Routledge
Taylor & Francis Group

The Quarterly Journal of Experimental Psychology

ISSN: 1747-0218 (Print) 1747-0226 (Online) Journal homepage: http://www.tandfonline.com/loi/pqje20

Prevalence of face recognition deficits in middle childhood

Rachel J Bennetts, Ebony Murray, Tian Boyce & Sarah Bate

To cite this article: Rachel J Bennetts, Ebony Murray, Tian Boyce & Sarah Bate (2016): Prevalence of face recognition deficits in middle childhood, The Quarterly Journal of Experimental Psychology, DOI: 10.1080/17470218.2016.1167924

To link to this article: http://dx.doi.org/10.1080/17470218.2016.1167924

Publisher: Taylor & Francis & The Experimental Psychology Society

Journal: *The Quarterly Journal of Experimental Psychology*

DOI: 10.1080/17470218.2016.1167924

Prevalence of face recognition deficits in middle childhood

Rachel J Bennetts, Ebony Murray, Tian Boyce, and Sarah Bate

Bournemouth University

Author note

Rachel J Bennetts, Ebony Murray, Tian Boyce, and Sarah Bate, Department of Psychology, Faculty of Science and Technology, Bournemouth University.

Correspondence concerning this article should be addressed to Rachel Bennetts, Department of Psychology, Faculty of Science and Technology, Bournemouth University, Poole House, Talbot Campus, Poole, UK, BH12 5BB.

Email rbennetts@bournemouth.ac.uk.

Word count:

Abstract: 250

Manuscript: 11,785 (excludes references, figure and table captions)

Abstract

Approximately 2-2.5% of the adult population is believed to show severe difficulties with face recognition, in the absence of any neurological injury – a condition known as developmental prosopagnosia (DP). However, to date no research has attempted to estimate the prevalence of face recognition deficits in children, possibly because there are very few child-friendly, well-validated tests of face recognition. In the current study, we examined face and object recognition in a group of primary school children (aged 5-11 years), to establish whether our tests were suitable for children; and to provide an estimate of face recognition difficulties in children. In Experiment 1 (n = 184), children completed a pre-existing test of child face memory, the CFMT-K, and a bicycle test with the same format. In Experiment 2 (n = 413), children completed three-alternative forced choice matching tasks with faces and bicycles. All tests showed good psychometric properties. The face and bicycle tests were well-matched for difficulty and showed a similar developmental trajectory. Neither the memory nor matching tests were suitable to detect impairments in the youngest groups of children, but both tests appear suitable to screen for face recognition problems in middle childhood. In the current sample, 1.2-5.2% of children showed difficulties with face recognition; 1.2-4% showed face-specific difficulties – that is, poor face recognition with typical object recognition abilities. This is somewhat higher than previous adult estimates: it is possible that face matching tests overestimate the prevalence of face recognition difficulties in children; alternatively, some children may "outgrow" face recognition difficulties.

Keywords: Prosopagnosia; face recognition; object recognition; cognitive development; developmental disorders

Prevalence of face recognition deficits in middle childhood

The ability to recognise faces develops substantially throughout the early years of life. From as early as 1-4 days of age, infants are able to learn and discriminate between a small number of faces (Pascalis, de Schonen, Morton, Deruelle, & Fabre-Grenet, 1995; Turati, Macchi Cassia, Simion, & Leo, 2006), and by the time we reach late adolescence, most people are "experts" at recognising a large number of familiar faces with little more than a glance (Hancock, Bruce, & Burton, 2000). However, some individuals never develop the ability to recognise faces – a condition known as developmental prosopagnosia (DP). DP (also known as congenital prosopagnosia or "face-blindness"; see Susilo & Duchaine, 2013 for a discussion of terminology) is a condition characterised by a severe, relatively selective deficit in face recognition, in the absence of any neurological injury (e.g., Bate & Cook, 2012; Bate, Haslam, Jansari, & Hodgson, 2009; Behrmann, Avidan, Marotta, & Kimchi, 2005; Bennetts, Butcher, Lander, Udale, & Bate, in press; Duchaine, Germine, & Nakayama, 2007; Duchaine & Nakayama, 2006). Studies in adult populations suggest that DP occurs in roughly 2% of the population (Bowles et al., 2009; Kennerknecht et al., 2006; Kennerknecht, Ho, & Wong, 2008), but there have been relatively few studies on DP in childhood (Brundson, Coltheart, Nickels, & Joy, 2006; Dalrymple, Corrow, Yonas, & Duchaine, 2012; Dalrymple , Garrido, & Duchaine, 2014; Jones & Tranel, 2001; McConachie, 1976; Schmalzl, Palermo, Green, Brundson, & Coltheart, 2008; Wilson, Palermo, Schmalzl, & Brock, 2010) and no published studies estimating the prevalence of DP in children.

Given that DP can be associated with social anxiety (Davis et al., 2011), feelings of embarrassment, avoidance of or anxiety in social situations (Dalrymple, Fletcher, et al., 2014; Yardley, McDermott, Pisarski, Duchaine, & Nakayama, 2008), and safety concerns (e.g., children misidentifying carers or teachers), it is important to understand how many children are affected by this condition and to develop reliable, valid screening tools with appropriate norms for multiple age groups. Consequently, Experiment 1 presents a new test of object memory (children's bicycles), matched in format to a pre-existing test of child face memory (the Cambridge Face Memory Test – Kids, or CFMT-K; Dalrymple, Garrido, et al., 2014); Experiment 2 presents new tests of face and object matching suitable for children. We used these tests to examine face and object processing in a large sample of UK primary school children (5 -11 yrs). There were two main aims: firstly, to examine the performance and psychometric properties of the face and object tests; and secondly, to provide an initial estimate of the prevalence of DP in children.

Identifying cases of developmental prosopagnosia

Traditionally, prosopagnosia has been identified using behavioural tests of face memory (generally following self-report, e.g., Bate et al., 2009; Behrmann et al., 2005; Bennetts et al., in press; De Gutis, Cohan, Mercado, Wilmer, & Nakayama, 2012; Duchaine, Yovel, & Nakayama, 2007) or, in some cases, self-report of face recognition difficulties alone (e.g., Kennerknecht et al., 2006). In adults, behavioural tests of face memory can be broken into two categories: familiar faces tests (specifically famous faces; e.g., Behrmann et al., 2005; Duchaine, Germine, & Nakayama, 2007) and standardised tests of face learning (e.g., the Benton Facial Recognition Test, Benton, Sivan, Hamsher, Varney, & Spreen, 1983; the Cambridge Face Memory Test or CFMT, Duchaine & Nakayama, 2006). Familiar faces tests are relatively ecologically valid and simple to administer when applied to a limited or homogenous population (i.e., in which all individuals would be expected to show similar levels of familiarity with the majority of famous faces used in the test), but practical difficulties can emerge when trying to test different populations (e.g., different age ranges, different cultural backgrounds) (Bowles et al., 2009). As children have variable and relatively limited exposure to famous faces, these types of test are impractical to design and administer with young children. Previous studies have used tests of personally familiar face recognition with children (e.g., Brundson et al., 2006; Schmalzl et al., 2008), but these can be labour-intensive to develop and present practical difficulties (e.g., finding and accessing a control group appropriately familiar with the same faces). While this approach may be useful for detailed case studies and theoretical research, it is simply not practical for general screening, especially as awareness of prosopagnosia grows and more cases are reported.

Another approach to assessing face memory is standardised tests of face learning. The most well-known of these tests is the CFMT, which tests participants' ability to learn six Caucasian young adult male faces. All images are cropped to exclude hair and other external features, and participants complete a three-alternative forced choice (3AFC) test (see methods for more details). The CFMT has very good psychometric properties (Duchaine & Nakayama, 2006; Bowles et al., 2009; Wilmer et al., 2012), and is widely used to assess individuals for DP and estimate face recognition abilities in the general population (e.g., Bowles et al., 2009; Bate, Parris, Haslam, & Kay, 2010; Germine, Duchaine, & Nakayama, 2011; Wilmer et al., 2012). Recently, several child-focussed variations of the CFMT have been introduced. The CFMT-K (Dalrymple, Garrido, et al., 2014) uses children's faces, as they may be remembered more accurately than adult faces (Anastasi & Rhodes, 2005); and varies the number of faces to be learnt (four faces for younger children and the usual six faces for older children). In contrast, the CFMT-C (Croydon, Pimperton, Ewing, Duchaine, & Pellicano, 2014) retains the original adult face stimuli; but reduces the number of learning faces to five for all age groups, and limits the number of choices in the test phase (two faces rather than three as in the adult and CFMT-K versions of the test). Like the original CFMT, both tests show relatively good psychometric properties (Croydon et al., 2014; Dalrymple, Garrido, et al., 2014), although as yet the CFMT-K has only been tested in a relatively small sample of comparatively older children (142 children aged 7-12 yrs).

While these adapted versions of diagnostic tests are a good starting point, some theories of development in face recognition would suggest that it may be difficult, if not impossible, to detect impairments in face memory at a very young age. Some authors suggest that face-specific processing only emerges late in childhood (e.g., Diamond & Carey, 1977) or follows a protracted developmental trajectory (e.g., Carey & Diamond, 1994; de Heering, Rossion, & Maurer, 2012) – in other words, these theories suggest that face-specific processing is absent or reduced in children. Work with adults has found that, in many cases, people with DP may also show deficits or abnormal patterns of performance on tasks assessing face-specific mechanisms (e.g., Avidan, Tanzer, & Behrmann, 2011; Behrmann et al., 2005; De Gutis et al., 2012; Duchaine et al., 2007, Palermo et al., 2011). This has led some authors to suggest that people with DP process faces in a similar way as objects (i.e., focussing on individual features rather than the whole face) (Behrmann et al., 2005). If children also show reduced face-specific processing (de Heering et al., 2012), or are more likely to process faces like objects in memory tasks (Diamond & Carey, 1977), it would be very difficult to detect differences between typically developing children (who might show minimal face-specific processing in these tasks) and those with DP (whose levels of face-specific processing may be reduced even further).

Other researchers have argued that face-specific processes are in place and fully developed at a very young age, and that all subsequent development is a factor of general cognitive improvements (e.g., Crookes & McKone, 2009; McKone, Crookes, Jeffery, & Dilks, 2012; McKone, Crookes, & Kanwisher, 2009; Want, Pascalis, Coleman, & Blades, 2003). Indeed, an overwhelming amount of evidence has found that face-specific processes are present even in very young children (see McKone et al., 2009), and many studies that claim to support the late maturity hypothesis suffer from restriction of range effects that limit their findings (Crookes & McKone, 2009; McKone et al., 2012). This suggests that it should be theoretically possible to detect face recognition impairments at a young age. However, children have shorter attention spans and more limited memory compared to adults (Anderson, 2002; Schneider & Presley, 1997), which could result in floor effects for younger age ranges if tests are not carefully designed or adapted to take account of these factors (Crookes & McKone, 2009). In line with this, typically developing young children (5-6 yrs) perform relatively poorly on the CFMT-K and CFMT-C, and show relatively high variability (Croydon et al., 2014; Dalrymple, Garrido, et al., 2014), making it difficult to apply the standard diagnostic criteria of two standard deviations below the mean (as this is generally at or below chance levels of performance). This may explain the very small number of child DP cases reported in the literature (see Dalrymple et al., 2012; Dalrymple, Garrido, et al., 2014; Wilson et al., 2010); further, it is possible that these cases represent atypical or particularly severe impairments of perceptual processing (e.g., seven out of 10 cases reported in Dalrymple et al., 2012, also showed object recognition deficits). These factors suggest that the typical method of diagnosing DP in adults – based primarily around face memory – may not be sufficient to detect many cases of DP in childhood.

