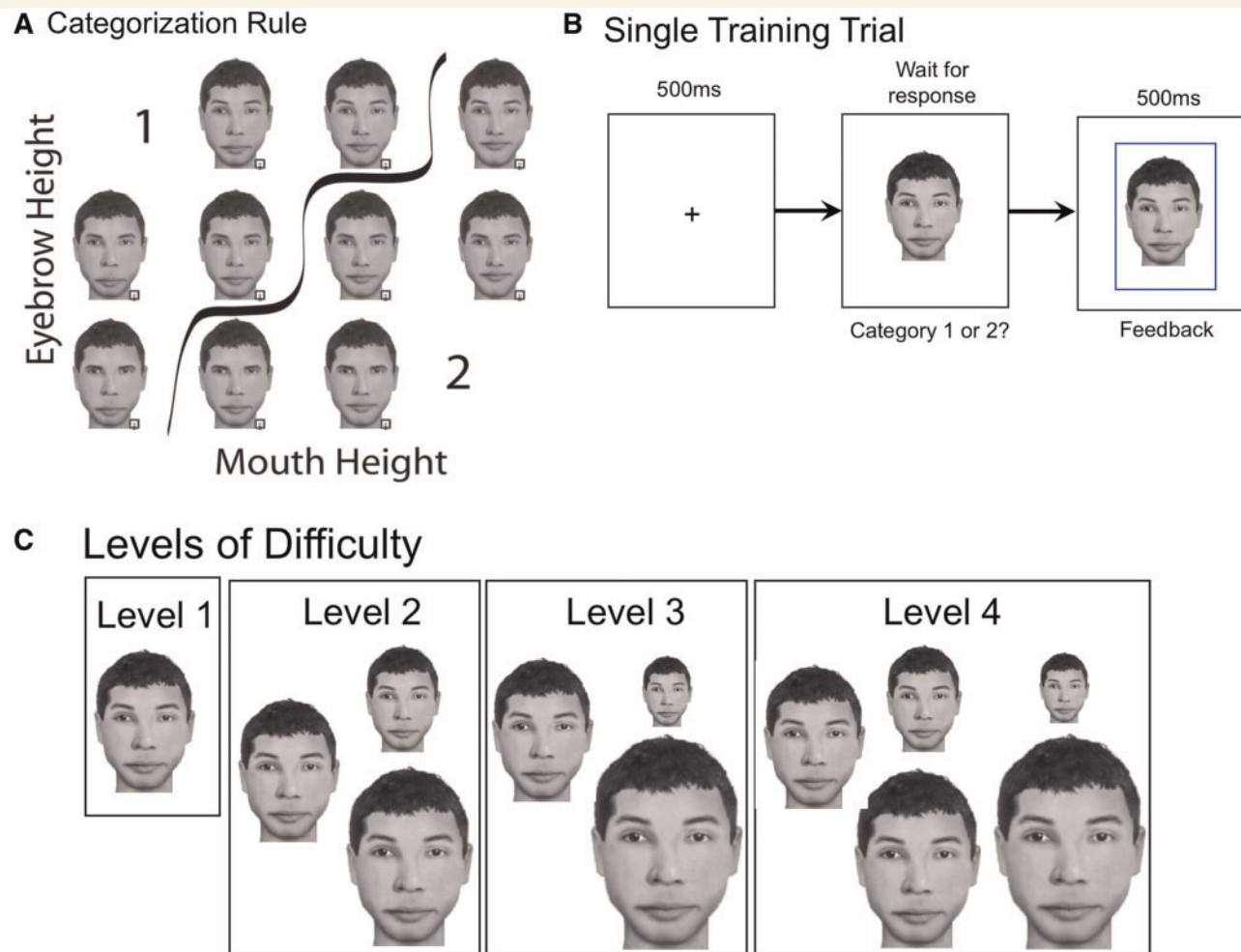
Figure 1 Examples of stimuli from the training task. (**A**) A matrix image of one template training face depicting the categorization rule: faces with higher eyebrows and lower mouths are in Category 1 and those with lower eyebrows and higher mouths are in Category 2. (**B**) Example of a single self-paced training trial with feedback indicating the trial was correct. (**C**) The face stimuli sizes used in each different difficulty level of training. Level 1 included faces $3.5° \times 5.2°$ in size (100% size). Level 2 included one-third of the stimuli at 100%, one-third at $2.6° \times 3.9°$ (75%), and one-third at $4.4° \times 6.6°$ (125%). Level 3 included one-third of the stimuli at 100%, one-third of the stimuli at a visual angle of $1.8° \times 2.6°$ (50%), and one-third of the stimuli at a visual angle of $5.3° \times 7.9°$ (150%). Level 4 included all five previous stimulus sizes distributed equally.

meaningful face processing tasks (e.g. measures of holistic face processing and discriminating faces across view-points) and self-report measures would allow a more thorough characterization of potential training-related improvements.

To address these outstanding issues, we recruited a group of 24 developmental prosopagnosics and had them perform a 3-week online face-training program modelled on DeGutis *et al.* (2007). Because piloting showed that this training program resulted in a skill that was relatively specific to the size of the trained face images, the current version of training allowed for different sizes of faces as training success increased (Fig. 1).

To determine the effects of training, we employed a waiting list control design in which 12 developmental prosopagnosics were assessed before and after training, while 12 developmental prosopagnosics were assessed before and after a waiting period, then performed the training, and were assessed again. This allowed us to not only assess the effects of training compared to a test–retest control, but also to assess the consistency of the training effect by seeing if the 'waiting list' group showed the same pattern of improvement after training as the 'training only' group. Finally, to measure potential training-related improvements, we administered assessments of four important domains of face processing: (i) front-view face discrimination; (ii) face discrimination from differing view-points; (iii) holistic face processing; and (iv) self-reported everyday face processing ability. Front-view face discrimination tasks measured generalization of training to new stimuli and task formats that were significantly different from the training procedure, whereas tasks involving varying view-points tested whether training (which only uses front-view faces) generalized to discrimination of faces from different view-points (this has shown to be particularly challenging for prosopagnosics; Marotta *et al.*, 2002; Lee *et al.*, 2010). To test our proposal that training specifically

enhances holistic face processing, we included classic measures of holistic processing (part-whole and face inversion). We also included a 5-day diary to measure generalization to real-world improvements.

Materials and methods

Participants

Twenty-four developmental prosopagnosics (14 female, seven females/group) with an average age of 35.38 years [standard deviation (SD) = 9.76 years] participated in the study and were randomly assigned to either the experimental or waiting list control group (Table 1).

When recruiting participants, we first began with a pool of ~4500 individuals who completed a survey at www.faceblind.org and complained of face recognition problems. This large group was then further pared down (Fig. 2). In particular, based on the following questions that our laboratory has found to reflect daily life face recognition abilities and proximity to location in Cambridge, MA, USA, we filtered the pool of 4500 down to 143 potential participants (participants' responses from 1–5, never = 1, always = 5): 1) I treat

strangers as if I know them to avoid offending people I might already know; 2) I find it hard to keep track of characters in TV shows or movies; 3) I try to remember non-facial information about people's appearance; 4) I can recognize family members and close friends out of context; 5) I can visualize the faces of family members and close friends. To be eligible for participation, Q1 + Q2 + Q3 − Q4 − Q5 + 12 must be ⩾17 [we have found this to be indicative of significant impairment on the Cambridge Face Memory Test (CFMT), see below]. Next, we contacted the eligible participants via email and set up telephone interviews with those participants who met our criteria.

Any participant who had been diagnosed with or suspected themselves of having Asperger's syndrome or autism, or who stated they had any difficulty recognizing emotions from faces, was given the Autism Spectrum Quotient questionnaire (Baron-Cohen *et al.*, 2001). Those who scored above a clinical cut-off of 32 on the Autism Spectrum Quotient questionnaire were excluded (*n* = 6). We also excluded those who were no longer interested in participating in our cognitive training study (*n* = 6), those whose clinical interviews suggested their deficits were relatively minor (*n* = 6), and those who scored better than 1.7 SD below the mean on the Cambridge Face Memory Test (*n* = 13). As in previous developmental prosopagnosic studies (DeGutis *et al.*, 2012b, 2013), we chose 1.7 SD because we recognize the limitations of the CFMT in that some people report symptoms of severe prosopagnosia, but score close to the normal

Table 1 Subject Demographics, raw scores for the CFMT and CFPT with z-scores in brackets, accuracy scores on the FOBPT, and assignment to experimental training-only or waiting list control group

Subject	Age	Sex	CFMT	CFPT	Faces	Objects	Bodies	Group
501	21	F	36 (−2.77)	46 (−0.76)	68%	84%	81%	Training
502	48	M	28 (−3.78)	80 (−3.55)	65%	99%	75%	Training
505	35	F	37 (−2.65)	70 (−2.73)	70%	90%	78%	Training
509	24	M	40 (−2.27)	52 (−1.25)	75%	94%	94%	Training
514	35	F	35 (−2.9)	66 (−2.4)	68%	79%	64%	Training
515	47	F	34 (−3.03)	60 (−1.91)	59%	85%	79%	Training
519	19	M	44 (−1.76)	52 (−1.25)	70%	91%	84%	Training
521	32	M	36 (−2.77)	70 (−2.73)	80%	94%	91%	Training
527	33	F	38 (−2.52)	78 (−3.39)	66%	92%	89%	Training
530	39	F	31 (−3.4)	56 (−1.58)	76%	90%	66%	Training
531	40	M	37 (−2.65)	76 (−3.22)	63%	83%	78%	Training
533	38	F	36 (−2.77)	54 (−1.42)	63%	86%	88%	Training
503	22	M	31 (−3.4)	76 (−3.22)	58%	89%	76%	Waiting list
506	35	M	27 (−3.91)	80 (−3.55)	61%	84%	86%	Waiting list
507	32	M	41 (−2.14)	62 (−2.07)	56%	84%	73%	Waiting list
508	46	F	38 (−2.52)	54 (−1.42)	69%	80%	84%	Waiting list
512	52	F	32 (−3.28)	92 (−4.53)	61%	70%	76%	Waiting list
516	44	F	21 (−4.67)	54 (−1.42)	54%	94%	78%	Waiting list
522	49	M	29 (−3.66)	44 (−0.6)	70%	93%	80%	Waiting list
523	27	F	30 (−3.53)	72 (−2.89)	58%	94%	85%	Waiting list
524	29	F	36 (−2.77)	68 (−2.57)	76%	94%	86%	Waiting list
525	47	M	42 (−2.01)	62 (−2.07)	71%	90%	80%	Waiting list
528	24	F	42 (−2.01)	38 (−0.11)	56%	95%	70%	Waiting list
529	31	F	41 (−2.14)	64 (−2.24)	76%	94%	81%	Waiting list
Mean	34.25		36.0	63.33	68%	89%	80%	Training
SD	9.26		4.05	11.52	6%	6%	9%	
Mean	36.50		34.17	63.83	64%	88%	80%	Waiting list
SD	10.53		6.86	15.22	8%	8%	5%	

