Running Head: FACIAL IDENTITY TRAINING

Guess Who? Facial Identity Discrimination Training Improves Face Memory in Typically Developing Children

Sarah Bate[1], Amanda Adams[1] & Rachel J. Bennetts[2]

[1]Department of Psychology, Bournemouth University, UK;

[2]College of Health and Life Sciences, Division of Psychology, Brunel University, UK.

Address for correspondence:

Professor Sarah Bate

Department of Psychology

Faculty of Science and Technology

Poole House

Bournemouth University

Poole

BH12 5BB

United Kingdom

Email: sbate@bournemouth.ac.uk

Abstract

While vast individual differences in face recognition have been observed in adults, very little work has explored when these differences come online during development, their domain-specificity, and their consistency across different aspects of face-processing. These issues do not only have important theoretical implications for the cognitive and developmental psychological literatures, but may reveal critical windows of neuroplasticity for optimal remediation of face recognition impairments. Here, we describe the first formal remedial face training programme that is suitable for children, modifying the popular game *Guess Who.* Eighty-one typical children aged 4-11 years were randomly allocated to an experimental or active control training condition. Over 10 training sessions, experimental participants were required to discriminate between faces that differed in feature size or spacing across 10 levels of difficulty, whereas control participants continuously played the standard version of *Guess Who* within the same timeframe. Improvements in face memory but not face matching were observed in the experimental compared to the control group, but there were no gains on tests of object matching or memory. Face memory gains were maintained in a one-month follow-up, consistent across age, and larger for poorer perceivers. Thus, this study not only presents a promising means of improving face recognition skills in children, but also indicates a consistent period of plasticity that spans early childhood to pre-adolescence, implying early segregation of face versus object processing.

Keywords: Face recognition; face perception; development; training; individual differences; prosopagnosia.

Guess Who? Facial Identity Discrimination Training Improves Face Memory in Typically Developing Children

It is becoming increasingly clear that there are vast individual differences in the face recognition skills of adults, suggesting a broad continuum of ability (e.g. Bruce, Bindemann & Lander, 2018; Wilmer, 2017). At the bottom end of this distribution are people with a condition known as "developmental prosopagnosia", who experience profound difficulties with facial identity recognition (for an overview see Bate & Tree, 2017). Much less work has investigated whether the same individual differences occur in childhood, although there are case reports of children as young as four years with the developmental form of prosopagnosia (Dalrymple, Garrido & Duchaine, 2014). Despite widespread theoretical and clinical interest in prosopagnosia across the lifespan, and evidence that large numbers of people may be affected (Bowles et al., 2009), very little work has examined its remediation (Bate & Bennetts, 2014). The work that has been carried out to date has mostly attempted to improve face recognition skills in adults, with little-to-moderate levels of success (Davies-Thompson et al., 2017; DeGutis, Bentin, Robertson & D'Esposito, 2007; DeGutis, Cohan & Nakayama, 2014). Critically, gains in adults may be curtailed by limited neuroplasticity within the mature face recognition system – a viewpoint shared by the general neurorehabilitation literature, which posits greater gains from cognitive training following early intervention (e.g. Elbert et al., 2001; Huttenlocher, 2002). Yet, no study to date has carried out a remediatory face training programme in children, despite its ability to address important theoretical questions within the cognitive, developmental, individual differences and rehabilitation literatures.

Instead, there are isolated attempts to improve face recognition skills in single-cases of childhood prosopagnosia. For instance, Ellis and Young (1988) administered four training programmes to a child with acquired prosopagnosia between the ages of 8 and 11 years. These programmes involved simultaneous matching of photographs of familiar and unfamiliar faces, paired discriminations of computer-generated schematic faces, paired discriminations of digitized images of real faces, and learning face–name associations. Unfortunately, none of these programmes yielded any improvement with practice. Two studies successfully trained an 8 year-old and a 4-year old with developmental prosopagnosia to recognise distinctive features of familiar faces (Brunsdon, Coltheart, Nickels & Joy, 2006; Schmalzl, Palermo, Green, Brunsdon & Coltheart, 2008; but see Dalrymple, Corrow, Yonas & Duchaine, 2010, for a child who did not gain from this intervention). However, this technique did not attempt to improve underlying deficits in processing strategy, but simply taught the child to compensate and circumvent recognition difficulties. Such compensatory approaches are not only laboured and challenging to implement, but also require intensive support from a carer (DeGutis et al., 2014).

The only existing remedial training study in a child was recently reported by our group (Bate et al., 2015): we described the case of EM, a 14 year-old female who acquired prosopagnosia following encephalitis at the age of eight years. EM underwent 14 weeks of perceptual training via an online face perception programme that attempted to improve her ability to make fine-grained discriminations between faces, progressing across 10 levels of difficulty. EM's face perception skills improved post-training, and she spent more time viewing the inner facial features. The gains transferred to new faces, and laboratory assessments also indicated improvements in her recognition of personally-known faces, although this did not transfer to everyday life. Importantly, this study raises the possibility that more formal perceptual face training programmes can have at least some success, particularly given the participant's age (intervention of this kind could readily be applied to much younger children when packaged in an appropriate format) and widespread brain damage.

Encouragingly, modest gains have also been observed in similar programmes that used adult participants. DeGutis and colleagues (2014; see also DeGutis et al., 2007) trained 24 adults with developmental prosopagnosia via a 3-week online programme that targeted relational processing – the ability to process spatial relationships between different facial features (Piepers & Robbins, 2012). Specifically, participants were required to make rapid category judgements about large numbers of faces (~30-40 minutes per day), by integrating the distance between the eyes and eyebrows with the distance between the mouth and nose. Compared with performance in a no-training waiting condition, participants showed moderate improvements on measures of front-view face discrimination, tests of holistic processing, and in self-reported diaries of everyday face recognition experiences. Pertinently, the largest improvements were observed in those who dedicated more time to training.

Davies-Thompson et al. (2017) applied a similar training programme to 10 adults with acquired prosopagnosia, over an 11-week period. These participants were required to discriminate whole-face differences over a variety of views and expressions for 30-40 minutes per session, three times a week, and a staircase design controlled the difficulty levels of subsequent trials. Gains generalized to new viewpoints and expressions of the trained faces and to untrained faces, and persisted for at least three months post-training. While there were minimal gains on standard tests of face-processing and in transfer to everyday life, gains were greater for individuals who had initially presented with more severe deficits in face perception.

The studies reviewed above suggest that adult cases of prosopagnosia can benefit to some degree from training, with both the severity of face-processing deficits at baseline and level of engagement influencing training outcome. Yet, no study to date has considered the influence of participant age on training outcomes, despite this being a renowned factor in the outcome of neurorehabilitation (Elbert et al., 2001; Huttenlocher, 2002). Therefore, identifying an age at which the face processing system is most responsive to training – i.e. is at its most plastic – is a key step in developing effective interventions for individuals with face processing deficits.

Despite some evidence of very early limits in plasticity (e.g. Geldart, Mondloch, Maurer, de Schonen & Brent, 2002; Pascalis et al., 2002), further evidence suggests that the face recognition system does not fully mature until later in development – offering a potential window for maximizing the gains of intervention during childhood. However, the age at which maturation occurs is theoretically contentious. To date, two conflicting viewpoints have been offered: one advocating slow maturation of face versus object recognition over the first 10+ years of life (Carey, 1992), and the other suggesting that the mechanisms underpinning face recognition are fully developed at an early stage (McKone, Crookes, Jeffery & Dilks, 2012). Weigelt et al. (2014) offer an intriguing attempt to reconcile these theories, reporting early adult-like processing of both face and object (cars, scenes and bodies) perception, but steeper developmental slopes for the memory of faces compared to objects. However, another study found that faces and bikes showed similar developmental trajectories for both memory and matching (Bennetts et al., 2017). Thus, while the proposed independence of face perception and face memory has long been investigated in the adult neuropsychological literature and is reflected in dominant theories of face-processing (e.g. Bruce & Young, 1986), it is much less clear how, or when, these two processes unfold in development, nor when adult-like individual differences in performance come online.

Further, the issue of domain-specificity has been debated in the adult cognitive neuropsychological literature for more than 50 years, but is far from resolved (Geskin & Behrmann, 2018). While many researchers support modular accounts of functionally distinct cortical face and object representations (e.g. Kanwisher, 2017), others advocate domain-general hypotheses of distributed cortical function (e.g. Behrmann & Plaut, 2015) and/or common underlying mechanisms for different object categories (e.g. Richler, Palmeri & Gauthier, 2012). Research on acquired and developmental prosopagnosia suggests that face and object processing deficits can dissociate (Bate, Bennetts, Tree, Adams & Murray, in press; Duchaine & Nakayama, 2005; Garrido et al., 2009; Rezlescu, Pitcher, & Duchaine, 2012; although see Geskin & Behrmann, 2017, for discussion), which supports the theory that at least some aspects of face processing are domain specific. Notably, these dissociations can even be observed in children (Bennetts et al., 2017, Dalrymple, Elison, & Duchaine, 2017), suggesting that domain specificity emerges relatively early in life. However, the point at which this occurs remains unclear. While it is possible that faces and objects rely on separate mechanisms from birth, it is also feasible that face and object recognition rely on shared mechanisms during early life (e.g. processing of individual parts), but segregate or become specialized at some point during development.

The current study aimed to address the substantive issues reviewed above by examining the efficacy of a facial identity discrimination programme in a large sample of typical children aged between four and 11 years. Investigating whether a face training programme brings about domain-specific gains in children in some/all aspects of face-processing offers a novel means to inform several key theoretical debates about the development of face processing, particularly if there is a critical age where gains for memory or matching can be maximised or become domain-specific. Our face training programme built upon the principles of existing remediatory programmes described above. To encourage engagement with the programme, we framed it within the traditional childhood game *Guess Who* (Hasbro Gaming) – a commercially available family game that is enjoyed by children worldwide.