An alternative approach to screening for DP in children is to examine face perception. In general, impairments to face perception are generally not considered necessary to meet the criteria for prosopagnosia in adults. However, the developmental trajectory of face matching may make it a more appropriate method for DP screening in children: A recent study found that face memory undergoes a prolonged period of development, independent of object memory abilities; whereas face matching is mature relatively early, and subsequent improvement is likely a consequence of general cognitive development (Weigelt et al., 2014). In other words, while face memory may develop slowly (as suggested by Diamond & Carey, 1977; de Heering et al., 2012), leading to floor effects and potential difficulties detecting impairments; face matching may be mature at a relatively early age (as suggested by Crookes

& McKone, 2009; McKone et al., 2012; Want et al., 2003). As such, face perception problems might be a more reliable indicator of DP in childhood than face memory problems. In support of this, Dalrymple, Garrido, et al. (2014) recently reported eight cases of DP in children (5-12 yrs old) and found that all cases showed significant impairments in tests of face perception. In contrast, only six out of 16 adults with DP showed similar impairments of face perception.

Currently there is only one standardised test of face perception that has been designed for children, the Dartmouth Face Perception Test (DFPT, Dalrymple, Garrido, et al., 2014). The DFPT involves children choosing which of three morphed faces looks "most like" a target face. The test has some theoretical strengths (e.g., presenting faces from different angles to lessen focus on individual features; cropping faces to reduce reliance on external cues) and relatively good levels of performance in the age range tested (7-12 years). However, the test also has several limitations: firstly, there is no published data for younger children (4-6 years), and pilot testing in our lab revealed that younger children (4-5 years old) had difficulty understanding the concept of "most like" as opposed to "same". Therefore, it is unlikely that the DFPT would be effective and appropriate for identifying atypical face processing in this age range. Secondly, there is no object test available using a similar format, possibly due to the use of morphed images: while it is technically possible to morph images of some objects (e.g., those used by Weigelt et al., 2014), photographic stimuli can end up looking quite unusual or unrealistic.

The lack of object memory and perception tasks is a significant problem in screening for and diagnosing DP in children: it is vitally important to develop tests of object processing that are well-matched to face processing tasks in regards to task demands and difficulty, but to date no well-matched object processing tests have been published that are suitable for children. Development of these tests is necessary for both theoretical reasons (i.e., understanding the development of face and object processing, Want et al., 2003; characterising the extent and nature of DP) and diagnostic reasons: without reliable, well-validated tests, relatively well-matched in format and performance level, it is difficult to determine whether poor performance in face processing tasks reflects a specific (or exaggerated) deficit for face processing, or a more general impairment of perceptual processing or memory. Furthermore, when working with children it is important to include control tests to rule out potential confounding factors such as difficulty understanding the task, problems with concentration, or more general memory or perceptual difficulties (Want et al., 2003). Ideally, this control test would use stimuli well-matched to faces in several ways: relatively familiar and interesting to children (preferably children of both sexes); sharing a first-order configuration (i.e., same "parts" in roughly the same place relative to one another); and not likely to encourage identification from a single feature (Crookes & McKone, 2009). Currently, there are no child-focussed tests of object perception and memory that fit these criteria. Consequently, for this study we developed tests of object memory (Experiment 1) and matching (Experiment 2) using children's bicycles as stimuli. We adopted the format of the CFMT-K (Dalrymple, Garrido, et al., 2014) for the child face and bicycle memory tests. To overcome the limitations of the existing perceptual tests, our matching tests adopted a simple 3AFC simultaneous matching procedure.

The current study

The current study examined face and bicycle memory and matching in a large sample of primary school-aged children (5-11 years old). Experiment 1 examined face and object memory; while Experiment 2 examined face and object matching. We used this data to assess the psychometric properties of the newly developed tests and provide some provisional norms for different age ranges. We were also interested in whether any of the tests align well with other methods of identifying face recognition impairments. Currently, the most common method of identifying face recognition difficulties in children is through parental report and, in some cases, tests of personally familiar face recognition (Dalrymple et al., 2012; Dalrymple, Fletcher, et al., 2014; Dalrymple, Garrido, et al., 2014; Schmalzl et al., 2008; Wilson et al., 2010). However, it is unclear whether these are effective methods – no studies to date have examined the link between parental report and child face recognition abilities, or the link between familiar face recognition and novel face recognition in children. As such, Experiment 2 also examined the relationship between different methods of identifying face recognition difficulties: for example, whether parental ratings of face recognition ability and children's identification of familiar faces aligned with more objective, standardised tests of face processing ability.

Finally, we used the data to provide an estimate of the prevalence of DP in primary-school children. Previous studies estimating the prevalence of DP have focussed on the adult population, using standardised interviews or tests administered to a broad group of participants unselected for face recognition ability (Bowles et al., 2009; Kennerknecht et al., 2006; Kennerknecht et al., 2008). Similarly, our sample of primary school children was unscreened and unselected for face recognition abilities, making it suitable to generate a preliminary estimate of DP in children. However, previous studies have focussed primarily on face processing, and as such, it is unclear whether the individuals in these samples showed more generalised difficulties with object recognition in addition to face recognition. Given the inclusion of matched object tests, we were able to offer a prevalence estimate for both general recognition difficulties (incorporating both face and object recognition difficulties) and face-specific difficulties.

Experiment 1: Face and Object Memory

Methods

Participants. 277 children in UK primary school Years 1-6 (aged 5-11-years) completed the face and bike memory tasks. An additional 38 students from Year 1 (aged 5.5-6.5 years) were tested but their data was unable to be analysed due to errors in data collection. Parental consent was obtained prior to children participating in the study.

The data were screened prior to inclusion in the analysis, based on background information and behaviour during the testing session. Prior to testing, parents completed consent forms which requested details about medical conditions (e.g., visual problems or developmental disorders). Participants who had problems with their vision (either in the past or at the time of testing) or who had a diagnosed or suspected developmental or neurological condition were excluded from analysis. Further, participants who were not attending to the tests, who did not speak sufficient English to follow the instructions, or who appeared to be pressing random keys throughout the testing also had their data excluded from analysis. This resulted in a total of 202 children in the analysis for the face memory task: 32 year one students (5.5-6.5 years old; M 6.0 y[1]; 15 female); 29 year two students (6.5-7.5 years old; M 6.76 y; 18 female); 37 year three students (7.5-8.5 years old; M 7.66 y; 21 female); 36 year four students (8.5-9.5 years old; M 8.68 y; 12 female); 34 year 5 students (9.5-10.5 years old; M 9.67 y, 17 female); and 34 year six students (10.5-11.5 years old; M 10.70 y; 16 female). Due to some missing data and some children only completing one of the two tests, a slightly different sample of 184 children was included in the analysis for bike memory: 32 year one students (5.5-6.5 years old; M 6.0 y; 14 female); 26 year two students (6.5-7.5 years old; M 6.61 y; 16 female); 37 year three students (7.5-8.5 years old; M 7.63 y; 19 female); 36 year four students (8.5-9.5 years old; M 8.68 y; 12 female); 31 year 5 students (9.5-10.5 years old; M 9.64 y, 15 female); and 22 year six students (10.5-11.5 years old; M 10.53 y; 12 female).

[1] In many cases, age was reported in whole years rather than exact years and months. As such, the mean ages reported here are likely to be lower than the actual mean age of children tested.

Of these, a subset of 43 children completed a familiar faces task, and 84 parents provided ratings of their child's face recognition abilities.

Stimuli and Measures.

Memory tasks. The memory tasks in this study were the CFMT-K (Dalrymple, Garrido, et al., 2014) and a test of memory for children's bicycles, exactly matched in format to the CFMT-K. For both stimulus categories, stimuli were greyscale, static, and of roughly equal size. The stimuli in the CFMT-K are faces of male Caucasian children, extracted from the Dartmouth Database of Children's Faces (Dalrymple, Gomez, & Duchaine, 2013). All faces display a neutral facial expression, and have been edited and cropped so that no facial hair, glasses, hair or ears (paraphernalia) could be seen. Bicycle stimuli were extracted from the webpage of a popular UK online store, converted to greyscale, and edited to remove any obvious brand or design elements (e.g., distinctive patterns or logos).

Both memory tests consisted of three stages: a learning stage; a test stage with novel images; and a test stage with novel images overlaid with visual noise. In the learning stage, participants were shown a single stimulus from three different viewpoints (for faces: facing left, facing front, facing right; for bikes: angled left, side view, angled right; see Figure 1). Each viewpoint was shown for three seconds. After viewing all three images, participants were presented with three test stimuli simultaneously onscreen (see Figure 1), and asked to choose which of three stimuli matched the one they just saw by pressing the 1, 2, and 3 keys on the keyboard. The test stimuli remained onscreen until a response was made. The target stimulus was always an identical image to one of the three learning images (i.e., the same side or front view of a face; the same angled or side view of a bike). This was repeated three times for each stimulus.

After the learning stage, there was a short revision period in which the learned faces were shown on the screen for 20 seconds. Following the revision, children completed the first test stage. In this stage, the format and procedure of the test images was identical to the learning stage, except that the images showed the faces from novel viewpoints or under novel lighting conditions, and the correct answer could be any of the faces presented in the learning phase. After the first test stage, children received another revision period and the final test stage. Once again, the format and procedure for each trial remained the same as in the learning and initial test stages, but the stimuli were presented from novel viewpoints and overlaid by visual noise.

Younger children (school years 1-3) completed a short version of both tasks – this involved learning four faces and bikes. The short version had a total of 48 trials (12 trials in the learning stage; 20 in the test stage with novel viewpoints; 16 in the test phase with noise overlaid). Older children (school years 4-6) completed the full version of each task, which involved learning six faces and bikes. The full version had a total of 72 trials (18 trials in the learning stage; 30 in the test stage with novel viewpoints; 24 in the test phase with noise overlaid). Scores (overall and for each individual stage) were converted to percentages to facilitate comparisons across the two versions of the test.

Familiar faces test. A subset of participants who completed the matching or memory tasks also completed a familiar faces test. This test used greyscale photographs of staff members from the participants' school, cropped to exclude hair and ears. Photographs were provided by the staff or school. For each version of the test (i.e., for each school), faces of 10 staff members were chosen (with the exception of one school that only provided seven photographs). These were presented alongside 10 distractor faces, which were matched as closely as possible in age, sex, and race to the staff faces. The photographs were presented onscreen in a random order, and participants were asked if the person was familiar, and if so, could they name the person. Stimuli remained onscreen until a response was made, and responses were recorded by the researchers. Scores for correct identification of familiar faces and rejection of unfamiliar faces were converted to *d'* (with loglinear correction for extreme values; Stanislaw & Todorov, 1999), a bias-free measure of sensitivity (McMillan & Creelman, 2005).