Raw scores for the CFMT and CFPT (with z-scores in brackets), accuracy scores on the FOBPT, and assignment to experimental training-only or waiting list control group. On the CFPT a lower score indicates better performance.
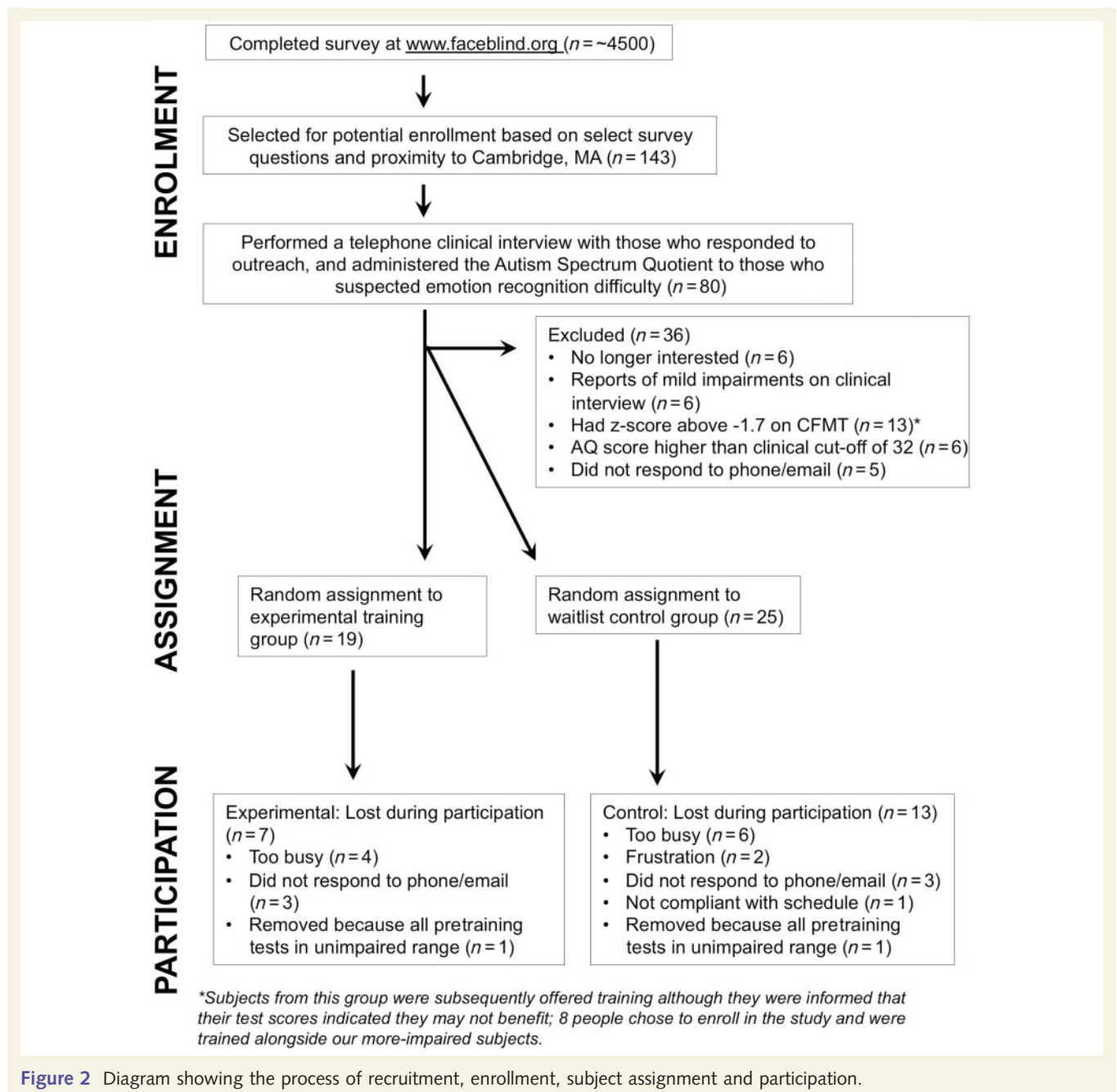
Figure 2 Diagram showing the process of recruitment, enrollment, subject assignment and participation.

range on the test. Because of this discrepancy, we wanted to be slightly more inclusive than a cut-off of 2.0 SD below the mean. After enrollment 18 participants chose not to continue participation in the study. Two subjects who, despite impaired scores on the CFMT, scored in the unimpaired range on almost all of the pretraining assessments, and were subsequently removed from the study. This resulted in 24 participants who successfully completed the training protocol in the experimental ($n = 12$) and waiting list control ($n = 12$) groups.

These 24 subjects were all tested and trained entirely online. When comparing our sample of 24 developmental prosopagnosics to a separate set of 19 similarly aged developmental prosopagnosic subjects, independent t-tests showed no significant differences in CFMT scores. Also supporting our use of an online sample, Wilmer et al. (2010) and Germine et al. (2012) have shown online testing to yield the same

mean performance and reliability as in-lab testing. Because of the challenge of completing this 4–6 week protocol remotely, great effort was made with each participant to ensure that they were motivated and compliant, including several email exchanges and occasional phone calls. Once consent for enrollment was received, subjects were oriented to the study procedure through a set of web pages, emails, and phone calls with the study coordinator. Details are available in the online Supplementary material.

Diagnostic assessment

Cambridge Face Memory Test

The CFMT is a widely used test of face recognition ability and an established method for determining the severity of

developmental prosopagnosia (Duchaine and Nakayama, 2006*b*; Lee *et al.*, 2010).

Repeated assessments

Part-Whole Face Task

We used the Part-Whole Face Task (PWFT) to both assess short-term memory for whole faces and to assess the ability to use the face context when discriminating changes in individual facial features (i.e. holistic face processing). After seeing a target face (e.g. Roger's face), unimpaired subjects demonstrate a subsequent memory test advantage for discriminating a feature change (e.g. discriminating Roger's nose from Ken's nose) when features are shown within the context of the target face compared to when discriminating features shown in isolation. We used the task from Tanaka *et al.* (2004) and further details are available in the Supplementary material.

Although both the training task and part-whole task involve discriminating changes to the eye and mouth regions and attending to both areas, there are several differences between these tasks that make the PWFT a good measure of transfer of training: (i) the use of different face stimuli (computer-generated faces for training versus composite real year-book faces in the PWFT); (ii) distinct task formats (categorization task for training versus short-term memory task in the PWFT); (iii) the nature of the discrimination task is different (discriminating feature spacing in training versus feature identities in PWFT); and (iv) differential memory demands (minimal long-term memory demands in training versus higher short-term memory demands in the PWFT).

Face Working Memory Task

The Face Working Memory Task (FWMT) is an adaptive measure designed by Garga Chatterjee and Ken Nakayama at Harvard University, used to estimate the number of faces (upright and inverted) that one can hold in visual working memory. In addition to scores on upright trials, by comparing upright and inverted performance we were also able to obtain a measure of the face inversion effect, which is thought to reflect face-specific configural and holistic processing. The stimuli used were a random assortment of 3241 computer-generated faces, created using GenHead software (GenHead by Genemation, www.genemation.com). A trial began by showing two repeated sequences of target faces to the participant. Each face was shown for 500 ms with a 200 ms interstimulus interval, followed by a 500 ms scrambled mask image. The participant was then presented with three stimuli simultaneously for 3000 ms (labelled 1, 2 and 3) in which they were asked to choose a target face from two foils by pressing 1, 2 or 3 on the keyboard. The subject was prompted for a response and after their response was recorded, feedback was provided and they pressed the spacebar to continue to the next trial.

To obtain an estimate of the average number of faces that an individual can hold in working memory, a staircase approach was implemented. In particular, once a participant provided correct responses on three trials in a row, the repeated sequences then consisted of two different face stimuli. At that point, if they again provided three correct responses in a row, the sequences consisted of three different face stimuli. The test progressed in this staircase pattern, but at any time should an incorrect response be given, the number of faces in the sequences was decreased by one. The dependent measure was proportional to the mean number of upright and inverted faces that the participant was shown.

Philadelphia Face Similarity Test

The Philadelphia Face Similarity Test (PFST) was modified from Thomas *et al.* (2008a). In this task, participants were presented with a central target face above two test faces and were asked to choose which test face was most similar to the target face. The dependent variable was how accurate participants were at judging face similarity. To better isolate face-specific processing mechanisms, we took the original version and removed colour information and tightly cropped the faces to prevent the use of colour and outward features as matching cues. Additionally, in an effort to make the task slightly more challenging, in our version the target face disappeared after 3 s instead of waiting for the subjects' response. Further details are provided in the Supplementary material.

Cambridge Face Perception Test

The Cambridge Face Perception Test (CFPT) is a computerized sorting task in which participants arrange six front-view faces according to their similarity to a three-quarter view target face (Duchaine *et al.*, 2007). Participants completed eight upright sorting trials. The dependent variable is the sum of the deviations from the correct order across all sorting trials. Further details are given in the Supplementary material.

Face-Objects-Bodies Perception Task

The Faces-Objects-Bodies Perception Task (FOBPT) is a same/different sequential matching task from Pitcher *et al.* (2009). It was used as a measure of face discrimination across views as well as a measure of non-face perceptual processing. In particular, participants were shown a target stimulus for 500 ms, a visual mask for 500 ms, and then were shown a probe stimulus for 500 ms and were asked to indicate if the probe was the same or different from the target. The dependent variable is per cent correct. Faces and bodies were presented from different views on each trial (but not objects, see below), but on any particular trial the two views are the same. Further details are provided in the Supplementary material.

For face blocks, 10 faces (varied in gender and viewing angle) were created using FaceGen software (Singular Inversions), and the component parts of these faces (eyes, mouth and nose) were then individually altered to create a second face. Each face pair was then used to create a morph series. For the different trials, the percentage morph difference between the two images was 50% (10 trials), 80% (20 trials), or 100% (10 trials). For object trials, a set of novel objects (Multipart Geon Objects) was downloaded from http://wiki.cnbc.cmu.edu/Novel_Objects. Each pair used for morphing was comprised of two visually similar objects seen from the same viewing angle that had the same overall shape, but varied in local details. For different trials the percentage difference between the two images was 20% (three trials), 30% (14 trials), 50% (six trials), 80% (seven trials), or 100% (10 trials). For bodies, 10 pairs of male bodies (varied in corpulence and muscle tone) wearing white shorts and seen from different viewing angles were created using Poser software (Smith Micro, Inc). Adobe Photoshop was used to remove the head. Body pose was the same for both images in each trial. For different trials the percentage difference between the two images was 50% (10 trials), 80% (20 trials), or 100% (10 trials).