Our rationale for using typical children was primarily motivated by sample size, given (a) relatively few children with face-processing deficits are currently known to researchers, and (b) a larger sample would allow us to address specific research questions regarding age (i.e. is the programme more successful in younger compared to older children?), individual differences in face-processing ability at baseline (i.e. are gains greater in poorer perceivers?), and the specificity of the developing face recognition system (i.e. are gains restricted to face perception and/or memory in comparison to other classes of objects?). Further, proof-of-concept of a new training programme could more convincingly be obtained from typical children, as extraneous factors that often complicate acquired and developmental disorders (e.g. accompanying cognitive and social deficits) could more readily be controlled.

Method

Participants

We set out to recruit a total of 80 typically developing children, equally split between the experimental and control training conditions. This sample size was calculated to give 80% power to detect moderate-to-large interaction effects (d = 0.70) and small (d = 0.30) within-subject main effects in the primary analysis (comparing performance on pre- and post-test measures), both of which are smaller than the training effects found by DeGutis et al. (2014) (power calculations carried out in G*Power 3.1). Because we selected a memory paradigm for pre- and post-assessment that differed in memory load (see below) for older (aged 7-11 years) and younger (aged 4-6 years) children, and we specifically set out to examine any age-related differences in training between these two age groups, we ensured equal numbers of children were recruited for each age group in each training condition. Thus, participants were allocated to the experimental and control training groups via pseudo random allocation.

Exclusion criteria were any history of face-processing deficits, uncorrected vision, or developmental, psychiatric or neurological disorder.

Recruitment paused when a minimum of 40 participants in each training condition had successfully completed the study through to the post-assessment phase. Due to a high drop-out rate (33%), 123 children were recruited in total. Forty-one participants dropped out at various phases: prior to the pre-assessments (N = 8; two from the experimental condition), during the pre-assessment tests (N = 17; six from the experimental condition), or during training or the post-test phase (N = 16; 10 from the experimental condition). Eighteen of the 41 participants who dropped out were from the younger age-group.

The final sample contained 82 participants (42 female), who were reported by their parents or guardians to have completed the training as requested. Forty-one were allocated to the experimental (M age = 7.2 years, SD = 2.3) condition, and 41 (M age = 6.9 years, SD = 2.2) to the training condition. Within the experimental condition, there were 21 younger (14 female; M age = 5.3 years, SD = 0.8, range = 4-6 years) and 20 older (nine female; M age = 9.4 years, SD = 1.3, range = 7-11 years) children; in the control condition there were 20 younger (seven female; M age = 5.0 years, SD = 0.8, range = 4-6 years) and 21 older (15 female; M age = 8.7 years; SD = 1.4, range = 7-11 years) children. Participants received a small financial incentive in exchange for their time, and were gifted their *Guess Who* game set. Ethical approval was granted by the institutional Ethics Committee.

Materials

Training: The training procedure adopted the format of the popular two-player children's game *Guess Who* (Hasbro Gaming). In the commercial version, the two players sit opposite each other, viewing the same set of 24 cartoon faces that are encased within a plastic frame, each behind a closable window (see Figure 1). Each image displays the entire face in colour from the neck upwards, and measures approximately 2.0 x 1.8 cm. A name is presented immediately below each image. Each player selects one face to be their chosen character, and ensures all the windows are open. They then take turns to ask yes/no questions that will help to reveal the identity of their opponent's character. For instance, if a player asks "Is your character female?" and receives an affirmative response, they would then shut all the windows covering male faces. The game proceeds until one player has only one window left open, and are able to guess the identity of their opponent's character. If they are correct, they win the game. Variation in characters' appearance are largely due to gender, ethnicity, the wearing of accessories (e.g. hats or spectacles), eye colour, facial hair, and hairstyle/colour (see Figure 1). These differences form the basis of the players' yes/no questions.

< Insert Figure 1 >

To create a perceptual face training programme within this game, we developed new insert cards that slot into the plastic frame, replacing the cartoon faces that are provided with the original game-set. Two new versions were created: one displaying male faces and the other displaying female faces (see Figure 2A and 2B). Each version was created by obtaining two colour photographic images of a model. In one image the model displayed a neutral facial expression, and in the other image the model displayed a happy facial expression. These base images were cropped around the neck and hair, adjusted to a size of approximately 1.7 x 2.3 cm, and presented on a white background. Twelve manipulated images were then created from each base image, using different combinations of four adjustments (see Figure 2A and 2B): two affecting the spacing of facial features (the distance between the eyes, or the distance between the eyes and mouth), and two affected the size of specific facial features (the eyes and nose). These manipulations were taken from the design of existing training programmes that have experienced at least some success: expression was manipulated by Davies-Thompson et al. (2017), spacing by DeGutis et al. (2007, 2014), and feature size by Bate et al. (2015). Together, a total of 24 new character images were created for each of the male and female insert cards. Each image was paired with a name taken from the annual UK list of popular baby names. Note that the external features of each character were identical and could not be used as cues to identity. Thus, the only questions that players could ask to discriminate identity refer directly to the expression, spacing and feature size manipulations described above.

< *Insert Figure 2* >

The images displayed in each version were then adjusted to become increasingly fine-grained in their differences over 10 levels of difficulty (see Figure 2C and 2D). That is, nine further image cards were prepared for each gender, where the manipulations became progressively less extreme (i.e. expressions became more ambiguous, or the differences in spacing or feature size became increasingly smaller). This resulted in 20 different image cards (10 male, 10 female; with 10 levels of difficulties for each gender). We then created a second set of the 20 cards where the character positions were dispersed across different locations. This provided two versions of the identical stimulus set: one for each opponent, in order to prevent location-based cueing to the correct answer during play.

Assessment tasks: Four existing tasks with high reliability (Bennetts et al., 2017) were used to assess the efficacy of training, assessing face and object (bike) processing, for memory and matching. Face memory was assessed with the Cambridge Face Memory Test – Kids (CFKT-K; Dalrymple et al., 2014; see Figure 3). The CFMT-K is adapted from the Cambridge Face Memory Test (CFMT; Duchaine & Nakayama, 2006), which is commonly used to measure individual differences and identify cases of prosopagnosia in the adult population (e.g., Bowles et al., 2009; Dalrymple & Palermo, 2016). The CFMT-K requires children to learn and then identify a number of target faces. The task has three stages. In the first (learning) stage, the face of a Caucasian male child is presented three times, from three different viewpoints (left, frontal, right; for three seconds per image). Faces are presented in greyscale, and cropped to remove hair or other identifying features. Subsequently, children are presented with three trials which assess their learning of the face. In each trial, children view three faces simultaneously, one of which is identical to the learning images, and two which show a different male Caucasian child from the same viewing angle. Children are asked to select which face they saw in the learning phase by pressing the 1, 2, or 3 key on their keyboard. Stimuli remain onscreen until a response is made, and this procedure is repeated for each of the target faces. The learning stage is preceded by a practice session which adopts the same format but uses cartoon faces instead of real images of children.

< Insert Figure 3 >

Following the learning stage, children review the target faces for 20 seconds (all faces are presented on screen simultaneously, from a frontal view), before proceeding to the second stage. The procedure for the second stage is identical to the test trials from the first stage: three faces are presented simultaneously, and children are asked to choose which identity is one of the target faces. However, in this stage the target images are presented from a previously unseen viewpoint, and participants are not aware which of the target faces will appear in each trial. After the second stage, participants review the faces once more for 20 seconds (identical to the post-learning review), before proceeding to the final stage. In the final stage, the procedure for each trial is identical to the second stage, except the faces are overlaid with visual noise.

The bike memory test (Bennetts et al., 2017) follows an identical structure to the CFMT-K, except all of the stimuli are bicycle images, extracted from an online store. All bicycle images were converted to greyscale and edited to remove any distinctive branding or design elements (e.g. logos; see Figure 3).

The authors of the CFMT-K produced two versions of the test. Following their recommendations, the number of target faces and bicycles and trials differed for younger children (under seven years of age at beginning of testing) and older children (those aged seven years or older). Younger children learnt four target faces/bicycles, and completed 48 trials in total (12 in the learning stage; 20 in the test stage with novel images; and 16 in the test stage with noise). Older children learnt 6 target faces/bicycles, and completed 72 trials in total (18 in the learning stage; 30 in the test stage with novel images; and 26 in the test stage with noise). This change was implemented because younger children are likely to perform worse than older children at memory-based tests due to differences in general cognitive factors (e.g., concentration) as well as any potential differences in face-specific processing. Using identical tests can result in restriction-of-range effects (i.e. ceiling or floor performance in some groups), which limit comparisons between groups and hinder interpretation of results, especially when the measure of interest is a difference between conditions (e.g. pre- and post-training scores; see Crookes & McKone, 2009, for a discussion of restriction of range effects in developmental studies of face recognition). Thus, one way to minimise these effects is to match tasks according to difficulty by varying the number of faces to be memorised for different age groups (e.g. Crookes & Robbins, 2014) - the approach adopted in the CFMT-K and the bike memory test[1].

---

[1]The choice to change the number of target faces at seven years of age was also supported by previous work using the CFMT-K and the bike memory test (Bennetts et al., 2017). That study administered the easier (four-item) versions to children in UK Year 3 (roughly aged 7.5-8.5 years), who showed a reasonably high level of performance in both the face and bicycle versions of the tests. However, we chose to use the six-item versions with this age group since (a) we were not aiming to identify children with significant deficits in this study;

The face and bicycle matching tasks were previously validated by Bennetts et al. (2017), adopting a 3AFC simultaneous matching design (see Figure 4). A single target image is presented at the top of the screen, and three test images are presented below. The target image is always a frontal face or side-view of a bike; test images differ from targets in viewpoint and/or lighting conditions, to prevent matching based on low-level image characteristics. Participants are asked to pick which of the test images is the same identity as the target image, and respond using the 1, 2, and 3 keys on their keyboard. Images remain onscreen until the participant responds. All of the images used in the face and bicycle matching tasks were extracted from the memory tasks. There were 30 trials in total for all participants, regardless of age. Both accuracy and reaction time were recorded for this task.