Parental report rating. Prior to testing, children were provided with an information and consent form for parents to complete and sign. Part of this form asked parents to rate their children's face recognition skills on a scale of 1 (Poor) to 5 (Excellent). To assist with the rating, the following anchor points were provided: 1: Struggles to recognise people when seen out of context or external features such as hairstyle or glasses are changed; 3: Occasionally struggles to recognise people when seen out of context or external features are changed; 5: Always recognises faces, even when a person has only been seen once before or has dramatically changed in appearance.

Procedure. The majority of children were tested at their primary school. A small minority of children who took part in the memory tasks completed the tasks in a lab as part of a larger study. Prior to testing, information and consent forms were distributed by the school. These forms (including parental report rating) were collected on the day of testing. The participants were brought into the research session in small groups and briefed about the tasks. The instructions to the tasks were presented to all participants before the tasks began and explained verbally by the researchers. Each child completed both the face and bike versions of the memory tasks and, in most cases, the familiar faces test. Order of the tasks was counterbalanced across participants. Tasks were completed either on a laptop or desktop computer, and participants sat approximately 80cm away from a computer screen where the tasks were presented. While the participants completed the tasks, they were monitored closely by the researchers to ensure they were following the instructions and attending to the tasks.

Participants were rewarded with stickers for completion of each test. At all times during testing, the ratio of researcher to participant was a minimum of 1:3. Ethical approval was granted by the School Research Ethics Committee at Bournemouth University.

Results

The initial aim of this study was to examine whether currently existing memory tasks are appropriate for screening for, and potentially diagnosis of, developmental prosopagnosia in children. As such, our first step was to examine the general pattern of performance across the face and bicycle memory tasks, and to assess their psychometric properties – specifically, we were interested in the developmental trajectory of face and object memory in childhood; whether any of the tasks showed floor or ceiling effects in any age group; the reliability of the tasks; and the relationship between face memory and other measures of face recognition ability. Finally, we used the dataset to derive a memory-based estimate the prevalence of DP in children at various ages.

Throughout the results, children were split into different age groups and analysis was carried out based on their school year (1-6).

Overall performance. Descriptive statistics for the face and bicycle versions of the memory tasks, including mean, standard deviation, and summary information on distribution, can be found in Table 1. Frequency distribution graphs for each year group are presented in the Supplemental material. The means for the four item version of CFMT-K in this study are quite similar to those found by Dalrymple, Garrido et al. (2014) in age groups where the tasks overlap: for example, in their sample 7 and 8 year olds had an average score of 59.0% and 70.2% respectively; in the current sample Year 2 and 3 students (roughly equivalent in age range) average scores were 64.9% and 72.3%. Unsurprisingly, overall performance for both face and bicycle tasks dropped slightly when the six-item test was introduced (Year 4, around 9 years old). Performance in the six item version was somewhat lower than previous work:

the sample of 10 and 11 year olds tested by Dalrymple, Garrido, et al. (2014) had an average score of 84.7% and 78.4%, compared to 64.7% and 69.9% in the current sample of Year 5 and 6 students. This may be an artefact of testing environment (classroom compared to lab) – while every attempt was made to keep the classroom testing quiet and controlled, the simple fact of having multiple children in the room at once, and the fact that the rooms were generally classrooms, school libraries, or computer rooms (which generally contain more visual clutter than a typical lab) meant the environment could have been more distracting for some children. Alternatively, the discrepancy may reflect the difference between a relatively unselected school sample and a self (or parent) selected lab sample motivated enough to visit a university lab.[2]

Despite these relatively low scores, one-sample *t*-tests (Bonferroni corrected for multiple comparisons) conducted on the overall scores for the memory tests confirmed that participants across all age ranges were performing above chance levels (33.3%) and below ceiling (100%). The *t* tests revealed no floor or ceiling effects in any age group, all ps < .0005, suggesting that our results do not suffer from a restriction of range problem.

One of the main aims of this study was to create an object memory test that was matched in difficulty to the face memory test, since it can be very difficult to compare performance across different object categories without matching tests for difficulty in at least one age group (Weigelt et al., 2014). To determine whether the tests were actually well-matched for difficulty, we compared performance on the face and bicycle tests for the Year 3

[2] Certainly, children who completed the tests in our lab scored somewhat higher than their school-based counterparts: there was an average difference of 6% for Year 1-3 students, and 11.37% for Year 4-6 students. However, the lab sample was relatively small (N < 10 per age group), and therefore statistical comparisons between the groups lacked power to detect differences.

and Year 4 students. These age groups were selected based on Weigelt et al.'s rationale that matching performance on a middle age group is most likely to avoid spurious age effects (this also had the advantage of matching performance in both versions of the tests). Paired *t* tests revealed that overall accuracy between the bikes and faces did not differ in either group: Year 3: $t(29) = 1.78$, $p = .085$, $d = 0.32$; Year 4: $t(30) = 1.89$, $p = .068$, $d = 0.34$.

TABLE 1 ABOUT HERE

Relationship between age and face and object memory. Having established that the face and bicycle tasks were matched for difficulty and did not suffer from floor and ceiling effects, we carried out a mixed ANOVA with age group (Year 1, 2, 3, 4, 5, 6) as a between subjects factor and stimulus (face, bicycle) as a within-subjects factor, to examine whether the pattern of results for faces and bicycles differed. The ANOVA revealed a significant main effect of stimulus, $F(1, 144) = 20.99$, $p < .0005$, $\eta^2 = .13$, reflecting the fact that average scores on the bicycle memory task, $M = 73.62$, $SD = 16.03$, were significantly higher than scores on the face memory task, $M = 66.45$, $SD = 17.63$. There was also a significant main effect of age group, $F(5, 144) = 4.81$, $p < .0005$, $\eta^2 = .14$, reflecting a gradual increase in scores between Years 1 and 3 and Years 4 and 6, although the majority of pairwise comparisons between age groups were not significant after correction for multiple comparisons. Notably, there was no interaction between age group and stimulus, $F(5, 144) = 0.73$, $p = .732$, $\eta^2 = .02$, suggesting that the effect of age was similar for both face and bicycle memory.

Reliability. Reliability of the tasks was assessed in two ways: first, we examined correlations between the two test stages of each format of the task; secondly, we calculated the internal consistency of the tasks using Cronbach's alpha. As participant numbers were not sufficient to perform many of the analyses separately for each year group, data for younger children (years 1-3) and older children (years 4-6) were collapsed.

Correlation between test stages. This method of assessing reliability has been used frequently in other tasks that have adopted the CFMT format (e.g. the original CFMT, Duchaine & Nakayama, 2006; Bowles et al., 2009; the CFMT-C, Croydon et al., 2014; the Cambridge Cars Memory Test or CCMT, Dennett et al., 2012), and examines whether the two test phases measure distinct or overlapping processes (i.e., does adding visual noise change the perceptual, mnemonic, or decision process). Correlations were calculated separately for each version of the tasks (short and full), collapsed across age groups.

Performance in the two test phases of the face memory task was significantly correlated in both the short and long versions of the task, short: $r(86) = .76$, $p < .0005$; long: $r(89) = .78$, $p < .0005$. These relatively strong correlations are similar in size to those reported for the adult CFMT ($r = .74$, Duchaine & Nakayama, 2006; $r = .73$, Bowles et al., 2009), and higher than that reported for the CFMT-C ($r = .54$; Croydon et al., 2014).

Likewise, performance in the two test phases of the bicycle memory task was significantly correlated in both the short and long versions of the test, short: $r(83) = .71$, $p < .0005$; long: $r(72) = .69$, $p < .0005$.

Internal reliability. Both versions of the CFMT-K showed good internal consistency, short: Cronbach's $\alpha = 0.91$; long: Cronbach's $\alpha = 0.92$. These results are comparable to previously reported results for the CFMT-K ($\alpha = 0.89$ for both versions; Dalrymple, Garrido, et al., 2014), and the adult version of the CFMT ($\alpha = 0.89$; Bowles et al., 2009). The bicycle memory tests has similarly high internal consistency, short: Cronbach's $\alpha = 0.91$; long: Cronbach's $\alpha = 0.88$. For all scales, item analysis revealed no individual items where removal would improve the reliability of the scale. As such, all four versions of the memory tests meet the basic standards for reliability of clinical tests ($\alpha > 0.85$) as laid down by Aiken (2003).

Relationship between face memory and other measures.

Measures of face recognition. Descriptive statistics for the familiar faces test and parental ratings of face recognition abilities are shown in Table 2. It is notable that no parents reported poor face recognition skills in their children – all ratings fell between 3 (average) and 5 (excellent). Performance on the short and long versions of the CFMT-K did not correlate with parental report of face recognition skills, short: $r(40) = .166$, $p = .312$; long: $r(30) = .00$, $p = .998$. Given the small number of participants in these samples and the extremely limited range of the parental report scale, it is possible that these correlations underestimate the true relationship between parental report and face memory. Due to the very small number of participants who completed both the memory tasks and familiar faces test, we were unable to examine the relationship between face memory and familiar face recognition in this sample.

Face and bicycle memory. Both the short and long versions of the memory tasks showed significant correlations between face and bike memory scores, short: $r(77) = .46$, $p < .0005$; long: $r(73) = .34$, $p = .004$. Although these correlations are significant, they are relatively weak, similar to the correlations found between object and face memory tests in the adult population (e.g., CFMT and CCMT $r = .37$, Dennett et al., 2012).

TABLE 2 ABOUT HERE

Prevalence of face memory impairments in children. The standard diagnostic criterion for prosopagnosia in the adult population is performance >2 *SD*s below the mean for a matched control sample (Bowles et al., 2009), although some research groups use a less conservative criterion and include individuals who show performance > 1.7 *SD*s below the mean (e.g., DeGutis et al., 2012). Based on these criteria, the cut-off for suspected impairment for each age group (percentage correct for -2 *SD*s and -1.7 *SD*s) is presented in Table 3. Despite all tests showing relatively good psychometric properties, the large amounts of variability in the younger age groups combined with lower overall scores, resulted in the more conservative cut-off being below chance levels of performance (33.33%) for some age groups – specifically, Years 1, 2, and 5 (see Table 3 for z-scores for chance performance in each age group). For Year 1 children, even the less conservative cut-off fell below chance. Similar results were obtained when estimating cut-off scores for impairment using alternate techniques such as the percentile rank method (Crawford, Garthwaite, & Slick, 2009; see Supplemental material for further details). Even children with severe face processing deficits may be expected to perform at or slightly higher than chance levels – consequently, Year 1 students were excluded from the prevalence estimates. The remainder of this section focusses only on the remaining groups of children (n = 170). Year 2 and 5 students who scored more than 1.7 *SD*s below the mean were included in the prevalence estimates, but their results should be interpreted with caution.