Face diary

Both before the first assessment and after the completion of training, subjects were instructed to fill out an online diary form at the end of five consecutive days (Supplementary material). The diary form was structured to assess subjective daily experiences with face recognition, and included questions about levels of face recognition confidence and

anxiety (on a scale of 1–10); social avoidance (scale of 1–4); number of successful face recognitions of personally-known individuals and famous individuals (in person, photos or videos); use of alternative strategies for recognition; mental imagery of faces; face recognition failures; and open text boxes to report their positive and negative daily experiences, as well as their activities/encounters throughout the day. The form was explicitly discussed with each subject before diary entry began, and particular care was paid to their understanding of the difference between recognizing someone successfully by their face, by alternative measures, and what it meant to fail to recognize someone familiar. After subjects' first diary entry, their responses were reviewed carefully to ensure they understood how to use the form, and feedback was provided if necessary. Because of the timing of the protocol, participants in the waiting list group completed the diary before the waiting period and after they completed training (third assessment session), but did not complete diary entries after their second assessment session. Diary entries were skipped at this time so that, similar to the training only group, the waiting list group could start training immediately after their second assessment session.

Training procedure

The rationale behind the training procedure is that prosopagnosics can apply some holistic processing to faces, but only over a spatially limited area (Barton *et al.*, 2003; DeGutis *et al.*, 2012b) and that this perceptual limitation is a major contributor to prosopagnosics' face recognition deficits. Therefore, the aim of the current training was to enhance prosopagnosics' ability to perceive internal feature spacing information across a greater spatial extent of the face. To accomplish this, we designed a challenging task requiring prosopagnosics to make category judgements based on integrating two vertical feature spacings: the distance between the eye and eyebrows, and between the mouth and nose (Fig. 1A). Performing this task quickly and accurately (>93% accuracy and <1 s reaction time) is demanding and requires one to perceive the overall configuration of the mouth/nose/eyes/eyebrows. The logic was that prosopagnosics would become faster and more accurate at making serial judgements about each feature spacing and, after thousands of trials, could potentially learn to allocate attention to both feature spacings simultaneously, resulting in greater sensitivity to configural information across the inner components of the face (i.e. greater holistic processing).

Training took place over a 15-day period (~30–40 min/day) using five different lifelike face templates created from the 'Faces' composite face-making software (Faces version 3.0, 1998). For each face template, eyebrow height and mouth height were parametrically varied in 2-mm increments to make 10 versions of the template face. Before training began, participants were emailed and instructed to print out the matrix (Fig. 1A) of Face 1 for use on their first day of training to learn the rules of the task. Day 1 consisted of two rounds of training, 300 faces each. Days 2 to 15 consisted of three rounds of training, 300 faces each. All training was self-paced, and visual feedback was provided immediately after each trial (Fig. 1B). After each training round, participants (and the experimenter) were shown a feedback matrix image displaying the accuracy and reaction time for each face as well as overall performance. The experimenter used this information to coach participants (e.g. go more for speed or pay closer attention to the eyebrow/eyes). Over the 15-day training period, the set of five individual template faces was cycled through three times. Subjects were also instructed to recreate laboratory conditions to the best of their ability and to perform the training at a time in their day when they were not tired.

After the first day, the majority of participants continued training without the aid of the matrix image printout. A few participants who particularly struggled on their first day used the matrix image on the second day of training. As of the third day of training, none of the participants were allowed to rely on the matrix printout. For the first few days, all participants were coached to increase their overall accuracy regardless of reaction time. At the point when participants' overall accuracy increased to approximately the mid-80s consistently, the coaching from the experimenter began to also include focus on reaction time. If participants struggled in balancing improvements to both their accuracy and reaction time, they were coached to count how many trials in a row they were responding to correctly, and once they reached six or seven in a row, begin to try to increase their speed. If they then responded incorrectly on two or three trials in a row, they should decrease their speed slightly until they respond correctly six or seven times in a row again, and repeat that process.

At the point where a participant performed at or above 93% overall accuracy and faster than 1000 ms overall reaction time for three training rounds in a row, they were moved up to the next difficulty level and began their next round of training. These thresholds were based on previous pilot studies showing that developmental prosopagnosics who demonstrated training-related improvements all achieved >93% accuracy and <1000 ms reaction time. Thus, we wanted subjects to achieve this level of performance before moving up to more challenging difficulty levels. Higher levels of difficulty (four difficulty levels in total) included faces of more varying sizes (Fig. 1C). Level 1 displayed all the face training stimuli at a visual angle of 3.5° × 5.2° (100%), approximately the peak of holistic/configural processing abilities based on recent findings in healthy control subjects (McKone, 2009). The idea behind Levels 2–4 is to increasingly vary the size of the training faces while still allowing participants to learn the particular spacing in new sizes of faces. Level 2 included one-third of the stimuli at 100%, one-third at 2.6° × 3.9° (75%), and one-third at 4.4° × 6.6° (125%). Level 3 included one-third of the stimuli at 100%, one-third of the stimuli at a visual angle of 1.8° × 2.6° (50%), and one-third of the stimuli at a visual angle of 5.3° × 7.9° (150%). Level 4 included all five previous stimulus sizes distributed equally. In each difficulty level, the visual angles were equally distributed across the 10 different face stimuli included in a round.

Study protocol

As shown in the training timelines in Fig. 3, the study protocol began with screening, consent, and random assignment, as discussed above. For randomization, a random number generator was used to create a list of spots assigned to either 'experimental' or 'waiting list control' and as participants enrolled they were assigned to the next condition on the list.

After random assignment, participants filled out the diary form at the end of each day, for 5 days. On some occasions participants would take notes throughout or at the end of the day and fill in their diary form in the morning for the previous day. After their first (and sometimes second) day of diary entry, the researcher would review their entry to ensure they understood the form properly, and to ask any follow-up questions that could help clarify the numbers that they reported. On the sixth day of their schedule, participants took the battery of pretraining tests in the following order: CFPT, PFST, PWFT, FWMT and FOBPT. Participants were instructed that the test battery session would take ~2–2.5 h, and that it was important that they were able to set aside one block of time during that day in which to complete all the tests together, at a time when they would not be particularly fatigued or distracted. They were instructed to 'find a quiet
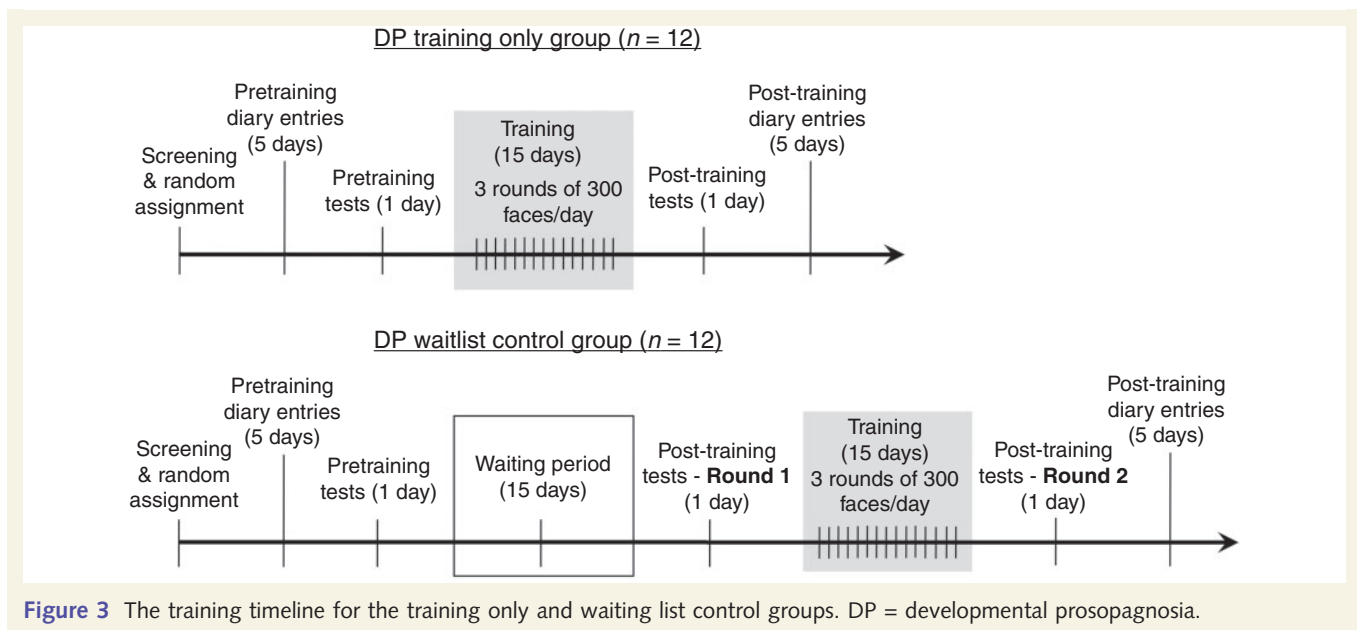
Figure 3 The training timeline for the training only and waiting list control groups. DP = developmental prosopagnosia.

place free from distractions and interruptions,' and to 'sit upright and square to the computer, so that your eyes are approximately two feet away from the computer screen'. Upon completing the test battery, participants emailed the researcher to provide notification that their testing session was finished.

At this point, those participants in the control group began a waiting period of 15 days, whereas experimental training only participants performed the training task (see above) for 15 days. Participants in the training only condition were instructed to perform training each day at a time during which they would not be particularly fatigued, distracted or interrupted. At times there were unforeseen circumstances arising in participants' schedules that caused them to skip a training day. These participants still completed 15 total days of training, although sometimes it was over the course of 16 or 17 days. On their last day of training, or Day 15 of waiting for the control participants, all participants were reminded of the post-training test session that was to take place the following day, and were reinstructed regarding their order of tests and the setting (quiet, free from distraction, etc.) they should prepare for themselves. Post-training tests were performed in the same order as the pretraining tests and participants were instructed to attempt to perform this testing session at the same time of day they performed their pretraining session. The following day, training only participants began five final days of diary entry, which completed their training protocol. At this time, the waiting list control subjects performed the training task for 15 days, a second round of post-training tests, and a final 5 days of diary entry, following the same protocol as the experimental training only subjects described above.

Statistical analysis

The main goal of this study was to measure the impact of face training in developmental prosopagnosics on various measures that are different from the training task itself. One issue with evaluating many measures before and after training is susceptibility to type I errors—finding improvement in individual measures due to chance. To assess the effects of training while accounting for potential type I errors associated with analysing multiple measures, our main analyses focused on four

repeated-measures between groups (training/waiting list control) multivariate analysis of variance models (MANOVAs) for the four domains of face processing in which we were interested: front-view face discrimination, face discrimination from varying view-points, holistic face processing, and everyday face recognition. Another reason we used this MANOVA approach is that it is more powerful at detecting group differences than performing many individual ANOVAs or analyses of covariance (ANCOVAs). To further characterize the data after finding a significant group × pre/post MANOVA, we performed ANCOVAs of individual measures Bonferroni-corrected for multiple comparisons. We used this ANCOVA approach whenever possible because it has shown to be more powerful than a repeated measures ANOVA approach and preferable for smaller cohorts, such as in the current study (Van Breukelen, 2006). For all ANCOVAs reported, we confirmed that the regression slopes were not significantly different between groups and, using Levene's test of homogeneity of variances, that variances did not significantly differ between groups.

As we only hypothesized that training would be more beneficial than waiting and had no reason to consider that waiting would be advantageous over training, we used one-tailed tests when comparing the training period to the waiting period (as recommended by Knottnerus and Bouter, 2001). However, we used two-tailed tests for all other comparisons (e.g. comparing 'better' trainees to 'worse' trainees).