< *Insert Figure 4* >

As these tasks are known to be highly reliable (Bennetts et al., 2017) and are the best available resources for measuring face memory/perception in children, they were selected for this study. However, multiple versions are not currently available, so we re-administered the identical tasks at each assessment stage of the study (see below). While test-retest effects were expected, the use of a control training condition, and object as well as face tasks, allowed us to examine whether performance on the face tests improved in the experimental group over and above the control group, and in faces over and above objects.

(b) we planned on administering the tests multiple times (and hence wanted to leave adequate room for practice effects as well as training effects); and (c) pilot testing (12 children, not included in the training study) indicated that 7- and 8-year olds were capable of completing the six-item versions with accuracy substantially above chance levels (mean accuracy 65-69%).

Procedure

Participants received a new unboxed version of the *Guess Who* game, alongside detailed instructions summarising the training and assessment procedure. Those in the experimental condition additionally received the pack of 40 new stimulus cards. Parents and guardians initially discussed the study protocols with an experimenter, and could contact a member of the research team with additional questions at any point during the process.

All participants initially completed the four assessment tasks (face and bike memory and matching) online. The two memory tasks were always completed first, with the order of the face and bike tests counterbalanced for both memory and matching. Parents and guardians were asked to ensure that their child understood the instructions of each task and maintained attention, but not to cue them to correct answers. Participants then immediately entered a 14 day training period, in which they were required to play the relevant version of *Guess Who* for at least half an hour per day, on any 10 of the 14 days. Most children played against their parents or guardians, and some played with appropriately-aged siblings.

Children in the control condition simply played the commercial version of the game, repeatedly using the same cartoon stimuli cards throughout the entire period (i.e. they did not progress to any further levels of difficulty). Implementing an active control condition which involved similar social interactions to the training (i.e. it required children to discuss and discriminate between faces in the context of the game) ensured that any differences in outcome were attributable to the training materials, rather than the game format causing increased attention to faces.

Children in the experimental condition used the set of 40 new stimuli cards. They began with the male version of Level 1, and followed the standard instructions of the game (i.e. selecting their character, and then taking turns to ask yes/no questions about their opponent's character until only one window remained open). The participant was deemed to have correctly performed the task whenever a successful guess was made by either of the two players (the participating child would need to answer the questions of their opponent correctly in order for them to win). When the game was successfully completed on two consecutive occasions, the players switched their cards to the Level 1 female version. When that level had been won successfully on two consecutive occasions they proceeded to Level 2, and so on, alternating between the male and female versions at each level.

Following the 14 day training period, participants in both conditions immediately completed the four face and bike memory and matching assessment tasks a second time (the same protocols were used for order and counterbalancing as in the pre-assessment). Both experimental and control participants were then asked to refrain from playing any version of *Guess Who* for the next four weeks, after which they completed the face (but not the bike) memory and matching tasks a final time (memory test first), to see if any face-processing gains were maintained. The entire procedure is summarised in Figure 5.

< Insert Figure 5 >

Statistical analyses

All accuracy scores were converted to percentage correct, allowing us to directly compare performance across the two age-appropriate versions (aimed at 4-6 or 7-11 year-olds) of the memory tasks. Because we adopted a pseudo random allocation procedure at recruitment, approximately equal numbers of children took part in these two age-versions across the control and experimental training conditions. Initial analyses explored whether the change in age-version was appropriate. Next, a MANOVA was carried out on face and bike memory and matching scores (entering age version as a fixed factor), across pre- and post- assessment results. This allowed us to assess whether post-training gains in performance were greater in the experimental compared to the control condition, and whether they were greater for faces than for bikes.

Because accuracy can obscure more fine-grained differences in performance in face perception tasks (e.g. Rossion & Michel, 2018), we also considered reaction times in the matching tasks in a separate ANCOVA (controlling for age). For each participant, average response latency was calculated for correct trials, excluding those that differed from the participant's mean response latency by more than three standard deviations, or were less than 150 ms in duration.

To assess whether any gains in face-processing performance were maintained, planned linear and quadratic contrasts, with appropriate follow-up analyses, were performed on the accuracy and reaction time face memory/matching measures across the three time-points (pre-training, post-training and one-month follow-up). Finally, we examined whether age or individual differences in face-processing ability at entry (i.e. performance on the pre-assessment tasks) influenced training gains in the experimental group. All data are available for public download (https://osf.io/2gjbr/).

Results

Engagement and progression

All participants in both conditions were reported by their parents or guardians to have completed the training as instructed (i.e. 30 minutes per day, on any 10 days within the 14 day period). Most children in the experimental condition progressed to the highest levels within this time period. Thus engagement with training, either via time spent training or the highest level achieved, could not be reliably analysed in relation to gains.

One child in the control group did not complete the pre-training face or bike memory assessments, and one participant in the experimental group did not complete the post-training face matching assessment. Eight children (five in the experimental condition) did not complete the follow-up assessment tasks. All other data were retained from these participants to protect sample size. One participant from the experimental group achieved very low accuracy scores (more than three SDs from the mean) on most assessment tests, and was excluded from all analyses.

Before carrying out further analysis, we examined whether the change in version of the memory test (i.e. the increase of two further target stimuli for children aged seven years and older) would adversely influence data interpretation. Figure 6 displays performance on the two memory tasks at the first assessment stage by age (i.e. before any training had begun), indicating that there is no obvious difference in face memory between the two ages where the version switch occurred (i.e. between six and seven years). This was supported by a non-significant independent samples *t*-test ($t(23) = 0.479$, $p = .636$), bolstering our decision to increase memory load at this age. However, there was a sharp increase in accuracy for the bike task, $t(23) = 7.035$, $p = .001$, $d = 2.44$, indicating that seven year old children found the task easier than younger children, even when task demands increased. In support, a significant correlation with age was only observed for bike memory ($r = .751$, $p = .001$), and not for any of the other tests (all *r*s < .245; sequential Bonferroni correction applied). To further explore this effect, age group was entered as a fixed effect in all subsequent analyses.

< Insert Figure 6 >

Accuracy scores: pre- versus post-training

A 2 (condition: experimental, control) x 2 (age group: younger, older) x 2 (stimulus: faces, bikes) x 2 (assessment: pre-training, post-training) MANOVA, with repeated measurements on the “assessment” and “stimulus” factors, was carried out on memory and matching accuracy scores. The four-way interaction was not significant, $F(2,74) = 0.122$, $p = .885$; however, the predicted three-way interaction between condition, stimulus and time emerged, $F(2,74) = 5.676$, $p = .005$, $\eta\rho2 = .133$. This interaction superseded a significant interaction between stimulus and time, $F(2,74) = 44.310$, $p = .001$, $\eta\rho2 = .545$, and the main effects of stimulus and time: $F(2,74) = 35.254$, $p = .001$, $\eta\rho2 = .488$ and $F(2,74) = 94.215$, $p = .001$, $\eta\rho2 = .718$, respectively. There was also a significant interaction between stimulus and age group, $F(2,74) = 40.739$, $p = .001$, $\eta\rho2 = .524$, and a main effect of age group, $F(2,74) = 63.539$, $p = .001$, $\eta\rho2 = .632$. All other interactions, and the main effect of condition, were non-significant (all *p*s > .05).

Memory performance

Univariate analyses indicated that the multivariate three-way interaction between condition, stimulus and time was upheld for memory: $F(1,75) = 8.424$, $p = .005$, $\eta\rho2 = .101$ (see Figure 7A). The memory interaction was followed up with two 2 (condition) x 2 (time) ANOVAs, each considering the face or bike memory data. For face memory, there was a significant interaction between time and condition, $F(1,78) = 6.561$, $p = .011$, $\eta\rho2 = .078$, superseding a main effect of time but not condition: $F(1,78) = 171.176$, $p = .001$, $\eta\rho2 = .687$ and $F(1,78) = 1.660$, $p = .201$ (see Figure 7A). A follow-up *t*-test confirmed that the gain in face memory performance was greater in the experimental compared to the control group: $t(78) = 2.561$, $p = .012$, $d = 0.57$. The same ANOVA for bike memory revealed a significant main effect of time, $F(1,78) = 20.365$, $p = .001$, $\eta\rho2 = .207$, but no main effect of condition nor interaction between the two: $F(1,78) = 0.049$, $p = .825$ and $F(1,78) = 1.543$, $p = .218$, respectively.

< Insert Figure 7 >

The multivariate interaction between stimulus and time was supported by univariate memory data: $F(1,75) = 79.380$, $p = .001$, ηρ2 = .514. A follow-up ANOVA confirmed that the difference between pre- and post-assessment scores was larger for faces (M = 19.79, SE = 1.52) than bikes (M = 4.42, SE = 0.91): $F(1,76) = 78.007$, $p = .001$, ηρ2 = .507. This finding superseded a main effect of time that indicated higher accuracy scores for post- (M = 73.99, SE = 1.24) compared to pre- (M = 61.98, SE = 1.27) assessments: $F(1,75) = 174.927$, $p = .001$, ηρ2 = .70; and a main effect of stimulus that indicated higher scores for faces (M = 74.44, SE = 1.69) than bikes (M = 61.53, SE = 1.06: $F(1,75) = 66.015$, $p = .001$, ηρ2 = .468.