TABLE 3 ABOUT HERE

Table 4 provides details of the individual cases that fell below the cut-off scores for either of the face processing tasks. These are represented graphically in Figure 3, and in the frequency distribution graphs included in the Supplemental material. Only one individual (a child in Year 6) met the strict criterion for face recognition impairment (> 2 *SD*s below the mean); another five children met the less conservative criterion of > 1.7 *SD*s below the mean (Year 2: 1 child; Year 3: 2 children; Year 5: 1 child; Year 6: 1 child). This leads to a prevalence estimate of between 0.59% and 3.53%, based purely on face memory ability.

TABLE 4 ABOUT HERE

Face-specific memory deficits. As mentioned in the introduction, a well-matched object recognition test can act as an effective control measure, allowing researchers to identify children who have specific or disproportionate difficulty with one class of objects. To identify how many of the above cases show a disproportionate difficulty with faces, we examined their scores on the bicycle tests. Results are shown in Table 4. All but one of the six cases that showed moderate to severe impairment in the face memory task performed in the normal range on the bike memory task (< 1.7 *SD*s below the mean; represented as red circles in Figure 3), while one case (W_13, a child in Year 3) showed a severe problem with bicycle memory in addition to face memory. In other words, only one of the cases could be considered to have a more general object recognition deficit; or showed difficulties with attention, task instructions, or other general performance factors, relative to their age group. This leads to a prevalence estimate of between 0.59% and 2.94% for *face-specific* memory deficits.

FIGURE 3 ABOUT HERE

Discussion

The face and bicycle memory tasks show appropriate levels of difficulty for examining memory in children, at least in a typical population. They show no evidence of floor or ceiling effects in children from 5-11 years of age, and are well-matched in difficulty in the middle of this age range (7-8 years), and show high reliability. Both tasks appear to show a similar pattern of results across the age range tested, suggesting that face and bicycle memory show a similar developmental trajectory (Crookes & McKone, 2009; cf Weigelt et al., 2014). The CFMT-K does not correlate with parental report of face recognition abilities, and shows a modest correlation with bicycle memory.

Unfortunately, though, the face memory tests show limited effectiveness for identifying cases of developmental prosopagnosia in young children: the standard diagnostic criteria of > 2 *SD* below the mean falls below chance levels for children below 7.5 years old. These figures are in agreement with Dalrymple, Garrido, et al. (2014), who found similar floor effects for the CFMT-K in children younger than 8 years of age. We also found surprisingly poor performance in Year 5 children (9.5-10.5 years), which suggests that future screening studies should consider using the easier, four-item version of the test in this age group. Our results also suggest that researchers aiming to examine face recognition ability in children should use caution when interpreting low scores for these age ranges: while an individual child may not score in the “impaired” range for their age group, a score within the “normal” range may not rule out the presence of face recognition difficulties. In these cases, researchers may wish to adopt a less stringent criterion for impairment (e.g., > 1.7 *SD*s below the mean or the 5th percentile for the age group) for screening, and follow up potential cases with alternative testing methods (see General Discussion).

Based on the remaining sample, our results suggest that between 0.6% and 3.5% of children may show difficulties with face recognition. While this figure is broadly in line with previous prevalence estimates in the adult population (e.g., Bowles et al., 2009; Kennerknecht et al., 2006), the exclusion of the youngest age group and resulting low sample size suggest that it may not reflect the true proportion of children who show difficulties with face recognition. Face perception tests may be a more appropriate screening tool in this population for several reasons: firstly, the face perception system may be mature earlier than the face memory system (Weigelt et al., 2014); secondly, matching tests have less cognitive demands than memory tests, as such, they and may avoid the kind of floor effects found in our younger age groups. Consequently, Experiment 2 examined performance and prevalence rates based on face and bicycle matching tasks.

Experiment 2: Face and Object Matching

Methods

Participants. 547 children in UK school years 1-6 (aged 5-11-years) took part in this study. As for the memory tasks, parental consent and background information was obtained prior to children participating in the study. The data were screened to exclude participants who had problems with their vision and those with a diagnosed or suspected developmental or neurological condition. Participants who were not attending to the tests[3], who did not speak sufficient English to follow the instructions, or who appeared to be pressing random keys throughout the testing also had their data excluded from analysis. This resulted in a total of 413 children included in the analysis for the matching tasks: 83 year one students (5.5-6.5 years old; M 5.94 y; 44 female); 83 year two students (6.5-7.5 years old; M 6.80 y; 37 female); 61 year three students (7.5-8.5 years old; M 7.76 y; 30 female); 76 year four students (8.5-9.5 years old; M 8.87 y; 41 female); 52 year 5 students (9.5-10.5 years old; M 9.8 y, 21 female); and 58 year six students (10.5-11.5 years old; M 10.73 y; 34 female). Of these, 378 children completed a familiar faces task, and 402 parents provided ratings of their child's face recognition abilities.

Stimuli and Measures. The face and bike matching tasks were exactly matched in format, and used stimuli extracted from the memory tasks. Each test consisted of 30 trials of a 3AFC simultaneous matching task. A target stimulus was shown at the top of the screen, along with three test stimuli at the bottom of the screen. The target and test stimuli differed in viewpoint and/or lighting conditions, to prevent simple image matching. The target stimulus was always the same viewpoint (frontal for faces, side view for bicycles), and the three test stimuli always shared the same viewpoint and lighting conditions (e.g., all test faces in a single trial were shown facing rightwards). There were six target stimuli, each displayed five times per test. On each trial, participants were asked to choose which of the test stimuli was the same identity as the target stimulus, and to respond using the 1, 2, and 3 keys on the keyboard. The stimuli remained onscreen until a response was made.

[3] The decision to exclude based on attention was made either based on researchers' observations at time of testing, or if a child responded faster than 300 msec to 10 or more trials on either the face or bike matching test.

Trials on which the reaction time was less than 300msec were excluded from analysis, on the basis that the child was not paying attention (or looking at all the stimuli) on that trial. This resulted in the removal of less than one trial per child on average (faces: M = 0.28 trials, SD = 1.06; bicycles: M = 0.30 trials, SD = 1.19). If 10 or more trials were excluded in this manner, the whole participants' data was removed from the analysis, under the assumption that they were not paying attention for a significant portion of the test. Scores were converted to percentages based on the final number of trials included in analysis.

Familiar faces tests and parental ratings were identical to Experiment 1.

Procedure. The procedure was identical to Experiment 1: all children completed both the face and bicycle versions of the matching tests, along with the familiar faces test (order of tests was counterbalanced). Parental ratings were provided on consent forms.

Results

Results for the matching tasks mirrored those for the memory tasks. Since a higher number of participants also completed the familiar faces tests and parental rating scale, we were also able to examine the relationship between these measures and the matching task, both in the general sample and for individual cases who showed difficulties with face recognition.

Overall performance. Descriptive statistics for the face and bicycle versions of the memory tasks, including mean, standard deviation, and summary information on distribution, can be found in Table 5. Frequency distribution graphs for each year group are presented in the Supplemental material. Unlike the memory tasks, all participants completed the same version of the matching tests.

Due to significant departures from normality, scores on matching tests were trimmed and transformed prior to analysis (with the exception of the reliability analyses). Participants who scored less than 25% correct on either test (five children in total) were excluded from the analysis on two grounds: firstly, these results were outliers for all age groups; secondly, these participants were performing distinctly below chance levels (33%), which may indicate a fundamental misunderstanding of the task requirements. Subsequently, the data underwent a square root transformation (including reflection). Descriptive statistics and summary information on distributions of the transformed data are available in the Supplemental material. All numerical data reported in the results section has been retransformed to the original scale (percentage correct).

One-sample *t*-tests (Bonferroni corrected for multiple comparisons, corrected $p < .004$) were conducted on the overall scores for the matching tests to confirm that participants across all age ranges were performing above chance levels (33.3%) and below ceiling (100%). As in the memory tests, the *t* tests revealed no floor or ceiling effects in any age group, all ps $< .0005$, suggesting that our matching results do not suffer from a restriction of range problem.

As in the memory tasks, we chose to compare the difficulty level of the tests for the middle of our age group (Years 3 and 4 students). Paired *t* tests revealed that overall accuracy between the bike and face matching tasks did not differ in either group: Year 3: $t(60) = 1.00$, $p = .323$, $d = 0.18$; Year 4: $t(76) = 1.84$, $p = .069$, $d = 0.21$.

TABLE 5 ABOUT HERE

Relationship between age and face and object matching. As in the memory tasks, we carried out an ANOVA and separate regressions to examine how performance on the face and object tests compared, and how it changed across the age range tested.

A mixed ANOVA with age group (Year 1, 2, 3, 4, 5, 6) as a between subjects factor and stimulus (face, bicycle) as a within-subjects factor revealed a significant main effect of stimulus, $F(1, 402) = 4.70$, $p = .031$, $\eta^2 = .01$, reflecting the fact that, on average, participants were significantly better at matching faces (*M* = 82.53, 95% *CI*: 81.12-83.88[4]) than bicycles (*M* = 80.61, 95% *CI*: 79.12-82.06). There was also a main effect of age, $F(5, 402) = 48.68$, $p < .0005$, $\eta^2 = .38$, reflecting a gradual increase in performance for older participants. Pairwise comparisons (Bonferroni-corrected) showed that in general, younger children performed worse than older children, although frequently the pairwise comparisons for adjacent age groups were not significant (e.g., Year 2 students performed significantly worse than Year 4, 5, and 6 students, *p*s < .0005, but not Year 3 students, $p = 1$). As in the memory tasks, there was no significant interaction between stimulus and age group, $F(5, 402) = .94$, $p = .455$, $\eta^2 = .01$, suggesting that the age-related gains in face matching performance were driven by similar processes to those underlying bicycle matching.

To examine the effect of age on matching performance, we carried out separate regression analyses on the face and bicycle versions of the tasks. In the face matching task, age significantly predicted scores, $F(1, 406) = 138.54$, $p < .0005$, accounting for 25.4% of the variance in performance. Each additional school year resulted in an average 5.60% increase in face matching scores.

The bicycle task showed a similar pattern of results: age significantly predicted bicycle matching scores, $F(1, 408) = 162.84$, $p < .0005$, accounting for 28.6% of the variance in performance. Similarly to the face matching task, each additional school year resulted in an average 6.14% increase in scores.

Reliability. Both the face and bike tasks showed high internal validity, faces: Cronbach's α = .87; bikes: Cronbach's α = .86. For both scales, item analysis revealed no individual items where removal would noticeably improve the reliability of the scale.

[4] Since transformations alter the relative distance between datapoints, 95% *CI*s are reported rather than *SD*s to provide a more accurate representation of the distribution of data after retransformation.

Relationship between face matching and other measures. A summary of performance on the familiar face recognition test and parental ratings of face recognition for each age group are shown in Table 6. The relationship between face matching scores and each measure is also shown for each age group.