Results

In the following results, we first compare the training only and waiting list groups at baseline to ensure that group differences are from training and not due to differences in demographics, prosopagnosia severity and type, or baseline performance. Next, we examine performance on the training task itself, looking at the degree to which developmental prosopagnosics were able to master this challenging task and progress to higher levels of difficulty. We then investigate the main goal of the study—to characterize the effectiveness of training and the degree to which

training generalized beyond the training task itself. To relate differing amounts of training task improvement to more or less improvement on assessments, we also performed *post hoc* analyses of 'better' and 'worse' trainees.
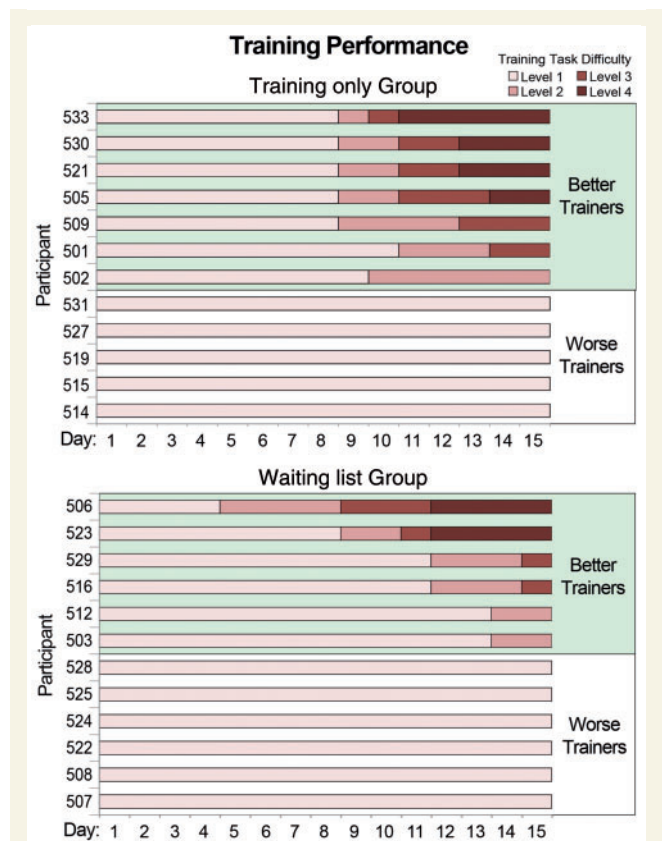
Comparing training and waiting list groups

We first sought to ensure that the participants in the waiting list and training only groups were similar with regards to demographics, prosopagnosia severity, and specificity of face deficits (face deficits only versus face, objects and body deficits). As can be seen in Table 1, the two groups did not significantly differ in age, numbers of males/females, CFMT scores, CFPT scores, or FOBPT (all *P*-values > 0.35). We also compared the waiting list and training only groups on the remaining pretraining measures including the PFST, PWFT whole trials, Part-Whole holistic processing, Face Working Memory upright trials, Face Working Memory inversion effect, and individual diary items and found no significant differences between the groups (all *P*-values > 0.15).

Training task performance

Next, we examined how successful developmental prosopagnosics were at progressing through the training task. All developmental prosopagnosics completed 15 days of training and spent an average time of 34.18 min/day (SD = 5.92 min) performing the training task. As can be seen in Fig. 4, slightly more than half of the developmental prosopagnosics made it beyond the first level of training difficulty (7 of 12 in the training only group and 6 of 12 in the waiting list control training group), completing an average of 5.85 days (SD = 2.44 days) of training with varied face sizes (i.e. difficulty levels 2–4).

Henceforth, we refer to these individuals as 'better' trainees whereas those individuals who did not make it past the first difficulty level are labelled 'worse' trainees. Because of worse trainees slower reaction times compared to the better trainees, the worse trainees performed training for a significantly longer time each day (mean = 37.27 min, SD = 6.38 min) than did better trainees (mean = 31.72, SD = 4.38) [$t(22) = 2.14$, $P < 0.05$]. Although the 'worse' trainees did not achieve the level of proficiency required to advance to the second level of difficulty, they still showed significant training task improvements when comparing the second day of training (the first day on which they did not use the face matrix) to the final day of training, both in accuracy [Day 2 mean = 0.74, final day mean = 0.88, $t(9) = 3.56$, $P < 0.01$] and reaction time [Day 2 mean = 3591 ms, final day mean = 1296 ms, $t(9) = 6.08$, $P < 0.01$]. Together, this suggests that despite not improving enough to advance to Level 2, the worse trainees were still motivated to complete training and improved at the training task. It should be noted that *post hoc* analyses did not show that better trainees were significantly different from worse trainees at the beginning of training (when comparing either the first or second day of training, all *P*-values > 0.15) and were not significantly different from worse trainees on any baseline assessments except for the CFMT. On the CFMT, better

Figure 4 Individual subject data showing levels of difficulty reached over the 15 days of training for the training-only and waiting list groups. We designate those who advanced past the first difficulty level (achieving > 93% accuracy and < 1 s reaction times on three training rounds in a row) as 'better' trainees and those who did not as 'worse' trainees.

trainees scored significantly poorer than worse trainees [better trainees mean = 32.08, worse trainees mean = 38.08, $t(20) = 2.43$, $P < 0.05$]. However, when Bonferroni-corrected for multiple comparisons, this difference was not significant.

In the following sections evaluating the pre/post assessments, we follow our planned analyses with *post hoc* analyses of differences between better and worse trainees to provide a sense of how important training task improvement is to assess improvements.

Assessing the effects of training

Front view face discrimination

We first sought to determine if training produced overall improvements in front-view face discrimination as measured by the PFST, PWFT and FWMT upright trials. To determine if there was an effect of training above and beyond an effect of retaking these assessments (comparing Session 1 to Session 2 after the waiting period in the control group), we performed a repeated-measures MANOVA using these three measures as dependent variables and pre/post × group (training/control) as factors, testing for a
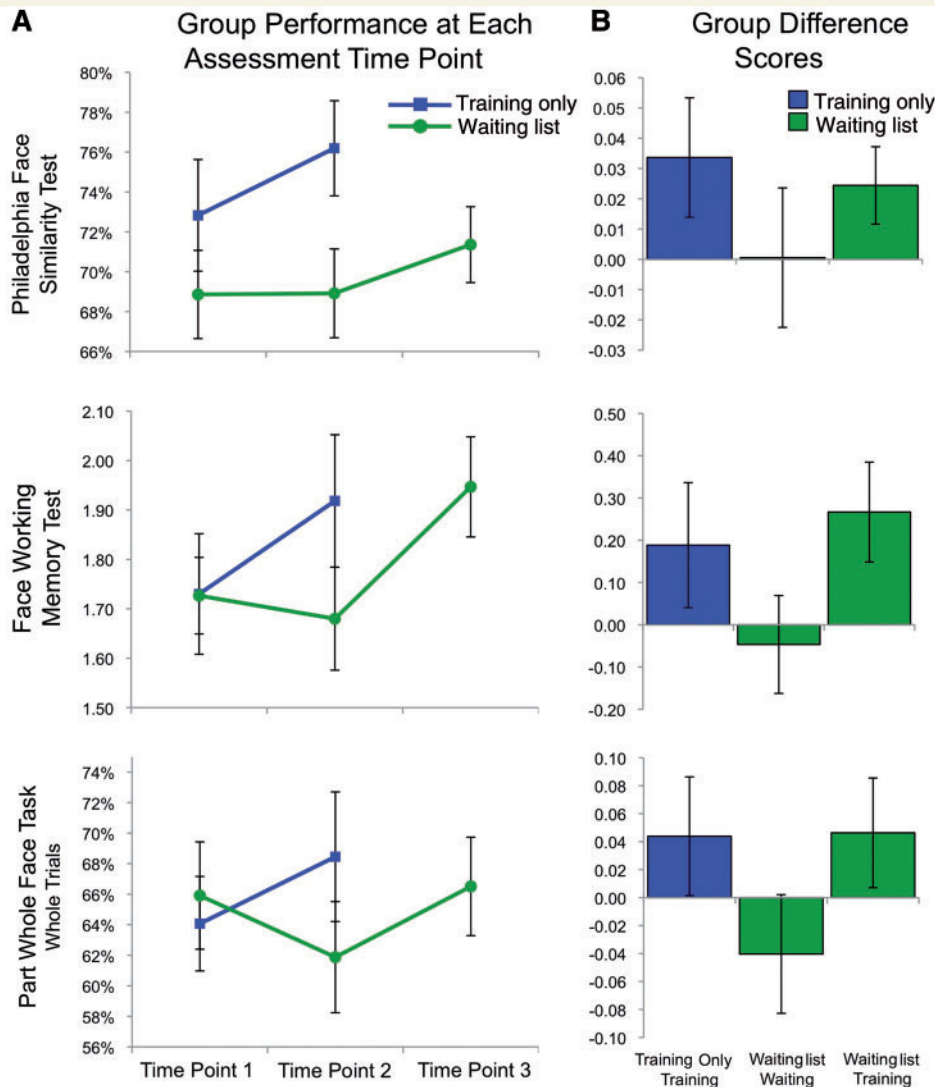
Figure 5 (**A**) Scores for the training only (blue) and the waiting list (green) groups on the three front-view face perception tests. Time point 1 indicates before training/waiting for both groups. Time point 2 is after training for the training only group, and after waiting for the waiting list group. Time point 3 is after training for the waiting list group. Error bars indicate the standard error of the mean. (**B**) Group difference scores on the three front-view face perception tests for training only group (time point 2 − 1), waiting list waiting period (time point 2 − 1), and waiting list training period (time point 3 − 2). Error bars indicate the standard error of the difference scores.

significant pre/post × group interaction. We found a significant overall pre/post × group interaction [$F(1,22) = 3.40$, $P < 0.05$ one-tailed, Wilks' $\lambda = 0.87$] that was driven by generally improved performance on all three tests in the training group after training (Fig. 5).

To follow-up this significant MANOVA, we performed between-groups univariate ANCOVAs for each of the three measures at Session 2 (post-training for training group and post-waiting for control group) controlling for Session 1 performance (pre-training and pre-waiting, respectively). We assessed these ANCOVAs at a P-value of 0.0167 after Bonferroni-correcting for multiple comparisons. The PFST demonstrated a significant between-groups difference only before correcting for multiple comparisons [PFST: $F(1,23) = 3.35$, $P = 0.04$, partial $\eta^2 = 0.14$], whereas the PWFT and FWMT only showed trends towards a

significant difference between groups [PWFT: $F(1,23) = 1.98$, $P = 0.09$, partial $\eta^2 = 0.09$; FWMT: $F(1,23) = 2.04$, $P = 0.08$, partial $\eta^2 = 0.09$]. The most likely reason that the overall MANOVA was significant and tests of these individual measures were not is the lack of power, but also may be due to the fact that the MANOVA test includes sub-hypotheses about linear combinations of the dependent variables that the follow-up ANCOVAs do not capture.