The significant multivariate interaction between stimulus and age group was also upheld by memory data: $F(1,75) = 82.078$, $p = .001$, ηρ2 = .523. Younger children scored significantly lower on bikes (M = 44.11, SE = 0.52) than faces (M = 71.41, SE = 2.58), but there was no difference between bike (M = 78.95, SE = 2.09) and face (M = 77.46, SE = 2.18) scores for older children, $F(1,38) = 139.023$, $p = .001$, ηρ2 = .785 and $F(1,37) = 0.468$, $p = .498$, ηρ2 = .012, respectively. This finding superseded the univariate main effect of age, $F(1,75) = 76.513$, $p = .001$, ηρ2 = .505; whereby older children (M = 78.21, SE = 1.67) achieved higher scores on the memory tests than younger children (M = 57.76, SE = 1.64).

Matching performance

Univariate analyses indicated that the multivariate three-way interaction between condition, stimulus and time was not upheld for matching: $F(1,75) = 2.037$, $p = .158$ (see Figure 7B). While the multivariate interaction between stimulus and time was supported by matching data: $F(1,75) = 4.999$, $p = .028$, ηρ2 = .062, a follow-up ANOVA did not find that the critical difference was between pre- and post-assessment scores for bike (M = 7.39, SE = 1.34) versus face (M = 4.36, SE = 1.22) matching, $F(1,76) = 3.563$, $p = .063$. However, a main effect of time indicated higher accuracy scores for post- (matching: M = 84.99, SE = 1.32) compared to pre- (M = 79.17, SE = 1.33) assessments: $F(1,75) = 33.290$, $p = .001$, ηρ2 = .307; and a main effect of stimulus indicated higher scores for faces (M = 85.01, SE = 1.47) than bikes (M = 79.15, SE = 1.31): $F(1,75) = 19.344$, $p = .001$, ηρ2 = .205.

The significant multivariate interaction between stimulus and age group was upheld by matching data: $F(1,75) = 9.474$, $p = .003$, ηρ2 = .112. While younger children scored significantly lower on bikes (M = 76.03, SE = 1.95) than faces (M = 85.99, SE = 2.16), there was no difference between bike (M = 82.28, SE = 1.75) and face (M = 84.03, SE = 1.98) scores in older children, $F(1,38) = 30.334$, $p = .001$, ηρ2 = .444, and $F(1,37) = 0.805$, $p = .375$, respectively. Finally, there was no main effect of age: $F(1,75) = 0.772$, $p = .382$.

To assess whether reaction times provide more sensitive insights into face matching performance, a 2 (stimulus) x 2 (time) x 2 (condition) ANCOVA was performed on the pre- versus post-training reaction time data, controlling for participant age. The three-way interaction was non-significant, $F(1,77) = 1.202$, $p = .276$ (see Figure 7C), as were all three two-way interactions (ps > .23).

A significant main effect of stimulus indicated that faces (M = 5059.16 ms, SE = 226.30) were matched more rapidly than bikes (M = 6159.45 ms, SE = 277.34), $F(1,77) = 21.242$, $p = .001$, ηρ2 = .216. Unsurprisingly, a main effect of time demonstrated that reaction times were more rapid in the post- (M = 5021.46 ms, SE = 210.96) compared to pre- (M = 6197.12 ms, SE = 319.86) assessment, $F(1,77) = 11.830$, $p = .001$, ηρ2 = .133. There was no main effect of condition, $F(1,77) = 1.295$, $p = .259$.

Face memory gains: follow-up

To assess whether gains in face memory were maintained over time, data from the three time points were entered into a 3 (time: pre-assessment, post-assessment, one-month follow-up) x 2 (condition: experimental, control) x 2 (age group: younger, older) ANOVA. The three-way interaction was non-significant, nor was the critical time*condition interaction, $F(2,136) = 0.735$, $p = .481$ and $F(2,136) = 2.926$, $p = .057$. Other than a main effect of time, $F(1,68) = 150.833$, $p = .001$, $\eta\rho 2 = .689$, all other interactions and main effects were not significant (all ps > .112). Likewise, planned linear and quadratic contrasts revealed no effect of age in the three-way interactions, $F(1,68) = 0.881$, $p = .351$ and $F(1,68) = 0.435$, $p = .512$, respectively. For the two-way condition*time analysis, both the linear and quadratic contrasts were again non-significant: $F(1,68) = 2.474$, $p = .120$ and $F(1,68) = 3.850$, $p = .054$, respectively (see Figure 8). A t-test on the difference between post-training and follow-up scores indicated no difference between the experimental and control groups: $t(71) = 0.634$, $p = .528$. That is, the slight further improvement in face accuracy scores in both conditions likely results from test-retest effects, and no drop-off in gains were observed in the experimental group.

< Insert Figure 8 >

Influences on gains

Our initial literature review identified two factors that may influence training outcome: participant age and individual differences in face recognition ability at entry. Thus, a multiple regression was performed to investigate whether participant age or score on the face memory pre-assessment task predicted memory improvement scores (i.e. the difference between post- and pre-training performance) for participants in the experimental training condition (N = 40). The model explained 43.4% of the variance, and was a significant predictor of training gains, $F(2,37) = 15.967$, $p = .001$. While individual differences in performance at entry significantly predicted training gains ($\beta = .696$, $p = .001$; see Figure 9), age did not contribute to the model, nor did it correlate with gains ($\beta = .109$, $p = .381$; $r = -.045$, $p = .785$).

< Insert Figure 9 >

Discussion

This study investigated the efficacy of a face training programme in improving face versus object memory and matching skills in typical children aged between four and 11 years. Compared to an active control condition, domain-specific gains in face memory but not face matching were observed post-training, and were maintained at a one-month follow-up. Further analyses revealed that the gains in face memory were consistent across age in the experimental condition, but were greater in the poorest perceivers.

This study makes a number of novel contributions to the literature on face recognition and rehabilitory training. First, this investigation provides a novel face training programme that is attractive to children, adapting a renowned game that encourages off-screen family interaction. Critically, this training programme resulted in an average 7.6% improvement in face memory performance over and above the active control condition (i.e. when the practice effect on the assessment task and/or gains from the control training procedure were eliminated; note that these cannot be dissociated from the available data). Given a significant improvement of this magnitude in typical children, coupled with the finding that gains were greater in the poorest perceivers at baseline, the programme presents a promising, family-friendly means of improving face memory in children with prosopagnosia.

Notably, all aspects of this research programme were completed within the child’s home environment, unaccompanied by a member of the research team. Instead, parents and guardians were asked to monitor compliance, to participate in or supervise training (e.g. when the child played with a sibling or friend), and to ensure understanding during the assessment tasks. There are clearly benefits of this model, particularly in terms of resourcing roll-out of the programme on a wider scale, and in overcoming geographical restraints. This is a particularly important point given increasing numbers of individuals are self-reporting with prosopagnosia (Bate et al., 2019; Bennetts et al., 2017), and some children experience severe socio-emotional consequences of the condition (Adams et al., in press; Dalrymple et al., 2014; Murray et al., 2018). However, it is also inevitable that this approach introduced some inaccuracies to the data, with parental feedback on participant engagement and level of assistance being particularly vulnerable to error. Yet, the reasonably generous sample size used here, the use of random allocation to training conditions, and the size of the improvement following training give us confidence in our data.

Indeed, the size of the improvement in face memory is particularly encouraging in relation to comparable findings reported by previous work. Davies-Thompson et al. (2017) reported an average ~10% improvement in CFMT scores in 10 adults with acquired prosopagnosia, but this was not significant in a group-level analysis. Although CFMT gains were not available in the studies reported by DeGutis and colleagues (2007, 2014), Bate et al. (2015) did not detect an improvement on this test for their adolescent case with acquired prosopagnosia. Whether the larger gains reported here result from the training strategy itself (i.e. the combination of strategies that have been somewhat successful in previous work), participant age (see discussion below), the inclusion of typical rather than impaired participants, the larger sample size, or a combination of some or all of these factors, is unknown. Further, one challenge that has not been addressed here is the transfer of gains into real-world everyday face recognition performance – a skill that is hard to tap in children, particularly those without impairments.

Second, we found evidence that the gains in face memory did not generalise to non-face objects (in this case, bicycles). Any increases in bike memory and matching did not significantly differ between the experimental and control training conditions, and were particularly small for the memory test. This finding indicates that training was targeting the actual visuocognitive processing strategies that underpin unfamiliar face memory, supporting domain-specific hypotheses of face-processing (e.g. Kanwisher, 2017). In support of the idea that visual experience or expertise in one category of objects does not necessarily have a domain-general effect on all visual discrimination abilities, a recent study found that extensive childhood visual experience with relatively homogenous object categories (Pokemon characters) can shape responses in the ventral temporal cortex during adulthood, while responses to face stimuli in face-selective areas were not affected by Pokemon expertise (Gomez, Barnett, & Grill-Spector, 2019). However, we only included one comparison object in the current study, and it cannot be ruled out that other categories of object may have also gained from training. Research into the development of the human visual cortex has proposed that certain visual qualities, such as eccentricity bias, rectilinearity, and animacy may determine the distribution of responses to different objects in the ventral temporal cortex (e.g. see Gomez et al., 2019). We did not attempt to match our non-face object on these qualities – instead, we focused on using stimuli that would be familiar and engaging to children in the age range we examined, that are clearly discriminable on an individual (as opposed to category) level, and that were reasonable well-matched in difficulty to the face tasks. However, future studies may consider using non-face comparison objects which vary parametrically on the properties identified above when examining the effects of face training programmes, as this has the capacity to shed further light on the question of domain specificity, as well as the development and organisation of the visual cortex.