Familiar faces test. Scores on the face matching task were significantly, albeit weakly, correlated with performance on the familiar faces test, $r(374) = .20$, $p < .0005$. It is unlikely that this correlation was artificially limited by restriction of range, as even the youngest children (who also showed the lowest recognition *d'*) performed significantly better than chance ($d' = 0$), $t(82) = 9.88$, $p < .0005$, $d = 1.08$; and the year 5 children (who showed the best recognition rates) did not perform at ceiling levels ($d' = 3.38$), $t(51) = -12.38$, $p < .0005$, $d = 1.71$.

Scores on the bicycle matching test also correlated with the familiar faces test to a similar degree, $r(374) = .19$, $p < .0005$, which suggests that the link between face perception and performance on the familiar faces test was not face-specific. Partial correlations showed that the relationship between familiar face recognition and both matching tasks disappeared when age was controlled for: face matching and familiar faces $r = .02$, $p = .689$; bicycle matching and familiar faces: $r = .02$; $p = .768$. This suggests that the (weak) relationship between familiar face recognition and both matching tasks is a by-product of general cognitive development – it is likely that older children were simply better at concentrating, focusing their visual attention, and following instructions; and this lead to a general increase in performance across multiple tasks.

Parental report. Similar to the face memory test, parental report of face recognition skills did not correlate with performance on the face matching task, $r(397) = .048$, $p = .340$. Once again, the vast majority of parents (400/402 valid datapoints) rated their child as having average (3) to excellent (5) face recognition abilities. No parents rated their child as having poor face recognition ability, and only two children (one in Year 1; one in Year 6) received a parental rating of 2. Neither of these children demonstrated poor objective face processing skills (Year 1: face matching z = 1.10, familiar faces z = 0.36; Year 6: face matching z = -0.16, familiar faces z = 0.82).

Face and bicycle matching. Correlations between face and bicycle matching for each individual age group are shown in Table 5. Overall, scores on the face and bicycle matching tasks correlated significantly, $r(408) = .46$, $p < .0005$, even after age was partialled out, $r = .26$, $p < .0005$. The correlation between matching tests was a similar magnitude to the correlations in the memory tests: a Fisher r-to-z transformation confirmed that the zero-order correlation was not significantly different to that found for either version of the memory test, *p*s > .15.

TABLE 6 ABOUT HERE

Prevalence of face matching deficits in children. The cut-off scores for suspected impairment for each age group (retransformed percentage correct for -2 *SD*s and -1.7 *SD*s, and z-scores for chance performance) are shown in Table 7[5]; cut-offs based on the percentile rank method are presented in the Supplemental material. Similar to the memory tests, the cut-off scores for the youngest age group (year 1: 5.5-6.5 years) were below chance levels of performance (33.33%). As such, these participants were excluded from the prevalence

[5] While an individual child may be classified as above/below the quoted cut-offs from raw scores alone, due to the nature of transformed scores (which change the relative distance between scores) we would not recommend estimating the magnitude of difference from the mean (i.e., z-scores or similar) based on an individual's raw score. Researchers wishing to quantify results for individual cases should either apply appropriate data transformations or rely on the percentile rank method (Crawford et al., 2009; see Supplemental material for more details).

estimates, and the subsequent analyses focus only on Years 2 - 6 (n = 329). As in the memory tasks, the more conservative cut-off for Year 2 students fell below chance levels. However, as the less conservative cut-off was still above chance, children scoring more than 1.7 SDs below the mean were retained in the prevalence estimates, but interpreted with caution.

TABLE 7 ABOUT HERE

Table 8 provides details of the individual cases who fell below the cut-off scores in the matching task; Figure 4 plots these cases graphically. A total of six children met the strict criterion for face recognition impairment (> 2 *SD*s below the mean). These cases were spread fairly evenly across the age range tested (Year 2: 1 child; Year 4: 2 children; Year 5: 2 children; Year 6: 1 child), suggesting that difficulties with face recognition are not more common or more apparent within a particular age group. Another 14 children met the less conservative criterion, > 1.7 *SD*s below the mean (Year 2: 3 children; Year 3: 5 children; Year 4: 2 children; Year 5: 1 child; Year 6: 3 children). This leads to a prevalence estimate of between 1.82% and 6.08%, based purely on face matching ability.

TABLE 8 ABOUT HERE

Face-specific deficits. Five out of the six cases who showed severe impairment on the face matching task performed in the normal range on the bike matching task, and a single case showed co-occurring severe impairment in the bike matching task (z = -2.52). Ten of the 14 cases who showed moderate impairment in the face matching task also performed within the normal range on the bike matching task. In total, then, between 1.52% and 4.56% of children showed face-specific perceptual deficits. These are represented as red circles in Figure 4. The remaining five children (represented as green triangles in Figure 4) may show a more generalised perceptual deficit, or could have simply had trouble with the task – further testing with different measures would be necessary to establish why each individual performed poorly across multiple tests. As can be seen from Table 8 and Figure 4, the level of face recognition deficit did not have a straightforward relationship with bicycle recognition problems – while it is true that the most severe face recognition deficit was accompanied by the most severe object recognition deficit (P5_05), in general, the cases who showed more general object recognition problems did not show more severe perceptual problems than those who showed a face-specific deficit.

FIGURE 4 ABOUT HERE

Converging measures. The majority of child DP cases reported in the literature to date have been referred to researchers via parental report and, until recently (e.g., Dalrymple, Garrido, et al., 2014), familiar faces have formed an important part of the diagnostic battery. Similarly, in adults, both self-report and multiple measures of face recognition ability are commonly taken into account when diagnosing DP (e.g., Avidan et al., 2011; De Gutis et al., 2012; Palermo et al., 2011; Yardley et al., 2008). However, it is unclear whether these are reliable measures, particularly in the case of children. The lack of correlation between face matching and parental report, and the relatively small correlation between face matching and familiar face recognition, suggest that these measures are not reliable at predicting performance within the general population. However, they may be more reliable at identifying cases at the lower end of the face recognition spectrum.

As can be seen from Table 8, none of the cases identified as having face matching difficulties received a low rating (1 or 2) on the parental report scale, and only one of the cases identified as having face matching difficulties showed a co-occurring deficit in the familiar faces recognition task (P2_15). As such, it appears that parental report and familiar faces tests (at least those involving faces of teachers) do not reliably predict poor performance in standardised tests of face matching.[6]

Prevalence of face processing difficulties across memory and matching tasks. Combining the two different types of test (memory and matching), 1.40% of children (n = 7) showed severe face processing difficulties; 5.21% (n = 26) showed moderate to severe difficulties. The lower end of this estimate is generally in line with prevalence estimates in the adult population (2-2.9%, Bowles et al., 2009; 2.47%, Kennerknecht et al., 2006), but the less conservative estimate is somewhat higher. This may simply reflect variability in face recognition performance in children: the cut-off points for the strict criterion are close to or below chance in some instances (particularly in younger age groups), which may mean more cases of DP fall into the "moderate" impairment range (> 1.7 *SD*s below the mean). However, this may also be because this estimate conflates three groups of children who performed poorly on the face recognition tests: those who performed poorly due to lack of attention or problems understanding the requirements of the tasks; those that performed poorly due to general perceptual or mnemonic problems; and those who truly show a face-specific deficit. While previous DP prevalence studies have not specifically examined object recognition (i.e., they may have also conflated the latter two groups of individuals), it is less likely that poor performance in adults could be attributed to lack of attention or difficulty with task requirements, and this may explain the markedly higher prevalence estimate for children in the current study. When children with co-occurring object recognition deficits are excluded,

[6] Unfortunately, familiar face score and parental ratings were not available for the majority of the children who showed face memory difficulties, so we were unable to determine whether these measures were a more effective predictor of memory difficulties.

1.20% of children (n = 6) showed a severe, face-specific deficit; 4.01% of children (n = 20) showed a moderate to severe face-specific deficit.

Discussion

The face and bicycle matching tests are of an appropriate level of difficulty for examining perception in primary school children. As in the memory tests, they show no floor or ceiling effects, and are well-matched for difficulty in the middle of the age range tested (5-11 years). Both tests show a similar, linear developmental trajectory, further supporting the hypothesis that face and object matching improve with age as a result of general cognitive development, rather than face-specific mechanisms (Crookes & McKone, 2009; Want et al., 2003). Both the face and bicycle matching tests show high reliability, and they appear to draw on partially overlapping mechanisms (as indicated by a significant correlation). However, the moderate size of this correlation suggests that face and object perception rely to some extent on discrete mechanisms, perhaps reflecting the input of face-specific processes. Face and bicycle matching correlate very weakly with familiar face recognition, but this is most likely driven by the fact that all three tasks improve with age, rather than a reflection of the overlap between the tasks themselves. There is no apparent relationship between parental report and face matching skills, either in the general population or in individual children who performed poorly in the face recognition tasks. In sum, the face and bicycle matching tests show good psychometric properties, and avoid floor effects across most of the age-range tested – as such, they are appropriate tools to detect potential face recognition deficits in children. Due to the non-normal distribution of scores for most ages groups, we would suggest that researchers who wish to make single-case comparisons use the percentile rank method (Crawford et al., 2009) – cut-offs are provided in the Supplemental material, and the information needed to calculate a child's exact percentile is available from the authors on request (or can be calculated from the frequency distribution tables, also in the Supplemental material).

Based solely on the matching tests, the upper estimate of prevalence of face recognition difficulties in children is higher than previously reported in adults: over 4.5% showed a face-specific deficit; and if results on object tests are not taken into account, over 6% of children show a substantial difficulty with face recognition. These figures remain relatively stable even when the data from the memory tests is included (resulting in a sample of 499 children): across the entire sample, over 4% of children showed moderate to severe face-specific recognition problems; if we include children who also show a more general problem with object and face recognition, this figure rises to over 5.2%.

General Discussion

Developmental prosopagnosia is a disorder of face recognition that is thought to affect around 2-3% of the adult population (Bowles et al., 2009; Kennerknecht et al., 2006). However, to date there have been no studies that have looked at the prevalence of DP in children. In this study we examined face and object recognition in primary school children (aged 5-11 years). Between 1.2% and 5.2% of our sample performed poorly in face recognition tasks, and around 1.2% to 4% of our sample showed deficits in face recognition in the presence of normal object processing – in other words, a face-specific deficit.

While our sample is not sufficiently large to provide a precise estimate of DP in childhood, these figures provide the first approximation of the percentage of children who might struggle with face recognition in general life. When object recognition is not taken into account (i.e., the 1.2-5.2% figure), the higher rate of face recognition difficulties children in general no doubt reflects the fact that some children simply did not understand or were not engaged with the tasks. While we attempted to exclude these cases (based on both observation and patterns of results), it is possible that some were still included in the final sample and led to inflated estimates of general recognition difficulties. As such, 5.2% should be considered a very high or upper-bound estimate of face recognition deficits in this age group. In contrast, the estimate of face-specific deficits in children (i.e., 1.2%-4%) is likely to be lower than the actual number of children who meet the criteria for DP: while there are a number of reports of face-selective deficits in DP (e.g., Dalrymple, Garrido, et al., 2014; Duchaine & Nakayama, 2005; Lee, Duchaine, Wilson, & Nakayama, 2010), many cases of DP in adults and children show co-occurring object recognition deficits (Duchaine et al., 2007; Dalrymple, Garrido, et al., 2014; Dalrymple et al., 2012; Behrmann et al., 2005). Consequently, 1.2%-4% could be considered a conservative or lower-bound estimate of face recognition difficulties in children.