Providing additional confidence in this pattern of results, training-related improvement on the combination of these measures was also replicated when the waiting list group performed training (Fig. 5). In particular, a repeated measures MANOVA on the three measures comparing immediately before training (post-waiting) with after training demonstrated a significant overall pre/post-training effect [$F(1,11) = 5.84$, $P < 0.05$, Wilks' $\lambda = 0.66$]. When examining individual measures Bonferroni corrected for multiple

comparisons ($P < 0.0167$), PFST [$t(11) = 2.31$, $P = 0.04$] and FWMT [$t(11) = 2.65$, $P = 0.02$] showed significant training effects only before Bonferroni correction whereas the PWFT only trended towards showing a significant effect [$t(11) = 1.63$, $P = 0.13$]. Together, these analyses suggest that training produced significant but modest improvements in front-view face discrimination across the entire group of developmental prosopagnosics.

Considering that only about half of the subjects made it past the first level of difficulty in the training task (i.e. better trainees), we sought to perform *post hoc* analyses to see if those individuals had greater face processing improvements than the worse trainees. To increase our power to detect potential subgroup differences, we collapsed before and after training data from the training only and waiting list groups. First we ran an exploratory repeated-measures MANOVA with pre/post-training and group (better/worse trainees) as factors that revealed a trend towards a significant pre/post × group interaction [$F(1,22) = 2.49$, $P = 0.06$, Wilks' $\lambda = 0.90$]. Next, to explore whether better and worse trainees differed in training-related improvements on individual measures, for each test we also ran ANCOVAs on the post-training data covarying out pretraining performance, Bonferroni-corrected for multiple comparisons ($P < 0.0167$). Although we did not find any subgroup differences in the PFST [$F(1,23) = 0.01$, $P = 0.93$, partial $\eta^2 = 0.00$], we found significantly greater improvements in the better trainees compared with worse trainees in the PWFT [$F(1,23) = 9.68$, $P < 0.005$, partial $\eta^2 = 0.32$], as can be seen in Fig. 6 and Table 2, and a trend towards this same pattern on the FWMT [$F(1,23) = 4.82$, $P = 0.04$, partial $\eta^2 = 0.19$].

The better trainees improved to such an extent on Part-Whole whole trials that after training there was no significant difference between their performance and unimpaired controls [better trainees: mean = 0.74, SD = 0.08; unimpaired controls: mean = 0.76, SD = 0.10; $t(49) = 0.32$, $P = 0.75$]. This demonstrates that those who achieved higher difficulty levels of training also showed more pronounced face processing improvements, particularly on the PWFT, and were driving the significant overall improvements on the measures of front face view discrimination.

Face discrimination from varying view-points

We next sought to determine whether training improvements generalized to improvements on tasks that require discriminating faces from different view-points, tasks that have shown to be particularly challenging for prosopagnosics (Marotta *et al.*, 2002; Lee *et al.*, 2010). When examining the faces condition in the FOBPT and the CFPT together in a repeated measures MANOVA or separately, we found no evidence of training-related improvements in either task or improvements in either the better or worse trainees (all *P*-values > 0.6) (Table 2). To assess whether this lack of transfer was because of poor internal consistency of the measurements, we calculated Cronbach's alpha for the faces condition in the FOBPT (alpha = 0.59, acceptable) and for the CFPT (alpha = 0.36, poor to unacceptable; although it should be noted that in a much larger sample of control subjects, Bowles *et al.*, 2009 found a considerably higher reliability of 0.74). This suggests that the lack of improvement on the faces condition in the FOBPT is less likely due to poor internal consistency whereas we cannot rule out this possibility for the CFPT. Considering this,

we cautiously suggest that the effects of training (which only included front-view faces) may not generalize to tasks with faces shown from varying view-points, even in developmental prosopagnosics who made it to higher levels of training difficulty.

Normal versus compensatory mechanisms

Another goal of the study was to quantify whether training enhanced either normal face processing mechanisms (e.g. holistic face processing) or whether it produced compensatory changes (e.g. enhanced parts and object processing). To quantify holistic processing as measured by the Part-Whole effect and Face Working Memory face inversion effect, we employed a regression-based approach (DeGutis *et al.*, 2012b, 2013). This approach provides a measure of holistic processing by examining the condition of interest (e.g. whole trials in PWFT and upright trials in FWMT) after statistically controlling for the control condition (e.g. parts trials in PWFT and inverted trials in FWMT). This regression-based approach is more in-line with the theory of these tasks as measures of holistic processing than more traditionally used subtraction-based approaches (DeGutis *et al.*, 2012b; Wilmer *et al.*, 2012). To create residuals for the part-whole effect and face inversion effect, using healthy control data we first calculated separate regression equations for each task with the control condition (part trials and inverted trials) predicting the condition of interest (whole trials and upright trials, respectively). Using control participants to calculate these regression equations allowed us to measure developmental prosopagnosics' holistic processing abilities in relation to healthy control subjects. We then computed developmental prosopagnosics' holistic processing residuals for each task by inputting developmental prosopagnosics' observed control condition data into separate regression equations predicting the task of interest and comparing the resultant predicted value to the observed value for each participant.

Using the regression residuals for the part-whole effect and face inversion effect, we next sought to quantify if training enhanced overall configural and holistic face processing. A repeated-measures MANOVA using both measures with group (training/control) and pre/post as factors failed to find a significant interaction [$F(1,22) = 0.27$, $P = 0.61$, Wilks $\lambda = 0.988$], suggesting that compared to the control group there was no overall improvement in holistic and configural processing after training.

We also performed exploratory analyses of better and worse trainees collapsing across the training only and waiting list group's pre/post-training data. We first ran an exploratory repeated-measures MANOVA with pre/post-training and group (better/worse trainees) as factors that did not reveal a significant pre/post × group interaction [$F(1,22) = 1.83$, $P = 0.19$, Wilks' $\lambda = 0.93$]. Next, to explore whether better and worse trainees differed in training-related improvements on individual measures, for each test we also ran Bonferroni-corrected ANCOVAs ($P < 0.025$) comparing better and worse trainees after training while co-varying out the pretraining performance. For the part-whole task, this revealed that better trainees significantly improved holistic processing more than worse trainees [$F(1,23) = 7.37$, $P < 0.025$, partial $\eta^2 = 0.26$]. As can be seen in Fig. 6A, improvements in better trainees were to the point where their holistic processing after
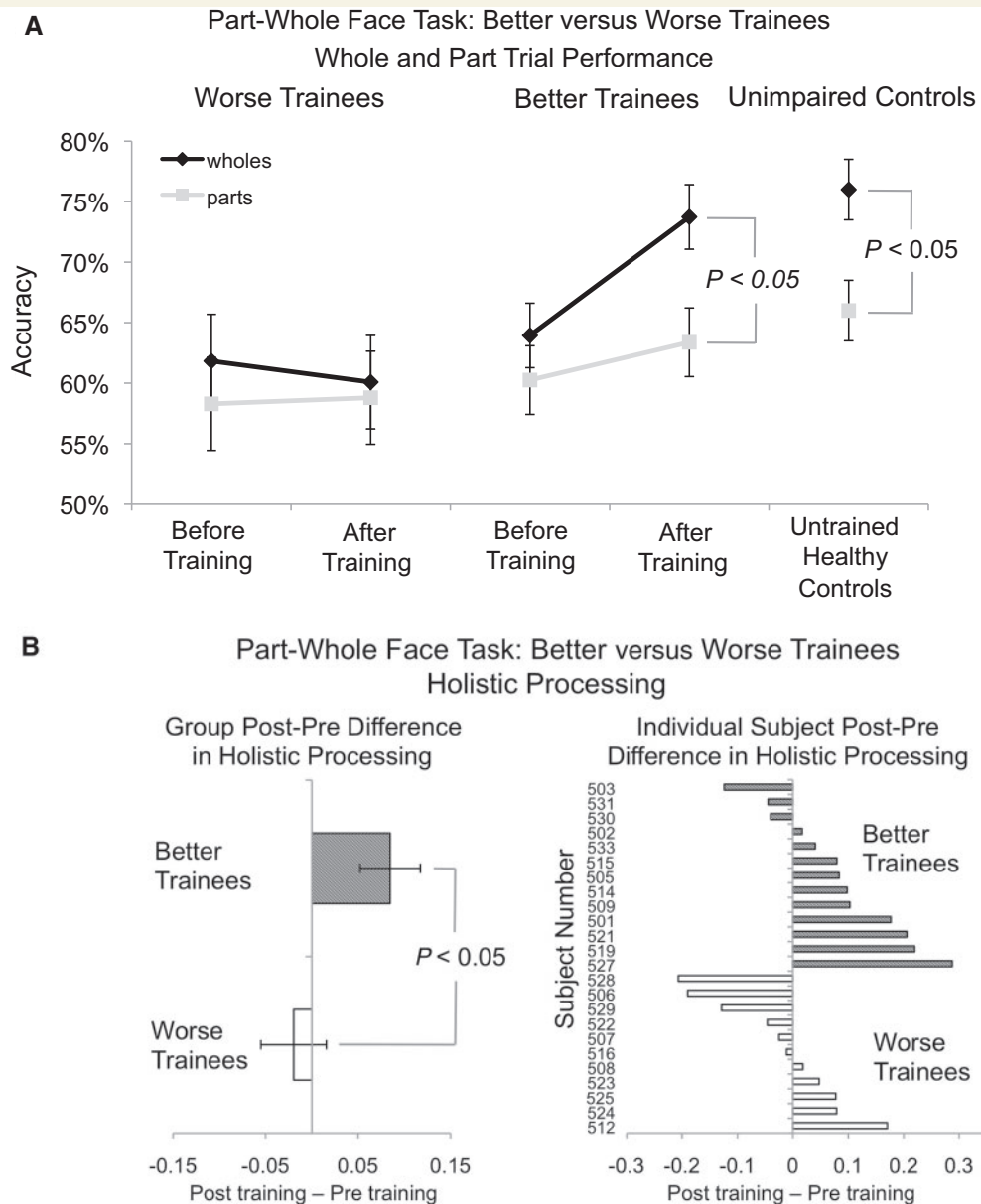
Figure 6 (**A**) Whole (black) and part (grey) trial accuracy for better and worse trainees before and after training compared to unimpaired controls. Error bars indicate the standard error of the mean for each condition. (**B**) Holistic processing measured by regression for better and worse trainees before and after training. Error bars indicate the standard error of the post minus pretraining difference scores.

training was indistinguishable from unimpaired controls (mean residuals of better trainees = −0.0071; mean residuals of unimpaired controls = 0). The face inversion effect showed a similar, though slightly weaker pattern of enhanced holistic processing in the better trainees compared to the worse trainees, with this effect failing to reach significance [$F(1,23) = 1.98$, $P = 0.17$, partial $\eta^2 = 0.09$]. This demonstrates that training enhanced holistic face processing, particularly as measured by the part-whole effect, but only in better trainees.