Third, it is pertinent that both performance at entry and training-gains in face memory were consistent across all ages, yet performance on the bike memory task improved with age. This finding indicates consistent plasticity for face memory across the target age range, and argues against evidence suggesting a limited window of plasticity in early infancy (e.g. Geldart et al., 2002; Pascalis et al., 2002). Instead, it is conceivable that the face-processing system is fully developed by the age of our youngest participants, supporting existing hypotheses of early maturation (e.g. McKone et al., 2012), yet still open to modification. Indeed, when examining face memory performance at entry to the programme (i.e. at the pre-training assessment), there was consistency in performance from the age of five years. This finding also bolsters claims of domain-specificity, at least in comparison to our target object category: neither face memory per se, nor gains in face memory, were assisted by age-related developments in more generalised processes, whereas this was the case for bike memory. Thus, face and object processing may at least begin to segregate before the age of five years, with early maturation for faces but continued development for objects as more generalized processes continue to come online.

In terms of neurorehabilitation, this finding suggests consistent levels of plasticity from early childhood through to pre-adolescence, with no benefits of earlier intervention. This suggests that the face recognition system remains open to modification during the target age range, and there is no critical window for younger children. However, it is possible that greater gains may be experienced in even younger children than those tested here. Having said this, we do not believe that this training strategy would be suitable for younger children, and alternative intervention techniques for pre-schoolers sorely need to be examined (e.g. by encouraging attention to faces). Further, it remains to be seen whether similar gains are experienced by adolescents and even adults, or whether some drop-off may occur at a particular age.

Pertinently, the greatest gains were observed in the poorest perceivers at entry, akin to the work with adults with acquired prosopagnosia reported by Davies-Thompson et al. (2018). While this finding is particularly encouraging for use of the training programme in clinical participants, it remains unknown whether greater engagement with training may have led to even larger gains. Indeed, we opted for experimental rigour by controlling participant engagement with training via our strict instructions, although admittedly we were dependent on parental feedback concerning adherence. As previous work has found a positive correlation between time spent training and post-training gains in adults with prosopagnosia (DeGutis et al., 2014), investigation of this factor in future work with atypical children may be particularly fruitful.

Fourth, the gains observed following training were specific to face memory, and did not extend to improvements in face matching tasks. It is unclear why we did not observe gains in face matching from the training, particularly as this process is arguably more targeted by the training task itself. It is possible that the face matching task lacks some sensitivity in detecting individual differences in performance in typical, or at least more proficient, children. While performance did begin to approach ceiling post-training in the experimental group, we would expect the analysis of reaction time to be sensitive to improvement even under conditions of high accuracy.

Instead, there are theoretical interpretations of the finding. Weigelt et al. (2014) posited that face perception is fully developed by a young age, but face memory continues to develop throughout childhood. While conflicting data suggests early maturation of both processes (i.e. in the current study and previous work, e.g. Bennetts et al., 2017), it is possible that early individual differences in face perception are either less varied, or more fixed and resilient to change, than those in face memory. Alternatively, it may be that each of these tasks depends on different processes, and training targeted one of these more than the other. Indeed, the manipulations applied to the training stimuli included both featural (i.e. changes to the size of specific facial features) and holistic (affecting the spacing of features) components. Some authors have claimed that face memory tasks (i.e. those using familiar or familiarized faces, including variants of the CFMT) tap holistic processes, whereas matching tasks depend more on image-level characteristics or featural processing (Hancock, Bruce, &

Burton, 2000; Megreya & Burton, 2006). While the training tasks varied both the spacing between features and the size of the features themselves, it is important to note that both of these manipulations affect the relationship between features – an important element of holistic or configural processing. It is possible that this focus on relational information in the training task strengthened holistic or configural processing to a greater extent than featural processing, resulting in the facilitation only in the memory task. Alternatively, it may be that featural processing or face perception is already fairly well-developed in children, but holistic/configural processing or face memory is still developing and more amenable to modification (de Heering, Rossion, & Maurer, 2012; Mondloch, Le Greand, & Maurer, 2002).

In sum, this study presents a new face training programme that is engaging for children, and can improve face memory by ~7% in typical children aged 4-11 years, with larger effects for poorer perceivers. Training outcomes advocate early independence of both face perception versus face memory, and face versus object processing.

Context of the Research

This investigation follows an ongoing line of research within our laboratory that has investigated the remediation of face recognition deficits. Our previous work has focused on adults and adolescents with acquired or developmental forms of prosopagnosia, and has resulted in mild-to-moderate gains in performance. Here, we were motivated to investigate whether greater gains could be made in children, inspired by neurorehabilitation theories of early plasticity. We also wished to develop a novel innovative face training programme that can be enjoyed within social off-screen contexts. The results reported here are promising, indicating that a relatively large gain in face memory can be observed in typical children following just ten 30-minute training sessions. Given the greatest gains were also observed in the weakest perceivers, it is possible that larger effects may be observed in children with prosopagnosia. Our ongoing work is now evaluating this possibility, together with the effects of extended training time. We hope that this work will deliver a novel, more successful training possibility that can be made available to children irrespective of their geographical location.

References

Adams, A., Hills, P., Bennetts, R., & Bate, S. (2019). Coping strategies for developmental prosopagnosia. *Neuropsychological Rehabilitation*. DOI: 10.1080/09602011.2019.1623824.

Bate, S., & Bennetts, R. J. (2014). The rehabilitation of face recognition impairments: a critical review and future directions. *Frontiers in Human Neuroscience*, *8*, 491.

Bate, S., Bennetts, R., Gregory, N.J., Tree, J., Murray, E., Adams, A., …Banissy, M. (2019). Objective patterns of face recognition deficits in 165 adults with self-reported developmental prosopagnosia. *Brain Sciences, 9*, 133.

Bate, S., Bennetts, R., Mole, J. A., Ainge, J. A., Gregory, N. J., Bobak, A. K., …Bussunt, A. (2015). Rehabilitation of face-processing skills in an adolescent with prosopagnosia: Evaluation of an online perceptual training programme. *Neuropsychological Rehabilitation: An International Journal*, *25*(5), 733-762.

Bate, S., Bennetts, R., Tree, J.J., Adams, A., & Murray, E. (in press). The domain-specificity of face matching impairments in 40 cases of developmental prosopagnosia. *Cognition*.

Bate, S., & Tree, J.J. (2017). The definition and diagnosis of developmental prosopagnosia. *Quarterly Journal of Experimental Psychology, 70*, 193-200.

Behrmann, M., & Plaut, D.C. (2015). A vision of graded hemispheric specialization. *Annals of the New York Academy of Sciences, 1359*, 30-46.

Bennetts, R. J., Mole, J., & Bate, S. (2017). Super-recognition in development: a case study of an adolescent with extraordinary face recognition skills. *Cognitive Neuropsychology*, *34*, 357-376.

Bennetts, R.J., Murray, E., Boyce, T., & Bate, S. (2017). Prevalence of face recognition deficits in middle childhood. *Quarterly Journal of Experimental Psychology, 70,* 234-258.

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

Bruce, V., Bindemann, M., & Lander, K. (2018). Individual differences in face perception and person recognition. *Cognitive Research: Principles and Implications, 3*, 18.

Bruce, V., & Young, A. (1986). Understanding face recognition. *British Journal of Psychology, 77*, 305-327.

Brunsdon, R., Coltheart, M., Nickels, L., & Joy, P. (2006). Developmental prosopagnosia: A case analysis and treatment study. *Cognitive Neuropsychology, 23*, 822-840.

Carey, S. (1992). Becoming a face expert. *Philosophical Transactions of the Royal Society of London, Series B, Biological Sciences, 29*, 95-102.

Crookes, K., & McKone, E. (2009). Early maturity of face recognition: No childhood development of holistic processing, novel face encoding, or face-space. *Cognition, 111*, 219-247.

Crookes, K., & Robbins, R. A. (2014). No childhood development of viewpoint-invariant face recognition: evidence from 8-year-olds and adults. *Journal of Experimental Child Psychology, 126*, 103-111.

Dalrymple, K.A., Corrow, S., Yonas, A., & Duchaine, B. (2012). Developmental prosopagnosia in childhood. *Cognitive Neuropsychology, 29*, 393–418.

Dalrymple, K.A., Elison, J. T., & Duchaine, B. (2017). Face-specific and domain-general visual processing deficits in children with developmental prosopagnosia. *The Quarterly Journal of Experimental Psychology, 70*, 259-275.

Dalrymple, K.A., Fletcher, K., Corrow, S., dasNair, R., Barton, J., Yonas, A. & Duchaine, B. (2014). "A room full of strangers every day": The psychosocial impact of developmental prosopagnosia on children and their families. *Journal of Psychosomatic Research, 77*, 144-150.

Dalrymple, K.A., Garrido, L., & Duchaine, B. (2014). Dissociation between face perception and face memory in adults, but not children, with developmental prosopagnosia. *Developmental Cognitive Neuroscience, 10*, 10-20.

Dalrymple, K. A., & Palermo, R. (2016). Guidelines for studying developmental prosopagnosia in adults and children. *Wiley Interdisciplinary Reviews: Cognitive Science, 7*(1), 73-87.

Davies-Thompson, J., Fletcher, K., Hills, C., Pancaroglu, R., Corrow, S.L., & Barton, J.J. (2017). Perceptual learning of faces: A rehabilitative study of acquired prosopagnosia. *Journal of Cognitive Neuroscience*, *29*, 573-591.

DeGutis, J.M., Bentin, S., Robertson, L.C., & D’Esposito, M. (2007). Functional plasticity in ventral temporal cortex following cognitive rehabilitation of a congenital prosopagnosic. *Journal of Cognitive Neuroscience*, *19*, 1790-1802.

DeGutis, J., Cohan, S., & Nakayama, K. (2014). Holistic face training enhances face processing in developmental prosopagnosia. *Brain*, *137*, 1781-1798.

De Heering, A., Rossion, B., & Maurer, D. (2012). Developmental changes in face recognition during childhood: Evidence from upright and inverted faces. *Cognitive Development, 27*(1), 17-27.

Duchaine, B., & Nakayama, K. (2005). Dissociations of face and object recognition in developmental prosopagnosia. *Journal of Cognitive Neuroscience, 17*, 249-261.