Prevalence of DP in children and adults

The upper range of the prevalence estimates for children – are relatively high compared to those found in the adult population (2-2.9%). While some of this discrepancy could be accounted for by lack of engagement in the tasks, even the more conservative estimate (including only face-specific deficits) – is noticeably higher in the current study when compared to past research (Bowles et al., 2009; Kennerknecht et al., 2006; Kennerknecht et al., 2008). This is particularly notable given that other prevalence studies have not included matched object recognition tests – in other words, their estimates are likely to be less conservative and include both face-specific and more general object recognition deficits.

The difference between our results and previous research may be because only one of these studies (Bowles et al., 2009) used behavioural measures of face recognition ability. Other research on prevalence has relied on self-report questionnaires followed by semi-structured interviews (Kennerknecht et al., 2006, 2008), which could be subject to completion or response biases (especially in a highly competitive population such as medical students, who made up a large portion of the cohort in both studies) – indeed, the authors acknowledge that their estimates should be considered minimal prevalence (Kennerknecht et al., 2006).

Alternatively, it is possible that the current results overestimate the number of children with face recognition difficulties, potentially due to the use of matching tasks rather than memory tasks in the majority of the sample. Some of the children who performed poorly at face matching may have shown minimal or no impairments in face memory – perhaps through the use of compensatory strategies – and consequently, these children may not represent true cases of DP. However, cognitive models of face perception (e.g., Bruce & Young, 1986) suggest that intact face perception skills are necessary for successful face recognition – in the absence of unusual strategies, poor face perception should make it difficult to extract a viewpoint and expression invariant representation of the face to store for future encounters (or match to a previously seen face). On the other hand, it is possible for individuals with DP to show poor memory in the context of typical perceptual abilities (e.g., Bate et al., 2014; Bennetts et al., in press; Eimer, Gosling, & Duchaine, 2012). Based on this model, we would expect matching tests to be more conservative than traditional memory-based tests of face recognition. Furthermore, recent research has suggested that face perception mechanisms mature earlier than face memory mechanisms (Weigelt et al., 2014), which means that it may simply be easier to detect abnormal performance in matching tasks (i.e., those relying on relatively mature processes) compared to memory tasks that may still be developing. Finally, while the majority of cases reported in the literature rely on memory tasks in the diagnosis of DP, adult prevalence estimates have included cases that have been identified solely on the basis of perceptual deficits (Bowles et al., 2009) – as such, our inclusion criteria are reasonably comparable with previous work.

Finally, it is possible that more children than adults experience difficulties in face recognition. This does not imply that all of these children should be classified as having DP: some children may simply show a delayed developmental trajectory for face recognition abilities, and in time these individuals may "outgrow" these deficits and catch up to their cohort. Other children may develop effective compensatory strategies that allow them to identify faces with a relatively normal level of accuracy (e.g., focussing on individual face parts). Given that the levels of face recognition problems did not decrease with age in the current sample, these processes presumably occur later in development, possibly during adolescence. Longitudinal studies of individual cases may help to determine what proportion of children outgrow or bypass their difficulties with faces, and whether there are any characteristics that can distinguish between these cases and individuals who continue to struggle with face recognition in adulthood.

Screening for and identification of face recognition deficits in children

Given the relatively high number of children who show face recognition difficulties, and the potential negative consequences such as social difficulties and safety concerns, it is crucial to develop effective methods of identifying individual cases. Early detection of face recognition problems is important for our theoretical understanding of typical and atypical visual development, as well as the development and implementation of interventions that could improve face recognition skills (Bate & Bennetts, 2014) or help mitigate negative socio-emotional consequences.

To date, the vast majority of the cases of DP in children reported in the literature have been brought to researchers' attention via parental report, and validated through tests of familiar faces recognition (e.g., Jones & Tranel, 2001; McConachie, 1976; Schmalzl et al., 2008) or, more recently, tests of face memory and perception (e.g., Brundson et al., 2006;

Dalrymple, Garrido, et al., 2014; Wilson et al., 2010). However, our results suggest that parental ratings of a child's face recognition ability may not be a reliable method of screening for face recognition deficits, at least in this simplified form. In the overall sample that completed the matching tasks, parental ratings did not correlate with objective measures of face recognition. Even within the sample of children who performed poorly on the face matching tests, no parent rated their child as being poor at face recognition. This may be due to the nature of the question, which focussed on children's ability to recognise faces in different contexts, and with different external features. However, the face memory and matching tests used in this study did not ask children to identify faces with different hairstyles or in different contexts, and it is possible that the discrepancy between ratings and performance arose because the measures do not tap overlapping skills in children. However, this question was chosen because many individuals with DP report failures of face recognition under these circumstances (Duchaine, 2011), suggesting a relatively good relationship between these skills and standard face recognition tests in adults with face recognition difficulties.

An alternative explanation is that most parents do not have accurate insight into their child's face recognition abilities, which could mean that many cases of DP in children go unidentified. This is not entirely surprising – even when children have been objectively confirmed to have DP, many parents struggle to understand the experiences of a child with DP and can show doubt about their impairment (Dalrymple, Fletcher, et al., 2014). Furthermore, even adults show poor insight into their face recognition skills: Bowles et al. (2009) found minimal to no correlation between self-report of face recognition skills and either the CFMT or CFPT. This is not to say that parental- or self-report measures cannot be useful in the identification of face recognition difficulties: Kennerknecht et al. (2006) note that a subsample of the cases identified in their study also performed poorly on a variety of behavioural tests; and the recently developed PI20 (a self-report measure of face recognition difficulties; Shah, Gaule, Sowden Bird, & Cook, 2015) correlates well with performance on famous face identification tests and the adult version of the CFMT. However, neither of these tests is appropriate for children, and until a reliable parental report scale is developed, behavioural tests appear to be the most appropriate way of identifying face recognition deficits in children.

It may also be difficult to identify cases of DP based on familiar face recognition tests. As noted in the introduction, these tests can be laborious and impractical to develop: given each child has exposure to a different circle of familiar faces, screening large groups requires finding a set of appropriately familiar faces and developing norming data for each subset of the population (e.g., different schools/ages for large scale screening). Even under these circumstances, it is difficult to rule out the possibility that some children may simply have less exposure to the "familiar" faces than others. This could explain why we found a relatively weak correlation between familiar face recognition (in this case, recognition of school teachers) and performance on the face matching test: despite attempting to choose highly familiar teachers, different class groups (and even individuals within the same class) had different levels of exposure to teachers. This further supports the idea that familiar faces tests are not ideal for screening large samples, although carefully constructed tests may be useful for confirming face recognition deficits in children who are identified in the screening process.

The difficulties inherent with parental report and familiar faces tests confirm that objective, standard tests of face recognition are the most appropriate method to screen for and identify individual cases of face recognition difficulties in children. In this study, both the CFMT-K and the face matching test showed good psychometric properties, and our data suggest that either test would be appropriate for identification of potential face recognition difficulties in older children (>7.5 years old). It could be argued that a major disadvantage of using a matching task alone as a screening tool is that it may fail to identify individuals who show deficits of face memory, but not face perception – a pattern of results found in some adults with DP and referred to as prosopamnesia (Bowles et al., 2009). However, recent research on children with DP has found that the vast majority exhibit difficulties in both perceptual and mnemonic tasks (Dalrymple, Garrido, et al., 2014). Taken together with the comparative ease of testing younger children on matching tasks (i.e., simpler instructions, one version for all age groups), we propose that matching tasks are an appropriate screening tool for children. Nonetheless, future studies examining children's performance on both matching and memory tests, and validation of performance in the different tests in participants with known face recognition difficulties (e.g., children with DP or autism spectrum disorder) will be needed to determine whether one testing format is more effective than another for identifying face recognition deficits in children. Future work should also explore tests outside the forced choice format – while these tests provide a valuable measure of identification, it is difficult to discriminate between different aspects of the identification process such as recognition (knowing a face belongs to a certain person) and discrimination (knowing that a face does not belong to a particular person). These elements may contribute to different failures of identification (e.g., walking past a familiar person; mistaking a stranger for a familiar person), and developing tests that can discriminate between these processes (e.g., Laurence & Mondloch, 2016) could provide additional information about the cognitive characteristics of DP in children, and identify potential areas of focus for rehabilitation.

It should be noted that none of the tests that have been developed to assess children's face recognition abilities are able to reliably detect impairments in very young children (6 yrs and below). Dalrymple, Garrido, et al. (2014a) did not present data for the CFMT-K or DFPT for children younger than seven years old, but noted that the CFMT-K showed floor effects (i.e., > 2 *SD*s from the mean is below chance levels of performance) for the seven-year olds, which is consistent with our current findings of floor effects in year 1 and near floor effects in year 2 (5.5-7.5 years old). Similarly, data from the norming sample of the CFMT-C (Croydon et al., 2014) shows that five- and six-year olds also show a floor effect. Even the very simplified matching task used in the current study showed floor effects in year 1 and near floor effects in year 2 students (5.5-7.5 years old). These findings suggest that simple forced choice behavioural tests may not be the most appropriate method of identifying face recognition difficulties in this age group, and alternative methods may need to be developed. One such technique could be eye-tracking: adults with DP often show atypical patterns of eye movements when looking at faces (e.g., Barton, Radcliffe, Cherkasova, & Edelman, 2007; Schwarzer et al., 2007), and this has also been observed in one child case (Schmalzl et al., 2008). Similarly, other disorders that have well-known abnormalities with face processing, such as autism spectrum disorder and Williams syndrome, have also been found to have atypical gaze patterns when viewing faces (Pelphrey et al., 2002; Riby & Hancock, 2008). Given that eye-tracking tests can be much shorter and require less input from the participants than traditional behavioural experiments, they can be used to examine face recognition in much younger populations (e.g., Liu et al., 2011; Turati, Di Giorgio, Bardi, & Simion, 2010), and consequently these tests may be a more appropriate and reliable method of identifying face processing difficulties in very young children.

Finally, as discussed in the introduction, it is important to include a well-matched object recognition test when screening for face recognition difficulties. While this is not a common procedure for screening adults with DP, the use of well-developed and normed object tests is particularly important to identify children who are not engaging with the tasks, and to discriminate between children who show specific or disproportionate deficits with faces as opposed to objects. This study introduces two tests of object recognition which are matched in format to the face recognition tests, show high reliability, and are relatively well-matched in performance across the age range tested. Given the excellent psychometric properties of these tests, we suggest that they (or similar object recognition tests) be incorporated into screening procedures for DP in children.