Recent work from our laboratory has shown that developmental prosopagnosics have particular deficits in holistic processing of the eye region in the part-whole task (DeGutis *et al.*, 2012b). To determine if we particularly remediated eye holistic processing

deficits we separately analysed the part-whole effect by eyes, nose and mouth. ANCOVAs of these separate face regions comparing training only versus waiting list control groups at the post-session whereas co-varying out pre-session scores failed to show significant group differences (all P-values > 0.21), suggesting that overall training did not remediate holistic processing of any particular feature. Furthermore, ANCOVAs comparing better and worse trainees after training on holistic processing of the eyes, nose and mouth when co-varying out the pretraining performance revealed only trends towards improvements for all three features. In particular, better trainees showed trends toward improving more at holistic processing than worse trainees on eyes trials [$F(1,23) = 3.52$, $P = 0.07$, partial $\eta^2 = 0.14$], nose

Table 2 Scores and comparisons of five training assessments between healthy controls, developmental prosopagnosics classified as better trainees, and those classified as worse trainees

	Controls	Better trainees			Worse trainees		
		Before training	After training	Difference	Before training	After training	Difference
Front view							
PFST	79% (6%)	73% (7%)	75% (7%)	2% (2%)	68% (10%)	72% (8%)	4% (2%)
PWFT							
Whole	76% (10%)[b]	**64% (11%)**	**74% (8%)**	**10% (3%)**	62% (12%)	60% (14%)	−2% (4%)
Parts	66% (10%)[b]	60% (7%)	63% (8%)	3% (3%)	58% (8%)	59% (10%)	1% (4%)
FWMT							
Upright	2.60 (0.64)	**1.73 (0.42)**	**2.09 (0.40)**	**0.36 (0.12)**	1.67 (0.35)	1.75 (0.34)	0.07 (0.13)
Inverted	1.70 (0.28)	1.46 (0.14)	1.66 (0.39)	0.21 (0.11)	1.37 (0.16)	1.45 (0.20)	0.08 (0.06)
Varied views							
CFPT	36.7 (12.2)[a]	73.4 (33.6)	59.8 (14.1)	−13.5 (9.8)	60.4 (11.5)	63.3 (13.3)	2.9 (3.1)
FOBPT							
Faces	78% (6%)	69% (7%)	66% (11%)	−2% (2%)	66% (4%)	67% (4%)	1% (2%)
Objects	84% (7%)	90% (6%)	88% (6%)	−2% (1%)	88% (5%)	89% (5%)	1% (1%)
Bodies	85% (5%)	81% (7%)	82% (7%)	1% (2%)	81% (7%)	81% (4%)	0% (2%)

[a]Duchaine *et al.* (2007).
[b]DeGutis *et al.* (2012b).
Pre/post-training comparisons significant at $P < 0.05$ are shown in bold.
A lower score on the CFPT indicates higher accuracy, and a negative difference indicates improvement. Standard deviation are in brackets.

trials [$F(1,23) = 3.91$, $P = 0.06$, partial $\eta^2 = 0.16$], and mouth trials [$F(1,23) = 3.57$, $P = 0.07$, partial $\eta^2 = 0.15$]. This suggests that greater holistic processing improvements in better trainees compared to worse trainees on the part-whole task were not specific to the eye region, but showed improvements across all features.

As a final exploratory analysis, we examined pre/post differences on the objects and bodies conditions in the FOBPT. Improvements on these conditions would suggest that training might enhance general object processing mechanisms rather than face-specific mechanisms such as holistic processing. As can be seen in Table 2, we found no evidence of training-related improvements in either object or body perception or improvements in either better or worse trainees (all P-values > 0.5). This could suggest that training improvements are specific to face processing tasks, although two other possibilities are that the FOBT task is generally insensitive to training-related improvements or that because developmental prosopagnosics generally performed well on the objects and bodies conditions before training, they had little room to improve performance.

Subjective face recognition ability

A final goal of the study was to determine whether the current training program generalized to improvements in self-reported face recognition ability. To accomplish this, we used the face diaries from the entire group of developmental prosopagnosics, combining the training only group who completed diaries before and after training with the waiting list group who completed diaries at the beginning, waited for 15 days, went through training, and then completed the diaries a second time. We ran a repeated-measures MANOVA on the quantitative questions (eight questions: items 1–6, 8, 9) (Supplementary material) before and after training (Fig. 7).
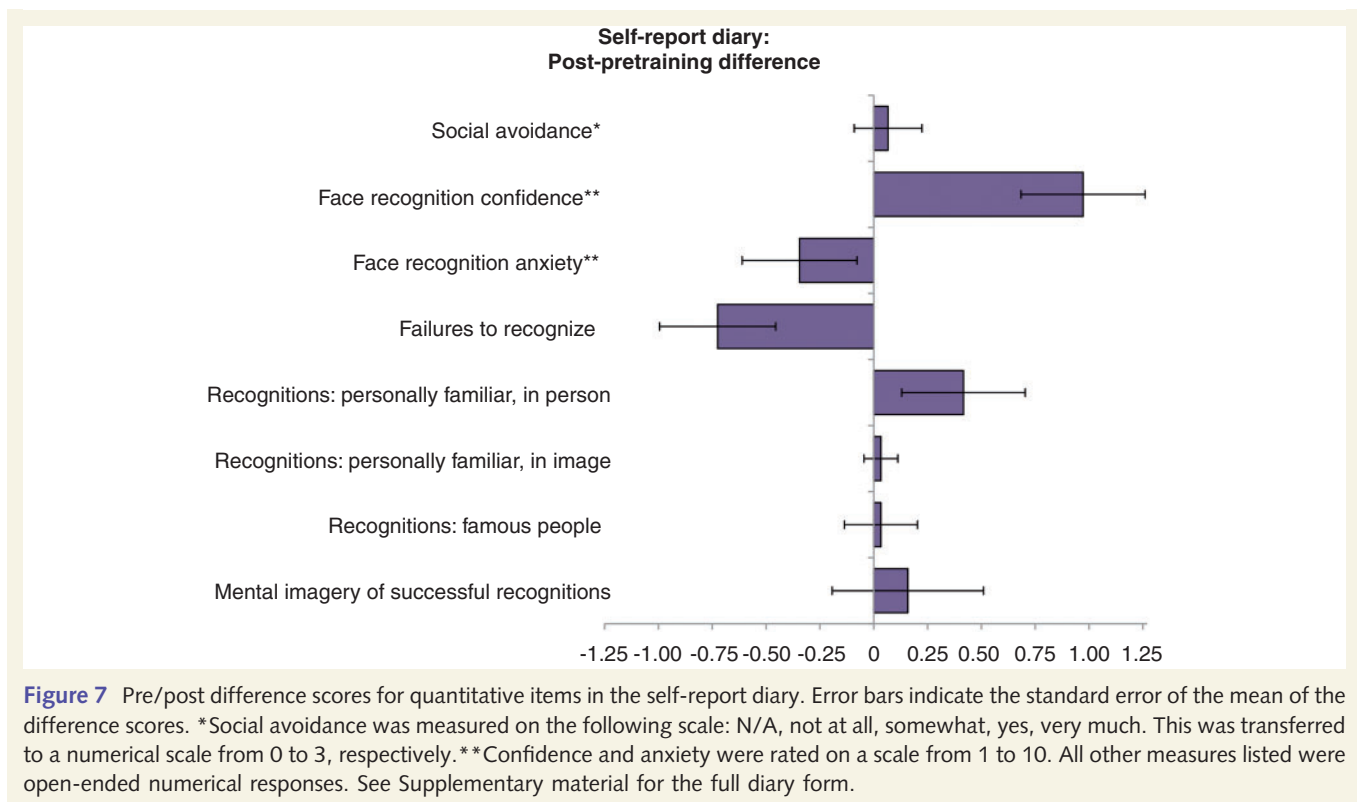
To understand if training-related everyday face recognition improvements differed between the training only and the waiting list

groups, we also included group (training only/waiting list) as a factor in this model. The MANOVA revealed a significant overall difference between pre- and post-training [$F(1,22) = 9.10$, $P < 0.01$, Wilks' $\lambda = 0.71$] that did not significantly vary by training only versus waiting list groups [$F(1,22) = 0.16$, $P = 0.69$, Wilks' $\lambda = 0.99$]. It also revealed a significant pre/post by item interaction [$F(7,16) = 2.64$, $P < 0.05$, Wilks' $\lambda = 0.46$]. When exploring individual items (Bonferroni-corrected for multiple comparisons $P < 0.00625$, one-tailed) we found that individuals significantly improved in face recognition confidence [$t(23) = 3.37$, $P < 0.005$] (Fig. 7) and showed a reduction in the number of daily recognition failures that was not quite significant after Bonferroni correction [$t(23) = 1.66$, $P < 0.01$], whereas the other items failed to reach significance (all P-values > 0.10).

We also explored whether the better trainees showed more everyday improvements in face recognition ability than the worse trainees. For each item above, we used ANCOVAs to compare the two groups' post-training scores when co-varying out their pretraining scores. Although the better trainees showed slightly larger improvements than the worse trainees, these differences failed to reach statistical significance (all P-values > 0.4). Thus, there was a significant overall improvement in daily face recognition across the entire group regardless of whether participants reached higher difficulty levels in training or not.

Discussion

The current results demonstrate that through cognitive training it is possible for developmental prosopagnosics to improve aspects of face processing. In particular, we showed that after completing ~8 h of face training aimed at enhancing holistic processing, developmental prosopagnosics improved on tests of front-view face matching and those who improved the most at the training task

Figure 7 Pre/post difference scores for quantitative items in the self-report diary. Error bars indicate the standard error of the mean of the difference scores. *Social avoidance was measured on the following scale: N/A, not at all, somewhat, yes, very much. This was transferred to a numerical scale from 0 to 3, respectively.**Confidence and anxiety were rated on a scale from 1 to 10. All other measures listed were open-ended numerical responses. See Supplementary material for the full diary form.

(better trainees) showed the greatest front-view face improvements and improved holistic face processing. In fact, on the PWFT, the better trainees showed performance on whole trials and holistic processing after training that was not significantly different from unimpaired controls. Unfortunately, training improvements did not generalize to improvements on face tasks that required view-point rotations, suggesting possible limits to the transfer of training. In addition to these objective measures of face processing, the results also provide preliminary evidence that training produced moderate but consistent improvements in developmental prosopagnosics' self-reported face recognition diaries. Together, these results suggest that developmental prosopagnosics' deficits are not intractable and that improving face processing in developmental prosopagnosics at the group level is possible.

An important aspect of the current study is that it demonstrates that training improvements can transfer to tasks with different stimuli and task formats. Although previous studies training prosopagnosics have shown some evidence of transfer (Schmalzl *et al.*, 2008), by incorporating a larger variety of tasks and stimuli this study was able to better demonstrate transfer across several tasks as well as explore the limitations of transfer.