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

Elbert, T., Heim, S., & Rockstroh, B. (2001). Neural plasticity and development. In C.A. Nelson & M. Luciana (Eds.), *Handbook of Developmental Cognitive Neuroscience* (pp. 191-202). Cambridge, MA: MIT Press.

Ellis, H.D., & Young, A.W. (1988). Training in face-processing skills for a child with acquired prosopagnosia. *Developmental Neuropsychology, 4*, 283–294.

Garrido, L., Furl, N., Draganski, B., Weiskopf, N., Stevens, J., Tan, G. C. Y., ... Duchaine, B. (2009). Voxel-based morphometry reveals reduced grey matter volume in the temporal cortex of developmental prosopagnosics. *Brain, 132*, 3443-3455.

Geldart, S., Mondloch, C.J., Maurer, D., de Schonen, S., & Brent, H.P. (2002). The effect of early visual deprivation on the development of face processing. *Developmental Science, 5*, 490-501.

Geskin, J., & Behrmann, M. (2018). Congenital prosopagnosia without object agnosia? A literature review. *Cognitive Neuropsychology, 35*, 4-54.

Gomez, J., Barnett, M., & Grill-Spector, K. (2019). Extensive childhood experience with Pokémon suggests eccentricity drives organization of visual cortex. *Nature Human Behaviour.* https://doi.org/10.1038/s41562-019-0592-8

Hancock, P. J., Bruce, V., & Burton, A. M. (2000). Recognition of unfamiliar faces. *Trends in Cognitive Sciences, 4*(9), 330-337.

Huttenlocher, P.R. (2002). *Neural plasticity: The effects of environment on the development of the cerebral cortex*. Cambridge, MA: Harvard University Press.

Kanwisher, N. (2017). The quest for the FFA and where it led. *Journal of Neuroscience, 37*, 1056-1061.

McKone, E., Crookes, K., Jeffery, L., & Dilks D.D. (2012). A critical review of the development of face recognition: Experience is less important than previously believed. *Cognitive Neuropsychology, 29*, 174-212.

Megreya, A. M., & Burton, A. M. (2006). Unfamiliar faces are not faces: Evidence from a matching task. *Memory & Cognition, 34*(4), 865-876.

Mondloch, C. J., Le Grand, R., & Maurer, D. (2002). Configural face processing develops more slowly than featural face processing. *Perception, 31*(5), 553-566.

Murray, E., Hills, P.J., Bennetts, R.J., & Bate, S. (2018). Identifying hallmark symptoms of developmental prosopagnosia for non-experts. *Scientific Reports, 8*, 1690.

Piepers, D., & Robbins, R. (2012). A review and clarification of the terms "holistic", "configural," and "relational" in the face perception literature. *Frontiers in Psychology, 3*, 559.

Pascalis, O., de Haan, M., & Nelson, C.A. (2002). Is face processing species-specific during the first year of life? *Science, 296*, 1321–1323.

Rezlescu, C., Pitcher, D., & Duchaine, B. (2012). Acquired prosopagnosia with spared within-class object recognition but impaired recognition of degraded basic-level objects. *Cognitive Neuropsychology, 29*(4), 325-347.

Richler, J.J., Palmeri, T.J., & Gauthier, I. (2012). Meanings, mechanisms, and measures of holistic processing. *Frontiers in Psychology, 3,* 553.

Rossion, B., & Michel, C. (2018). Normative data for accuracy and response times at the computerized Benton Facial Recognition Test (BFRT-c). *Behavior Research Methods, 50*, 2442-2460.

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*, 704-729.

Thomas, M.S.C. (2003). Limits on plasticity. *Journal of Cognition and Development*, *4*, 99-125.

Weigelt, S., Koldewyn, K., Dilks, D., Balas, B., McKone, E., & Kanwisher, N. (2014). Domain-specific development of face memory but not face perception. *Developmental Science, 17*, 47-58.

Wilmer, J.B. (2017). Individual differences in face recognition: A decade of discovery. *Current Directions in Psychological Science, 26*, 225-230.

Figure Captions

Figure 1: Image A shows an alternative version of the commercial *Guess Who* game, using cartoon faces (note that the facial images shown are very similar to those presented in the game itself, but the latter cannot be presented due to issues with permissions). The facial stimuli are printed on cards which slot into the accompanying plastic frame: players select their own character from the array at the top (their opponent's task is to guess the identity of this character by asking yes/no questions about their facial appearance), and see the same characters dispersed behind the windows below (windows are closed by the player following their opponent's yes/no responses to the player's yes/no enquiries, until only one window remains open). Image B displays how the plastic frame is positioned on a table, with each opponent viewing their stimuli set on opposite sides of the frame.

Figure 2: The new stimuli set developed for the experimental training condition. Images A and B display the initial set of manipulated male and female characters, respectively (these were designated to be the easiest level of difficulty, i.e. Level 1). Images C and D display the female characters at Levels 5 and 10 of difficulty, respectively.

Figure 3: Examples of the Cambridge Face Memory Test–Kids (CFMT-K) and bicycle memory tasks. CFMT-K images are adapted from Dalrymple, Garrido, et al. (2014).

Figure 4: Example trials from the face and bicycle matching task. Face images are adapted from Dalrymple, Garrido, et al. (2014).

Figure 5: A summary of the training and assessment procedure.

Figure 6: Mean percentage accuracy on the pre-assessment face and bike age memory tests according to age (collapsed across training condition). The increase in memory load is indicated by the dotted line (i.e. for children aged seven years and older).

Figure 7: Percentage accuracy on the face and bike (A) memory and (B) matching tasks pre- and post-training for the experimental and control training groups. Average reaction times for the matching tasks are shown in (C).

Figure 8: Percentage accuracy on the face memory test at the pre-training, post-training, and follow-up time points, for the experimental and control training conditions.

Figure 9: The relationship between pre-training face memory scores and the face memory gain post-training (i.e. post-training score - pre-training score).

Figure 1:

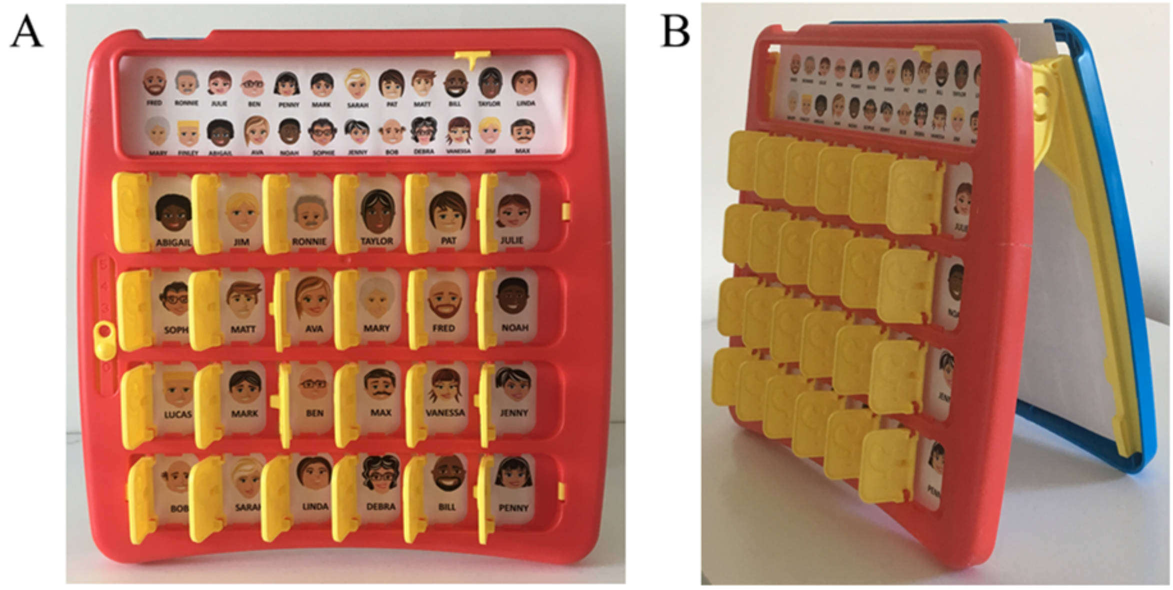

A
B
ABIGAIL
JIM
RONNIE
TAYLOR
PAT
JULIE
MATT
AVA
MARY
FRED
NOAH
LUCAS
MARK
BEN
MAX
VANESSA
JENNY
LINDA
DEBRA
BILL
PENNY

Figure 2:

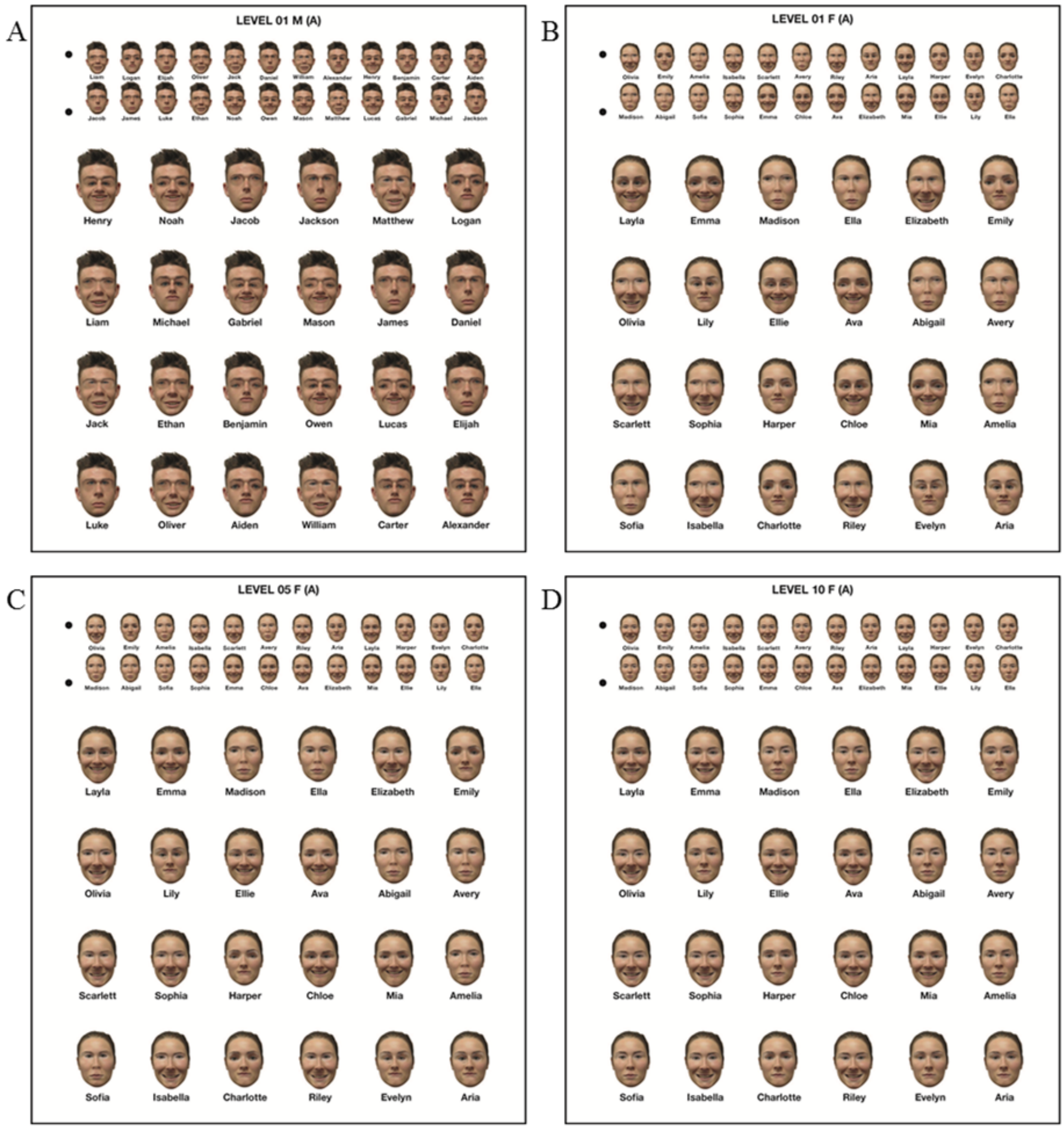

A
LEVEL 01 M (A)
Henry Noah Jacob Jackson Matthew Logan
Liam Michael Gabriel Mason James Daniel
Jack Ethan Benjamin Owen Lucas Elijah
Luke Oliver Aiden William Carter Alexander
B
LEVEL 01 F (A)
Layla Emma Madison Ella Elizabeth Emily
Olivia Lily Ellie Ava Abigail Avery
Scarlett Sophia Harper Chloe Mia Amelia
Sofia Isabella Charlotte Riley Evelyn Aria
C
LEVEL 05 F (A)
Layla Emma Madison Ella Elizabeth Emily
Olivia Lily Ellie Ava Abigail Avery
Scarlett Sophia Harper Chloe Mia Amelia
Sofia Isabella Charlotte Riley Evelyn Aria
D
LEVEL 10 F (A)
Layla Emma Madison Ella Elizabeth Emily
Olivia Lily Ellie Ava Abigail Avery
Scarlett Sophia Harper Chloe Mia Amelia
Sofia Isabella Charlotte Riley Evelyn Aria

Figure 3:

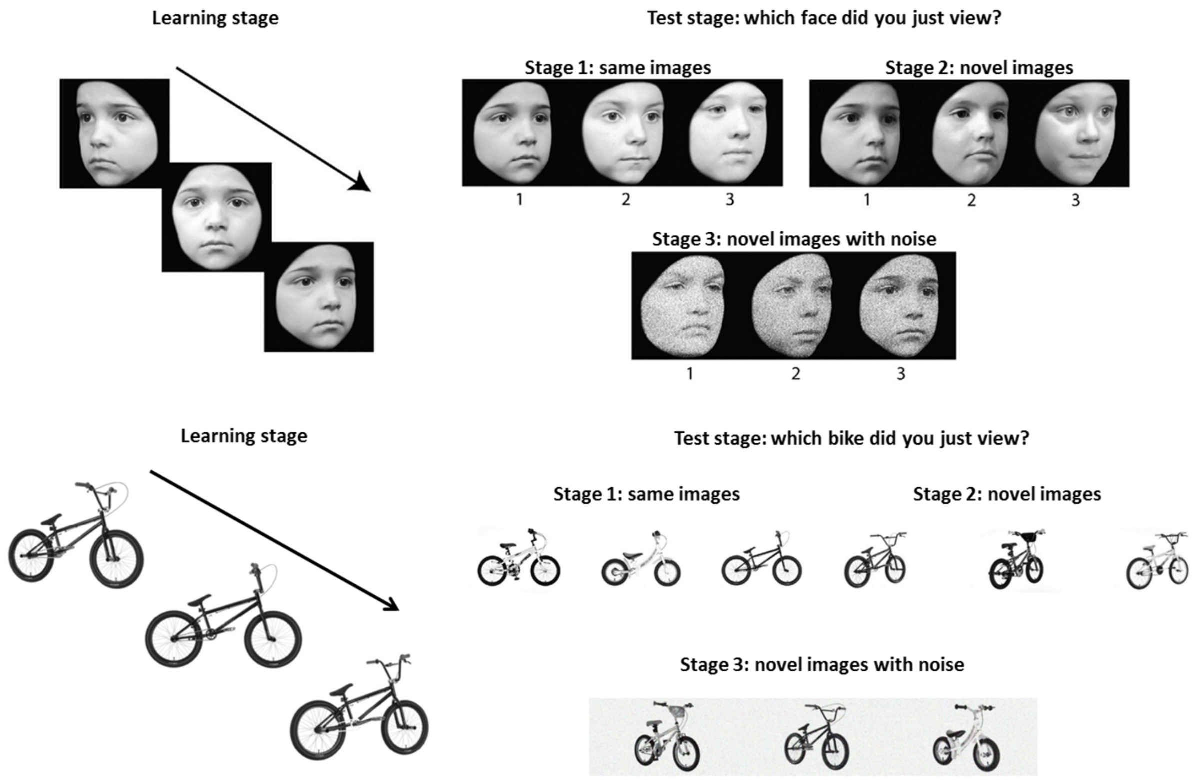

Learning stage
Test stage: which face did you just view?
Stage 1: same images
Stage 2: novel images
1
2
3
1
2
3
Stage 3: novel images with noise
1
2
3
Learning stage
Test stage: which bike did you just view?
Stage 1: same images
Stage 2: novel images
Stage 3: novel images with noise

Figure 4:

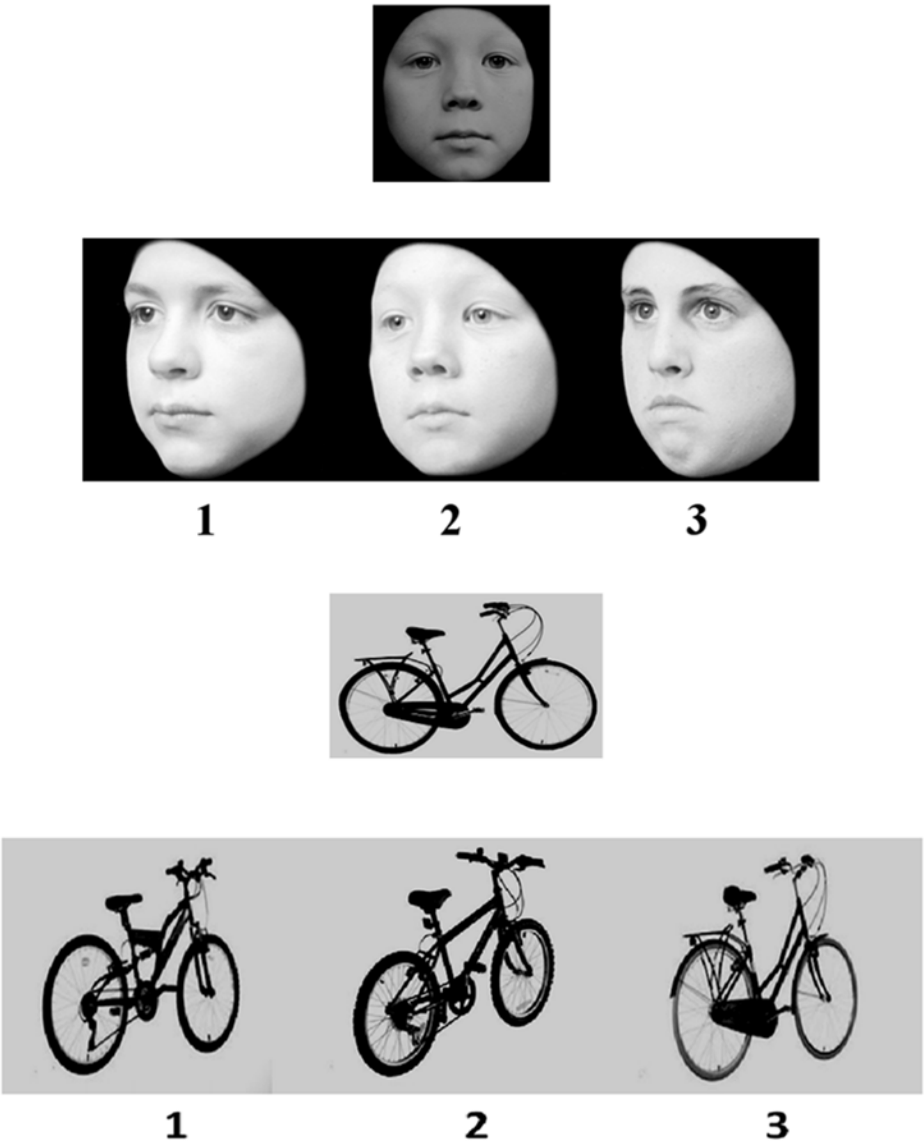


Figure 5:

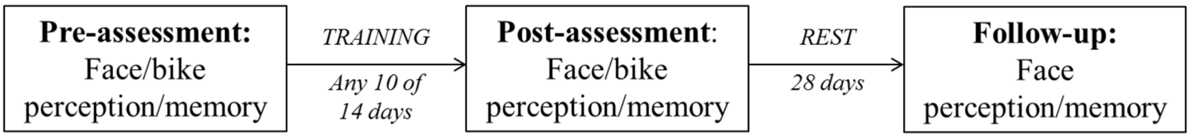

Pre-assessment:
Face/bike perception/memory
TRAINING
Any 10 of 14 days
Post-assessment:
Face/bike perception/memory
REST
28 days
Follow-up:
Face perception/memory

Figure 6:

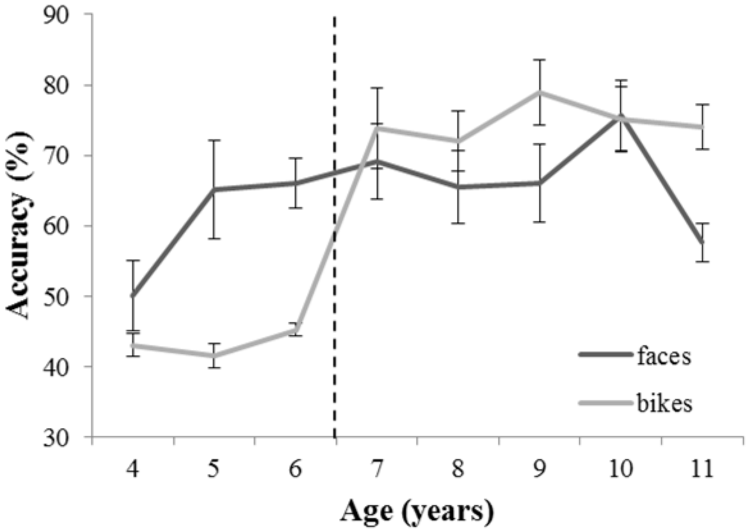

90
80
70
60
50
40
30
Accuracy (%)
4
5
6
7
8
9
10
11
Age (years)
faces
bikes

Figure 7:

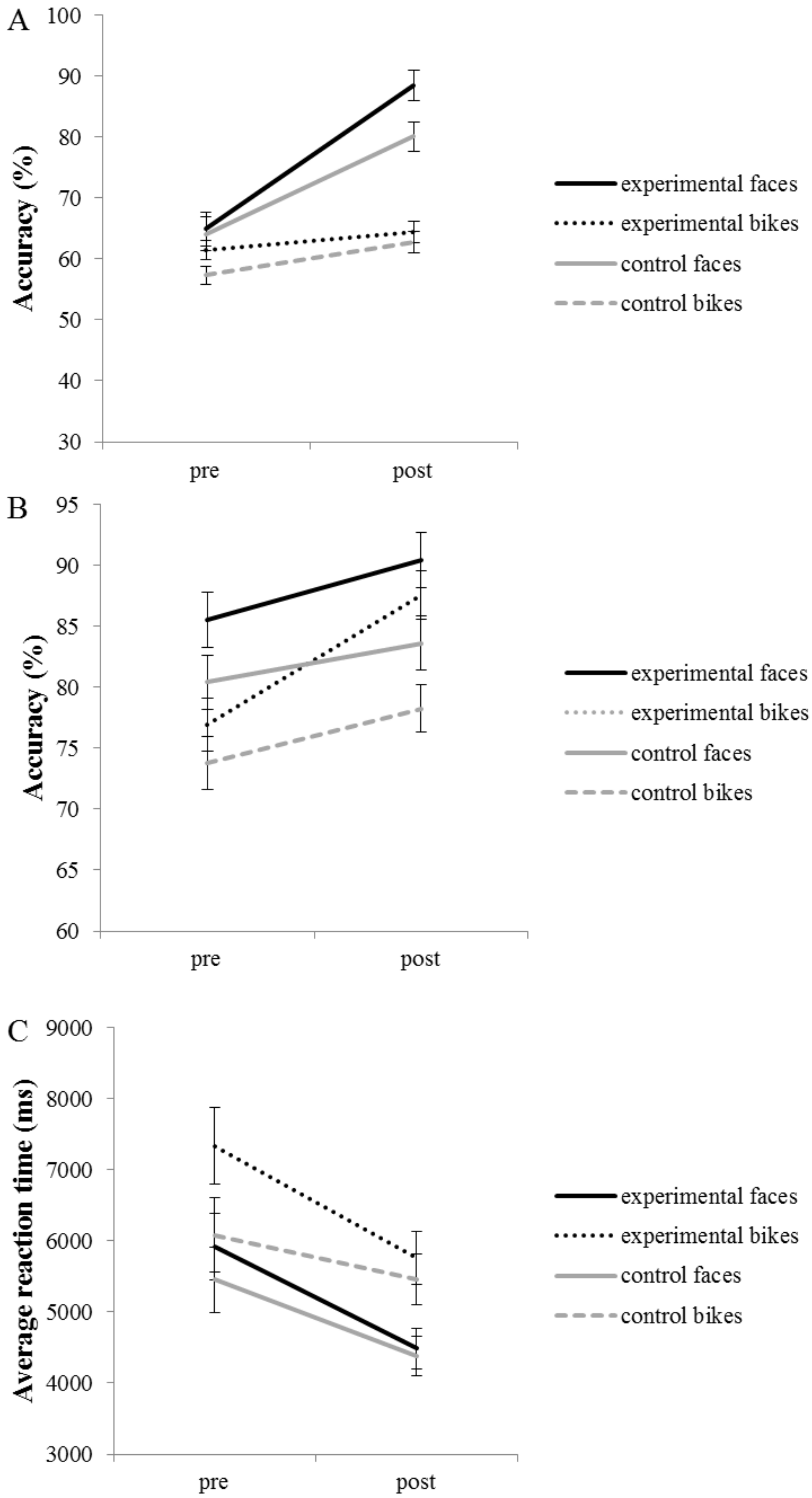

A
100
90
80
70
60
50
40
30
Accuracy (%)
pre
post
experimental faces
experimental bikes
control faces
control bikes
B
95
90
85
80
75
70
65
60
Accuracy (%)
pre
post
experimental faces
experimental bikes
control faces
control bikes
C
9000
8000
7000
6000
5000
4000
3000
Average reaction time (ms)
pre
post
experimental faces
experimental bikes
control faces
control bikes

Figure 8:

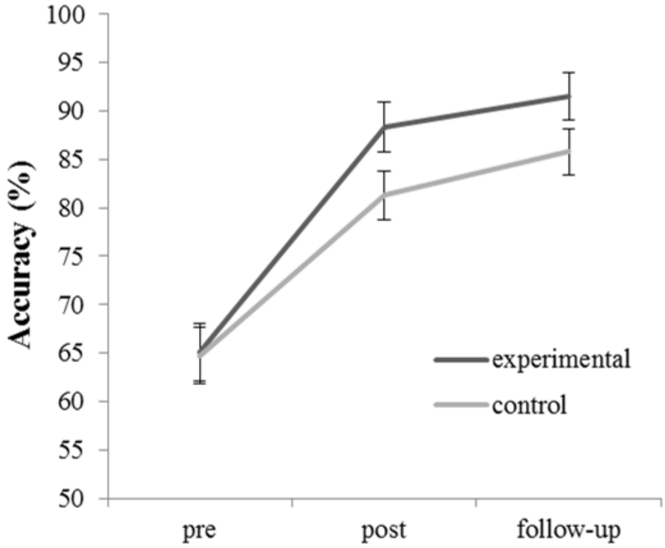
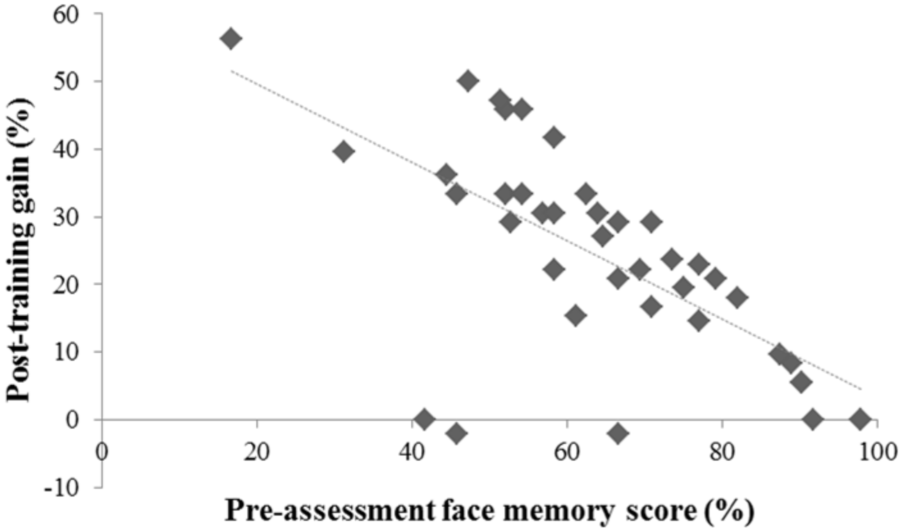

Accuracy (%)
100
95
90
85
80
75
70
65
60
55
50
pre
post
follow-up
experimental
control

Figure 9:



60
50
40
30
20
10
0
-10
0
20
40
60
80
100
Post-training gain (%)
Pre-assessment face memory score (%)