Development of face recognition

While the main aim of this study was to examine the prevalence of face recognition difficulties in childhood, the data also allows us to make some observations about the development of face recognition in children with typical face recognition abilities. Performance on both the memory and matching tests showed improvement throughout childhood, as would be expected from past research (e.g., Crookes & McKone, 2009; Croydon et al., 2014, Johnston et al., 2011; Weigelt et al., 2014). Due to the shift from a four-item to six-item memory test, the pattern of improvement is less clear for memory tests (similar to the norms reported in Dalrymple, Garrido, et al., 2014), but there was a significant linear trend for face matching to develop with age. Notably, there was an almost identical trend for bicycle matching to improve with age, and there was no interaction between face and bike scores in either the matching or memory tests. This argues against the idea that face recognition follows a protracted period of development, independent of general object recognition (e.g., Carey & Diamond, 1994; de Heering et al., 2012). Instead, our results are consistent with the view that face recognition is mature at an early age, and developmental improvements reflect general gains in cognitive performance (e.g., memory, sustained attention, ability to follow task instructions) (Crookes & McKone, 2009; Want et al., 2003). While the general pattern of our data is also consistent with the view that faces and objects are processed in a similar manner in children (Diamond & Carey, 1977), the fact that face and bike matching and memory showed a similar, modest level of correlation across the age range tested (with the exception of year 6 students in the matching task), and this correlation is quite close to that reported for face and object processing in adults (Dennett et al., 2012), suggests that face and object processing rely to a large extent on separate processes, similar to face and object recognition in adulthood. This conclusion is also consistent with many studies that have found face-specific processing in young children (see McKone et al., 2009, 2012, for a review).

It is important to note that both the memory and matching tests showed a similar developmental trajectory for faces and bicycles. Recently Weigelt et al. (2014) suggested that face memory and matching may undergo separate developmental trajectories, with face matching maturing early, and showing a similar developmental trajectory to object recognition; and face memory maturing more gradually, and therefore showing a different developmental trajectory to object recognition. This was one of the bases for choosing to rely primarily on matching tasks to screen for face recognition difficulties. However, our results indicate that both face memory and face matching show a similar developmental trajectory to object memory and matching throughout childhood – in other words, the developmental trajectory does not differ depending on the task. It is unclear why our results differed from Weigelt et al.'s (2014), as the methodologies of the two studies differed in a number of ways – for example, Weigelt et al.'s (2014) memory test required children to remember 10 items from each category (as opposed to 4-6 in the current memory tests), and their test phase did not incorporate changes in image quality or viewpoint. Consequently, further work with tests designed to investigate these specific methodological differences may be necessary to determine the factors that influence the developmental trajectories of face and object processing.

Nonetheless, the present results suggest that both matching and memory tests (at least, those using the CFMT format) are theoretically appropriate methods to diagnose face recognition deficits in children, as long as the tests are designed and administered to avoid floor effects.

Conclusions

The current study found that a surprisingly high percentage of children show significantly poor performance in face recognition tasks. Given the relatively high prevalence of difficulties and the increasing public recognition of DP, we expect that an increasing number of child DP cases will be brought to researchers' attention over the coming years. Consequently, it is important to develop effective screening and diagnostic procedures for face recognition impairments; incorporating reliable, valid tests, such as those used in the current study. Further work is needed to develop more effective parental report measures and screening tests for young children. In combination, these measures will allow researchers to identify cases of DP at an early age, and potentially begin interventions that could help improve face recognition in affected children.

Supplemental Material

The Supplemental material can be found at the address

References

Anastasi, J. S., & Rhodes, M. G. (2005). An own-age bias in face recognition for children and older adults. *Psychonomic bulletin & review*, *12*(6), 1043-1047.

Anderson, P. (2002). Assessment and development of executive function (EF) during childhood. *Child Neuropsychology*, *8*(2), 71-82.

Aiken, L.R. (2003). *Psychological testing and assessment* (11^{th} ed.). Boston, MA: Allyn and Bacon.

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

Barton, J. J., Radcliffe, N., Cherkasova, M. V., & Edelman, J. A. (2007). Scan patterns during the processing of facial identity in prosopagnosia. *Experimental Brain Research*, *181*(2), 199-211.

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., & Cook, S. (2012). Covert recognition relies on affective valence in developmental prosopagnosia: Evidence from the skin conductance response. *Neuropsychology, 26*, 670-674.

Bate, S., Cook, S.J., Duchaine, B., Tree, J.J., Burns, E.J., & Hodgson, T.L. (2014). Intranasal inhalation of oxytocin improves face processing in developmental prosopagnosia. *Cortex, 50*, 55-63.

Bate, S., Haslam, C., Jansari, A., & Hodgson, T. (2009). Covert face recognition relies on affective valence in congenital prosopagnosia. *Cognitive Neuropsychology, 26*, 391-411.

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

Bennetts, R. J., Butcher, N., Lander, K., Udale, R., & Bate, S. (In press). Movement Cues Aid Face Recognition in Developmental Prosopagnosia. *Neuropsychology.*

Benton, A. L., Sivan, A. B., Hamsher, K., Varney, N. R.,& Spreen, O. (1983). *Contributions to Neuropsychological Assessment.* New York, NY: Oxford University Press.

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

Bruce, V., & Young, A. W. (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., & Diamond, R. (1994). Are faces perceived as configurations more by adults than by children? *Visual Cognition, 2(3),* 253–274.

Crawford, J. R., & Garthwaite, P. H. (2009). Percentiles please: The case for expressing neuropsychological test scores and accompanying confidence limits as percentile ranks. *The Clinical Neuropsychologist*, *23*(2), 193-204.

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.

Croydon, A., Pimperton, H., Ewing, L., Duchaine, B. C., & Pellicano, E. (2014). The Cambridge Face Memory Test for Children (CFMT-C): A new tool for measuring face recognition skills in childhood. *Neuropsychologia*, *62*, 60-67.

Dalrymple, K. A., Corrow, S., Yonas, A., & Duchaine, B. (2012). Developmental prosopagnosia in childhood. *Cognitive Neuropsychology*, *29*(5-6), 393-418.

Dalrymple, K. A., Fletcher, K., Corrow, S., das Nair, R., Barton, J. 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*(2), 144-150.

Dalrymple, K. A., Gomez, J., & Duchaine, B. (2013). The Dartmouth Database of Children's Faces: acquisition and validation of a new face stimulus set.

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.

Davis, J. M., McKone, E., Dennett, H., O'connor, K. B., O'Kearney, R., & Palermo, R. (2011). Individual differences in the ability to recognise facial identity are associated with social anxiety. *PloS one*, *6*(12), e28800.

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

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

Dennett, H. W., McKone, E., Tavashmi, R., Hall, A., Pidcock, M., Edwards, M., & Duchaine, B. (2012). The Cambridge Car Memory Test: A task matched in format to the Cambridge Face Memory Test, with norms, reliability, sex differences, dissociations from face memory, and expertise effects. *Behavior Research Methods*, *44*(2), 587-605.

Diamond, R., & Carey, S. (1977). Developmental changes in the representation of faces. *Journal of Experimental Child Psychology*, *23*(1), 1-22.

Duchaine, B. (2011). Developmental prosopagnosia. In A. Calder, G. Rhodes, J Haxby, & M Johnson (Eds.), *Oxford handbook of face perception*. UK: Oxford University Press.

Duchaine, B., Germine, L., & Nakayama, K. (2007). Family resemblance: Ten family members with prosopagnosia and within-class object agnosia. *Cognitive Neuropsychology*, 24, 419-430.

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 participants. *Neuropsychologia*, *44*, 576-585.

Duchaine, B., Yovel, G., & Nakayama, K. (2007). No global processing deficit in the Navon task in 14 developmental prosopagnosics. *Social cognitive and affective neuroscience*, *2*(2), 104-113.

Eimer, M., Gosling, A., & Duchaine, B. (2012). Electrophysiological markers of covert face recognition in developmental prosopagnosia. *Brain*, *135*(2), 542-554.

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

Johnston, P., Kaufman, J., Bajic, J., Sercombe, A., Michie, P., & Karayanidis, F. (2011). Facial emotion and identity processing development in 5- to 15-year-old children. *Frontiers In Psychology*, *2,* 26.

Jones, R. D., & Tranel, D. (2001). Severe developmental prosopagnosia in a child with superior intellect. *Journal of Clinical and Experimental Neuropsychology*, *23*(3), 265-273.

Kennerknecht, I., Grueter, T., Welling, B., Wentzek, S., Horst, J., Edwards, S., & Grueter, M. (2006). First report of prevalence of non-syndromic hereditary prosopagnosia (HPA). *American Journal of Medical Genetics Part A*, *140*(15), 1617-1622.

Kennerknecht, I., Ho, N. Y., & Wong, V. C. (2008). Prevalence of hereditary prosopagnosia (HPA) in Hong Kong Chinese population. *American Journal of Medical Genetics Part A*, *146*(22), 2863-2870.

Lee, Y., Duchaine, B., Wilson, H. R., & Nakayama, K. (2010). Three cases of developmental prosopagnosia from one family: Detailed neuropsychological and psychophysical investigation of face processing. *Cortex, 46(8),* 949-64.

Macmillan, N. A., & Creelman, C. D. (2005). *Detection Theory: A User's Guide (2nd ed.).* Mahwah, N.J.: Lawrence Erlbaum Associates.

Laurence, S., & Mondloch, C. J. (2016). That's my teacher! Children's ability to recognize personally familiar and unfamiliar faces improves with age. *Journal of Experimental Child Psychology*, *143*, 123-138.

Liu, S., Quinn, P. C., Wheeler, A., Xiao, N., Ge, L., & Lee, K. (2011). Similarity and difference in the processing of same-and other-race faces as revealed by eye tracking in 4-to 9-month-olds. *Journal of experimental child psychology*,*108*(1), 180-189.

McConachie, H. R. (1976). Developmental prosopagnosia. A single case report. *Cortex*, *12*, 76-82.

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*(1-2), 174-212.

McKone, E., Crookes, K., & Kanwisher, N. (2009). The cognitive and neural development of face recognition in humans. *The Cognitive Neurosciences*,*4*, 467-482.

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

Pascalis, O., de Schonen, S., Morton, J., Deruelle, C., & Fabre-Grenet, M. (1995). Mother's face recognition by neonates: A replication and an extension. *Infant Behavior and Development*, *18*(1), 79-85.

Pelphrey, K. A., Sasson, N. J., Reznick, J. S., Paul, G., Goldman, B. D., & Piven, J. (2002). Visual scanning of faces in autism. *Journal of autism and developmental disorders*, *32*(4), 249-261.

Riby, D. M., & Hancock, P. J. (2008). Viewing it differently: Social scene perception in Williams syndrome and autism. *Neuropsychologia*, *46*(11), 2855-2860.

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.

Schneider, W., & Pressley, M. (1997). *Memory development between two and twenty*. Mahwah, NJ: Lawrence Erlbaum Associates.

Schwarzer, G., Huber, S., Grüter, M., Grüter, T., Groß, C., Hipfel, M., & Kennerknecht, I. (2007). Gaze behaviour in hereditary prosopagnosia.*Psychological Research*, *71*(5), 583-590.