The transfer of training shown here should be contrasted to the visual perceptual learning literature, which typically finds that learning is highly specific to the stimuli and training task (Fahle, 2005; but see Ahissar and Hochstein, 2004). It also contrasts a recent high-profile large scale study suggesting that web-based cognitive training programs generalize little, if any, beyond improvements on the training program itself (Owen *et al.*, 2010). One possible reason why the current training program may have

promoted transfer is because it trained developmental prosopagnosics to expand their attentional window to the whole face and attention training has been shown to transfer more than basic perceptual training (Ahissar and Hochstein, 2004). Another possible reason that training transferred is, rather than having developmental prosopagnosics learn specific aspects of faces, the current training targeted increasing sensitivity to spacing between features that can be applied to all faces. A final explanation of the transfer of the current training is that it included massive practice on a single skill so subjects could achieve a high level of proficiency, and high levels of proficiency and automaticity have been associated with increased transfer of learning (Ahissar and Hochstein, 2004; Burk and Humes, 2007).

Not only do the results suggest that training generalized to different face tasks, but they also suggest that it enhanced holistic face processing mechanisms. As the current training task requires efficiently discriminating feature spacing (i.e. performing second order configural processing within the eyes and mouth regions), some face theorists would predict that training enhances processing local feature spacing information only and does not generalize to improvements in holistic processing (Maurer *et al.*, 2002). In contrast to this prediction, the part-whole results in the better trainees suggest that training improved perceptual integration abilities across the entire face rather than specifically to the eyes and mouth regions. This improvement is highly relevant for prosopagnosia as several researchers have characterized this disorder as having a significant deficit in holistic face processing (Farah *et al.*, 1998; Busigny *et al.*, 2010; Ramon *et al.*, 2010). For example, several recent studies of prosopagnosics have consistently found significantly reduced holistic processing on the PWFT

(Busigny *et al.*, 2010; Ramon and Rossion, 2010; DeGutis *et al.*, 2012b). Thus, the better trainees significantly enhanced holistic processing on the PWFT to the point of being almost as good as unimpaired controls is notable because it demonstrates that it is possible to enhance a commonly reported, potentially core dysfunction in prosopagnosia. This greater holistic processing after training fits with a previous DP training result from our laboratory showing enhanced N170 selectivity to faces after training (DeGutis *et al.*, 2007). This event-related potential finding is particularly relevant as Jacques and Rossion (2009) demonstrate that the N170 is sensitive to both changes in identity and holistic face processing. Together with the current results, this suggests that those who successfully progress through the face training can achieve a more holistic style of face processing.

These part-whole task improvements in the current study are somewhat similar to those in a cognitive training study of autistic children by Tanaka *et al.* (2010) with some important differences. They found that after 20 h of various perceptual matching and expression recognition training tasks, autistic children demonstrated increased accuracy when averaging across part and whole trials and increased holistic processing of the eyes (but not increased overall holistic processing). Considering that the Tanaka *et al.* (2010) study used different stimuli and timing compared to the current study, it suggests that the Part-Whole paradigm may be generally sensitive to face training-related improvements. One important difference is that the current study found significant improvements in overall holistic processing in the better trainees, indicating enhanced holistic perceptual processing, whereas Tanaka *et al.*'s (2010) holistic processing improvements were specific to the eye region, which may be attributable to increased attention to the eyes. Another important point is that, in contrast to prosopagnosia, holistic face processing is not thought to be a core deficit in autism (for review see Weigelt *et al.*, 2012). Thus, improving holistic processing in developmental prosopagnosics is more relevant to their fundamental deficit than is the case with autism.

Despite the generalization of training-related improvements to front-view face perception and holistic processing, the current results also indicate potential limits to generalization. There was no evidence of training-related improvements on a number the tests, including the CFPT and faces portion of the FOBPT. This could be due to poor internal consistency of the CFPT in this sample. However, the faces portion of the FOBPT showed solid internal consistency and suggests that developmental prosopagnosics' training-related skills could not be successfully applied to faces seen from varied view-points. This could be because face view-point variation is particularly difficult for developmental prosopagnosics and may be resistant to cognitive training (Marotta *et al.*, 2002; Lee *et al.*, 2010). Alternatively, it could be that the current training program did not sufficiently train discriminating faces from different view-points. Either way it demonstrates that teaching developmental prosopagnosics to better process one view of a face does not necessarily lead to a skill that generalizes across view-points.

In addition to showing evidence of improvements on some, but not all, objective measures of face processing, the current study also provides preliminary evidence suggestive of self-reported training-related improvements in face recognition. Even when including studies of healthy participants, there have only been two previous studies that have attempted to quantify self-reported daily face recognition (Young *et al.*, 1985; Schweich *et al.*, 1992). This is likely because awareness of one's own face recognition ability is susceptible to mistaken attributions, which may be exaggerated in prosopagnosia. A further drawback is that in the current diary, several questions were not on a relative scale (e.g. Likert-type scale), but asked for absolute numbers (e.g. How many people who you know personally, did you recognize today by their face in PICTURES or VIDEO?). This resulted in these questions being particularly variable and highly dependent on the type of day the individual had (e.g. at a class reunion versus staying at home and working). Despite these limitations, developmental prosopagnosics did show a significant overall improvement across the entire face diary, demonstrating particular improvement in face recognition confidence. That said, those who showed the largest behavioural improvements (better trainees) did not show greater self-report diary improvements. Thus, self-reported improvements should be interpreted cautiously.

Taken together, the current results suggest that developmental prosopagnosics' deficits can be ameliorated and that their structural and functional neural face processing infrastructure is modifiable with training. Previous studies demonstrate that developmental prosopagnosics have subtle but significant volume reductions in regions associated with face processing in the right middle fusiform gyrus and right inferior temporal gyrus (Garrido *et al.*, 2009) and have shown reduced structural integrity between these regions, particularly the right ventral occipito-temporal white matter tracts (Thomas *et al.*, 2008b). In addition to these structural deficiencies, recent event-related potential evidence suggests that developmental prosopagnosics generally show atypical effects of face inversion on the N170 (Towler *et al.*, 2012) indicative of abnormal encoding of face identity. Functional MRI studies typically do not show abnormalities in core face processing regions such as the fusiform face area (for a counter-example, see Bentin *et al.*, 2007), but have shown abnormal responses in more 'extended' face processing regions in anterior temporal regions, typically involved in processing identity, name, and biographical information (Avidan and Behrmann, 2009; Avidan *et al.*, 2013). Considering these previous findings and our previous case study showing a more normal N170 and enhanced connectivity with right-lateralized posterior face-selective regions after a similar training procedure (DeGutis *et al.*, 2007), we suggest that in developmental prosopagnosics, face training targeting holistic processing may enhance signatures of normal face processing in relatively early stages of face processing.

Although the results demonstrate that training has a positive impact on face processing in developmental prosopagnosics, precisely what they are learning remains to be clarified. One possibility that we favour is that subjects with developmental prosopagnosia are learning to attend to the eye and mouth regions simultaneously rather than having to shift their attention between these regions. This interpretation matches developmental prosopagnosics' self-reports of glancing at the mouth region and then the eye region for the first several thousand trials and then becoming more able to perform the task quickly without having to

look directly at the eyes and mouth (DeGutis *et al.*, 2007, 2008). Recording eye movements early and late in training would be useful to determine whether developmental prosopagnosics are making more centralized (Orban de Xivry *et al.*, 2008) or fewer eye movements. In addition to expanding their window of attention, developmental prosopagnosics may also be better able to perceptually integrate both spacing and feature information within this larger window of attention (i.e. larger perceptual field; see Rossion, 2008). This is consistent with the Part-Whole Task results showing significant improvements in the holistic advantage for aspects of faces not specifically trained, including the nose and the entire eye region. Although the details of what developmental prosopagnosics are learning requires further clarification, the current results suggest that treatment approaches that aim to overcome the developmental prosopagnosics' propensity to attend to circumscribed aspects of face (Barton *et al.*, 2003) and facilitate expanded attention to the 'whole' face may be particularly effective.

One aspect of the current results that may shed additional light on the mechanisms of training is the marked differences between better and worse trainees. The better trainees showed substantially larger improvements on the PWFT and FWMT and larger improvements in holistic and configural processing. This begs the question of what differentiates better from worse trainees. One possibility is that better trainees went for speed and worse trainees took a more careful approach. Though better and worse trainees were equivalent at the beginning of training, by the third day of training better trainees were significantly faster at the training task than worse trainees. It may be that worse trainees' more careful approach does not allow training to become an automatic skill. Another possibility is that because the task was quite difficult and did not change based on worse trainees' performance, it may not have provided the proper level of challenge for worse trainees. This is relevant in that training programs that carefully match and adapt task difficulty to a subject's current performance on a trial-by-trial basis have shown to produce greater training effects than non-adaptive programs (Söderqvist *et al.*, 2012). An additional possibility is that training with varied face sizes was what led to larger improvements in the better trainees. In other words, it may have been that practicing the training task with faces that vary in size created a more robust, flexible skill and made it so that training generalized to other face tasks. For example, due to the location of the eyes and mouth being less predictable in the size change training conditions, it could be that these conditions fostered an approach where subjects make fewer saccades and more simultaneously process the eye and mouth regions. Because the worse trainees never received varied face sizes, this could explain why they did not experience as large improvements as the better trainees.

Although the current study sheds light on the potential for enhancing face processing in developmental prosopagnosics and provides direction for future research, there are many unanswered questions. First, we do not know the longevity of the observed training effects. Although in previous studies we found effects that lasted in the order of months (DeGutis *et al.*, 2007, 2008), it was not possible to conduct a formal follow-up testing with the current sample of developmental prosopagnosics. Another limitation is

that we did not include an active control training task and the observed training improvements could have been due to a placebo effect. One argument against this is that those who did not progress as far through the difficulty levels of the training program (worse trainees) showed more modest improvements than those who made sufficient progress through the difficulty levels of training (better trainees). If the current results were due to a placebo, one would expect that simply discriminating thousands of faces (as in the case of the worse trainees) would be sufficient to improve on the assessments, which is clearly not the case.

In summary, the current study provides evidence that developmental prosopagnosics' face recognition abilities are, at least, partially remediable. We show that ~8 h of face training targeting holistic processing is sufficient to produce overall improvements in front-view face discrimination. We also show that developmental prosopagnosics who made it through the more difficult levels of face training showed the greatest improvements in front-view face discrimination and clear evidence of holistic face processing. Training, however, did not transfer to face discrimination from varying view-points and showed only modest improvements in those developmental prosopagnosics that did not make it beyond the first difficulty level of training. In addition to these results on objective tests, the results also show preliminary evidence for self-reported diary improvements after training. Taken together, the current results provide guideposts for future investigations to further enhance and remediate face recognition deficits in group studies of developmental prosopagnosics (and potentially acquired prosopagnosics) and provide several fruitful approaches to achieve this goal.

Acknowledgements

Funding

Supplementary material

Supplementary material is available at *Brain* online.

References

Ahissar M, Hochstein S. The reverse hierarchy theory of visual perceptual learning. Trends Cogn Sci 2004; 8: 457–64.

Avidan G, Behrmann M. Functional MRI reveals compromised neural integrity of the face processing network in congenital prosopagnosia. Curr Biol 2009; 19: 1146–50.

Avidan G, Tanzer M, Behrmann M. Impaired holistic processing in congenital prosopagnosia. Neuropsychologia 2011; 49: 2541–52.

Avidan G, Tanzer M, Hadj-Bouziane F, Liu N, Ungerleider LG, Behrmann M. Selective dissociation between core and extended regions of the face processing network in congenital Prosopagnosia. Cereb Cortex 2013. Advance Access published on January 31, 2013, doi: 10.1093/cercor/bht007.

Baron-Cohen S, Wheelwright S, Skinner R, Martin J, Clubley E. The Autism-Spectrum Quotient (AQ): Evidence from Asperger syndrome/high-functioning autism, males and females, scientists and mathematicians. J Autism Dev Disord 2001; 31: 5–17.

Barton JJ, Cherkasova MV, Press DZ, Intriligator JM, O'Connor M. Developmental prosopagnosia: a study of three patients. Brain Cogn 2003; 51: 12–30.

Bentin S, Degutis JM, D'Esposito M, Robertson LC. Too many trees to see the forest: performance, event-related potential, and functional magnetic resonance imaging manifestations of integrative congenital prosopagnosia. J Cogn Neurosci 2007; 19: 132–46.

Beyn ES, Knyazeva GR. The problem of prosopagnosia. J Neurol Neurosurg Psychiatry 1962; 25: 154–8.

Bowles DC, McKone E, Dawel A, Duchaine B, Palermo R, Schmalzl L, et al. Diagnosing prosopagnosia: Effects of ageing, sex, and participant–stimulus ethnic match on the Cambridge Face Memory Test and Cambridge Face Perception Test. Cogn Neuropsychol 2009; 26: 423–55.

Brunsdon R, Coltheart M, Nickels L, Joy P. Developmental prosopagnosia: A case analysis and treatment study. Cogn Neuropsychol 2006; 23: 822–40.

Bukach CM, Bub DN, Gauthier I, Tarr MJ. Perceptual expertise effects are not all or none: spatially limited perceptual expertise for faces in a case of prosopagnosia. J Cogn Neurosci 2006; 18: 48–63.

Burk MH, Humes LE. Effects of training on speech recognition performance in noise using lexically hard words. J Speech Lang Hear Res 2007; 50: 25–40.

Busigny T, Joubert S, Felician O, Ceccaldi M, Rossion B. Holistic perception of the individual face is specific and necessary: evidence from an extensive case study of acquired prosopagnosia. Neuropsychologia 2010; 48: 4057–92.

Chatterjee G, Nakayama K. Normal facial age and gender perception in developmental prosopagnosia. Cogn Neuropsychol 2012; 29: 482–502.

Dalrymple KA, Corrow S, Yonas A, Duchaine B. Developmental prosopagnosia in childhood. Cogn Neuropsychol 2012; 29: 393–418.

DeGutis J, Robertson L, Nakayama K, McGlinchey R, Milberg W. Learning faces: plasticity and the rehabilitation of congenital prosopagnosia. J Vis 2008; 8: 184a.

DeGutis J, Chatterjee G, Mercado RJ, Nakayama K. Face gender recognition in developmental prosopagnosia: evidence for holistic processing and use of configural information. Vis Cogn 2012a; 20: 1242–53.

DeGutis J, Cohan S, Mercado RJ, Wilmer J, Nakayama K. Holistic processing of the mouth but not the eyes in developmental prosopagnosia. Cogn Neuropsychol 2012b; 29: 419–46.

DeGutis J, Wilmer J, Mercado RJ, Cohan S. Using regression to measure holistic face processing reveals a strong link with face recognition ability. Cognition 2013; 126: 87–100.

DeGutis JM, Bentin S, Robertson LC, D'Esposito M. Functional plasticity in ventral temporal cortex following cognitive rehabilitation of a congenital prosopagnosic. J Cogn Neurosci 2007; 19: 1790–802.

Duchaine BC, Nakayama K. Developmental prosopagnosia: a window to content-specific face processing. Current Opin Neurobiol 2006a; 16: 166–73.

Duchaine B, Nakayama K. The Cambridge Face Memory Test: results for neurologically intact individuals and an investigation of its validity using inverted face stimuli and prosopagnosic participants. Neuropsychologia 2006b; 44: 576–85.

Duchaine B, Germine L, Nakayama K. Family resemblance: ten family members with prosopagnosia and within-class object agnosia. Cogn Neuropsychol 2007; 24: 419–30.

Ellis HD, Young AW. Training in face-processing skills for a child with acquired prosopagnosia. Dev Neuropsychol 1988; 4: 283–94.

Fahle M. Perceptual learning: specificity versus generalization. Curr Opin Neurobiol 2005; 15: 154–60.

Farah MJ, Wilson KD, Drain M, Tanaka JN. What is "special" about face perception? Psychol Rev 1998; 105: 482–98.

Garrido L, Furl N, Draganski B, Weiskopf N, Stevens J, Tan GC, et al. Voxel-based morphometry reveals reduced grey matter volume in the temporal cortex of developmental prosopagnosics. Brain 2009; 132 (Pt 12): 3443–55.

Germine L, Nakayama K, Duchaine BC, Chabris CF, Chatterjee G, Wilmer JB. Is the Web as good as the lab? Comparable performance from Web and lab in cognitive/perceptual experiments. Psychon Bull Rev 2012; 19: 847–57.

Jacques C, Rossion B. The initial representation of individual faces in the right occipito-temporal cortex is holistic: electrophysiological evidence from the composite face illusion. J Vis 2009; 9 (8): 1–16.

Kennerknecht I, Grueter T, Welling B, Wentzek S, Horst J, Edwards S, et al. First report of prevalence of non-syndromic hereditary prosopagnosia (HPA). Am J Med Genet A 2006; 140: 1617–22.

Kennerknecht I, Ho NY, Wong VC. Prevalence of hereditary prosopagnosia (HPA) in Hong Kong Chinese population. Am J Med Genet A 2008; 146A: 2863–70.

Knottnerus JA, Bouter LM. The ethics of sample size: two-sided testing and one-sided thinking. J Clin Epidemiol 2001; 54: 109–10.

Lee Y, Duchaine B, Wilson HR, Nakayama K. Three cases of developmental prosopagnosia from one family: detailed neuropsychological and psychophysical investigation of face processing. Cortex 2010; 46: 949–64.

Marotta JJ, McKeeff TJ, Behrmann M. The effects of rotation and inversion on face processing in prosopagnosia. Cogn Neuropsychol 2002; 19: 31–47.

Maurer D, Grand RL, Mondloch CJ. The many faces of configural processing. Trends Cogn Sci 2002; 6: 255–60.

Mayer E, Rossion B. Prosopagnosia. In: Godefroy O, Bogousslavsky J, editors. The behavioral and cognitive neurology of stroke. Cambridge: Cambridge University Press; 2007. p. 315–34.

McKone E. Holistic processing for faces operates over a wide range of sizes but is strongest at identification rather than conversational distances. Vis Res 2009; 49: 268–83.

Orban de Xivry JJ, Ramon M, Lefèvre P, Rossion B. Reduced fixation on the upper area of personally familiar faces following acquired prosopagnosia. J Neuropsychol 2008; 2: 245–68.

Owen AM, Hampshire A, Grahn JA, Stenton R, Dajani S, Burns AS, et al. Putting brain training to the test. Nature 2010; 465: 775–8.

Palermo R, Willis ML, Rivolta D, McKone E, Wilson CE, Calder AJ. Impaired holistic coding of facial expression and facial identity in congenital prosopagnosia. Neuropsychologia 2011; 49: 1226–35.

Pitcher D, Charles L, Devlin JT, Walsh V, Duchaine B. Triple dissociation of faces, bodies, and objects in extrastriate cortex. Curr Biol 2009; 19: 319–24.

Polster MR, Rapcsak SZ. Representations in learning new faces: evidence from prosopagnosia. J Int Neuropsychol Soc 1996; 2: 240–8.

Ramon M, Busigny T, Rossion B. Impaired holistic processing of unfamiliar individual faces in acquired prosopagnosia. Neuropsychologia 2010; 48: 933–44.

Ramon M, Rossion B. Impaired processing of relative distances between features and of the eye region in acquired prosopagnosia–two sides of the same holistic coin? Cortex 2010; 46: 374–89.

Rossion B. Picture-plane inversion leads to qualitative changes of face perception. Acta Psychol 2008; 128: 274–89.

Schmalzl L, Palermo R, Green M, Brunsdon R, Coltheart M. Training of familiar face recognition and visual scan paths for faces in a child with congenital prosopagnosia. Cogn Neuropsychol 2008; 25: 704–29.

Schweich M, Vanderlinden M, Bredart S, Bruyer R, Nelles B, Schils JP. Daily-life difficulties in Person recognition reported by young and elderly subjects. Appl Cogn Psych 1992; 6: 161–72.

Söderqvist S, Nutley SB, Ottersen J, Grill KM, Klingberg T. Computerized training of non-verbal reasoning and working memory in children with intellectual disability. Front Hum Neurosci 2012; 6: 271.

Tanaka JW, Kiefer M, Bukach CM. A holistic account of the own-race effect in face recognition: evidence from a cross-cultural study. Cognition 2004; 93: B1–9.

Tanaka JW, Wolf JM, Klaiman C, Koenig K, Cockburn J, Herlihy L, et al. Using computerized games to teach face recognition skills to children with autism spectrum disorder: the Let's Face It! program. J Child Psychol Psychiatry 2010; 51: 944–52.

Thomas AL, Lawler K, Olson IR, Aguirre GK. The Philadelphia face perception battery. Arch Clin Neuropsychol 2008a; 23: 175–87.

Thomas C, Moya L, Avidan G, Humphreys K, Jung KJ, Peterson MA, et al. Reduction in white matter connectivity, revealed by diffusion tensor imaging, may account for age-related changes in face perception. J Cogn Neurosci 2008b; 20: 268–84.

Towler J, Gosling A, Duchaine B, Eimer M. The face-sensitive N170 component in developmental prosopagnosia. Neuropsychologia 2012; 50: 3588–99.

Van Breukelen GJ. ANCOVA versus change from baseline: more power in randomized studies, more bias in nonrandomized studies [corrected]. J Clin Epidemiol 2006; 59: 920–5.

Weigelt S, Koldewyn K, Kanwisher N. Face identity recognition in autism spectrum disorders: A review of behavioral studies. Neurosci Biobehav Rev 2012; 36 (3): 1060–84.

Wilmer JB, Germine L, Chabris CF, Chatterjee G, Williams M, Loken E, et al. Human face recognition ability is specific and highly heritable. Proc Natl Acad Sci USA 2010; 107: 5238–41.

Wilmer JB, Germine L, Chabris CF, Chatterjee G, Gerbasi M, Nakayama K. Capturing specific abilities as a window into human individuality: the example of face recognition. Cogn Neuropsychol 2012; 29: 360–92.

Yardley L, McDermott L, Pisarski S, Duchaine B, Nakayama K. Psychosocial consequences of developmental prosopagnosia: a problem of recognition. J Psychosom Res 2008; 65: 445–51.

Young AW, Hay DC, Ellis AW. The faces that launched a thousand slips: everyday difficulties and errors in recognizing people. Br J Psychol 1985; 76 (Pt 4): 495–523.