Shah, P., Gaule, A., Sowden, S., Bird, G., & Cook, R. (2015). The 20-item prosopagnosia index (PI20): a self-report instrument for identifying developmental prosopagnosia. *Royal Society Open Science*, *2*(6), 140343.

Stanislaw, H., & Todorov, N. (1999). Calculation of signal detection theory measures. *Behavior Research Methods, Instruments, and Computers*, *31*(1), 137-149.

Susilo, T., & Duchaine, B. (2013). Advances in developmental prosopagnosia research. *Current Opinion in Neurobiology, 23*, 423-429.

Turati, C., Di Giorgio, E., Bardi, L., & Simion, F. (2010). Holistic face processing in newborns, 3-month-old infants, and adults: Evidence from the composite face effect. *Child development*, *81*(6), 1894-1905.

Turati, C., Macchi Cassia, V., Simion, F., & Leo, I. (2006). Newborns' face recognition: Role of inner and outer facial features. *Child Development*, *77*(2), 297-311.

Want, S. C., Pascalis, O., Coleman, M., & Blades, M. (2003). Face facts: Is the development of face recognition in early and middle childhood really so special?. In O. Pascalis & A. Slater (Eds.), *The development of face processing in infancy and early childhood: Current perspectives* (pp. 207-221). Hauppauge, NY, US: Nova Science Publishers.

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*(1), 47-58. doi: 10.1016/j.dcn.2014.07.003

Wilmer, J. B., Germine, L., Chabris, C. F., Chatterjee, G., Gerbasi, M., & Nakayama, K. (2012). Capturing specific abilities as a window into human individuality: the example of face recognition. *Cognitive Neuropsychology*, *29*(5-6), 360-392.

Wilson, C. E., Palermo, R., Schmalzl, L., & Brock, J. (2010). Specificity of impaired facial identity recognition in children with suspected developmental prosopagnosia. *Cognitive Neuropsychology*, *27*(1), 30-45.

Yardley, L., McDermott, L., Pisarski, S., Duchaine, B., & Nakayama, K. (2008). Psychosocial consequences of developmental prosopagnosia: A problem of recognition. *Journal of Psychosomatic Research*, *65*(5), 445-451.

Figure captions

Figure 1: Examples of the CFMT-K and bicycle memory tasks. CFMT-K images adapted from Darymple, Garrido, and Duchaine (2014).

Figure 2: Example trials from the face and bicycle matching tasks.

Figure 3: Mean scores on the CFMT-K (face memory) task across school year. Children in Years 1-3 completed a 4-item version of the task; children in Years 4-6 completed the 6-item version of the task. Error bars indicate +/- 1 SEM. Chance level performance is 33.33%. Green triangles denote children who also performed poorly on the bicycle matching task (> 1.7 *SD* below the mean); red circles denote children who performed within the normal range on the bicycle matching test.

Figure 4: Mean scores on the face matching task across school year, retransformed to original scale. Error bars indicate 95% confidence intervals of the mean. Chance level performance is 33.33%. Individual cases who scored more than 1.7 *SD* below the mean for their year are marked in each year group. Green triangles denote children who also performed poorly on the bicycle matching task (> 1.7 *SD* below the mean); red circles denote children who performed within the normal range on the bicycle matching test.

Table 1: Results of children's face and bicycle memory tasks

		Face memory						**Bicycle memory**					
School Year	**Task format**	**N**	**Percentage correct M (SD)**	**Min.**	**Max.**	**Skew (std. error)**	**Shapiro-Wilk statistic**	**N**	**Percentage correct M (SD)**	**Min.**	**Max.**	**Skew (std. error)**	**Shapiro-Wilk statistic**
1	4-item	32	56.88 (16.86)	25	94	.51 (.43)	.96	32	63.69 (21.17)	27	92	-.24 (.44)	.92*
2		29	64.89 (17.58)	31	90	-.12 (.43)	.94	26	72.00 (15.99)	43	100	.02 (.46)	.97
3		37	72.29 (17.29)	42	100	-.40 (.41)	.94	37	80.83 (16.49)	33	98	-1.43 (.40)	.84*
4	6-item	36	64.37 (15.27)	39	96	.34 (.40)	.96	36	68.91 (12.00)	43	94	.10 (.40)	.98
5		3	64.67	31	99	-.13	.97	3	74.62	42	93	-.85	.92*

	2	(18.03)			(.41		1	(14.64)			(.43	
					)						)	
6	3	69.88	38	92	-.45	.95	2	79.90	51	92	-	.84*
	4	(14.89)			(.41		2	(11.60)			1.4	
					)						8	
											(.52	
											)	

Chance performance on the memory tasks was 33.33%

Table 2: Results of familiar face recognition task and parental rating of face recognition ability for Experiment 1.

	Familiar face recognition		Parental rating	
School Year	**N**	***d'*** **M (SD)**	**N**	**Rating (max 5) M (SD)**
1	12	1.71 (0.71)	24	3.83 (0.70)
2	3	3.38 (0.00)	11	3.91 (0.70)
3	8	2.59 (0.58)	18	4.33 (0.69)
4	8	2.19 (0.47)	15	4.27 (0.59)
5	5	2.24 (0.70)	5	4.00 (0.71)
6	7	2.92 (0.31)	11	4.27 (0.65)

Table 3: Cut-off scores in percentage correct and total correct (to the next lowest whole digit, in parentheses) for impairment on children's face and bicycle memory tasks. Z-scores indicate how many SDs below the mean chance performance falls

	Faces			**Bicycles**		
School Year	**-2 *SD***	**-1.7 *SD***	**Chance z-score** (33.33%)	**-2 *SD***	**-1.7 *SD***	**Chance z-score** (33.33%)
1	23.15* (11)	28.21* (13)	-1.40	21.36* (10)	27.71* (13)	-1.43
2	29.74*(14)	35.01 (16)	-1.80	40.03 (19)	44.82 (21)	-2.42
3	37.70 (18)	42.89 (20)	-2.25	47.86 (22)	52.80 (25)	-2.88
4	33.83 (24)	38.41 (27)	-2.03	44.91 (32)	48.51 (34)	-2.96
5	28.61* (20)	34.02 (24)	-1.74	45.34 (32)	49.73 (35)	-2.82
6	40.10 (28)	44.57 (32)	-2.45	56.69 (40)	60.17 (43)	-4.01

* Indicates cut-off score is below chance levels of performance

Table 4: Profile of individual cases who showed poor face memory ability

Participant	**School Year**	**Age (years)**	**Faces z**	**Bicycles z**	**Parental rating**	**Familiar faces test z**
W6_11	6	11	-2.17	.22	-	-
W2_06	2	7	-1.91	-	-	-
W_13	3	7	-1.77	-2.88	-	-
S3_08	3	8	-1.77	.65	4	-
R5_02	5	9	-1.89	.69	3	-0.79
W6_19	6	11	-1.71	1.16	-	-

Red indicates participant scored > 2 *SD*s below the mean. Orange > 1.7 *SD*s below the mean.

Table 5: Results of children's face and bicycle matching tasks, and correlation between face and bicycle matching

		Face Matching[1]					**Bike Matching[1]**					
Year	**N**	**Percentage correct M (SD)**	**Min.**	**Max.**	**Skew (std. error)**	**Shapiro-Wilk statistic**	**Percentage correct M (SD)**	**Min.**	**Max.**	**Skew (std. error)**	**Shapiro-Wilk statistic**	**Face and bicycle matching correlation[2]**
1	83	61.81 (19.56)	16	97	-.13 (.27)	.97*	61.14 (15.66)	27	93	-.21 (.27)	.98	.20
2	83	73.39 (18.42)	27	100	-.67 (.26)	.94*	70.66 (17.93)	30	100	-.47 (.26)	.96*	.34**
3	61	76.57 (18.07)	34	100	-.88 (.31)	.90*	73.81 (18.38)	24	100	-.71 (.31)	.94*	.27
4	76	82.27 (14.08)	47	100	-.88 (.28)	.93*	79.06 (15.71)	33	100	-.83 (.28)	.93*	.20
5	52	85.81 (13.03)	36	100	-1.54	.85*	82.50 (16.05)	27	100	-1.23	.88*	.33**

					(.33)					(.33)		
6	58		67		-1.08	.88*		63	100	-1.33	.84*	.12
		90.45 (8.43)		100	(.31)		91.72 (8.78)			(.31)		

Chance performance on the matching tasks was 33.33%

* Shapiro-Wilk statistics $p < .05$; indicates a significant departure from normality

** $p < .05$

[1]Raw data prior to transformation. For retransformed data, see Supplemental material

[2]Correlations calculated based on transformed means

Table 6: Results of familiar face recognition task and parental rating of face recognition ability for Experiment 1.

	Familiar face recognition			**Parental rating**		
School Year	**N**	***d'*** **M (SD)**	**Correlation with face matching[1]**	**N**	**Rating (max 5) M (SD)**	**Correlation with face matching[1]**
1	83	0.91 (0.84)	.045	82	3.93 (0.73)	.019
2	82	1.29 (0.86)	-.12	79	3.71 (0.72)	-0.16
3	39	1.88 (0.88)	.32**	60	3.93 (0.73)	-0.14
4	64	1.77 (0.89)	.27**	75	3.87 (0.68)	-0.11
5	52	2.04 (0.78)	-.31**	49	3.88 (0.67)	0.10
6	58	1.78 (0.86)	-.10	57	3.79 (0.75)	.080

** $p < .05$

[1]Correlations calculated based on transformed means

Table 7: Cut-off scores in percentage correct for impairment on children's face and bicycle matching tasks. Scores have been retransformed from square root transformation. Z-scores indicate how many SDs below the mean chance performance falls

	Faces			Bicycles		
School Year	**-2 *SD***	**-1.7 *SD***	**Chance z-score** (33.33%)	**-2 *SD***	**-1.7 *SD***	**Chance z-score** (33.33%)
1	20.31*	28.42*	-1.51	26.23*	32.65*	-1.67
2	27.90*	36.87	-1.82	24.93*	33.80	-1.72
3	34.91	43.39	-2.05	30.45*	39.19	-1.90
4	50.94	57.43	-2.73	39.40	47.36	-2.22
5	54.15	60.70	-2.84	42.53	50.86	-2.31
6	69.5	73.85	-3.99	70.46	74.96	-3.92

	Faces			Bicycles		
School Year	**-2 *SD***	**-1.7 *SD***	**Chance z-score** (33.33%)	**-2 *SD***	**-1.7 *SD***	**Chance z-score** (33.33%)
1	20.31*	28.42*	-1.51	26.23*	32.65*	-1.67
2	27.90*	36.87	-1.82	24.93*	33.80	-1.72
3	34.91	43.39	-2.05	30.45*	39.19	-1.90
4	50.94	57.43	-2.73	39.40	47.36	-2.22
5	54.15	60.70	-2.84	42.53	50.86	-2.31
6	69.5	73.85	-3.99	70.46	74.96	-3.92

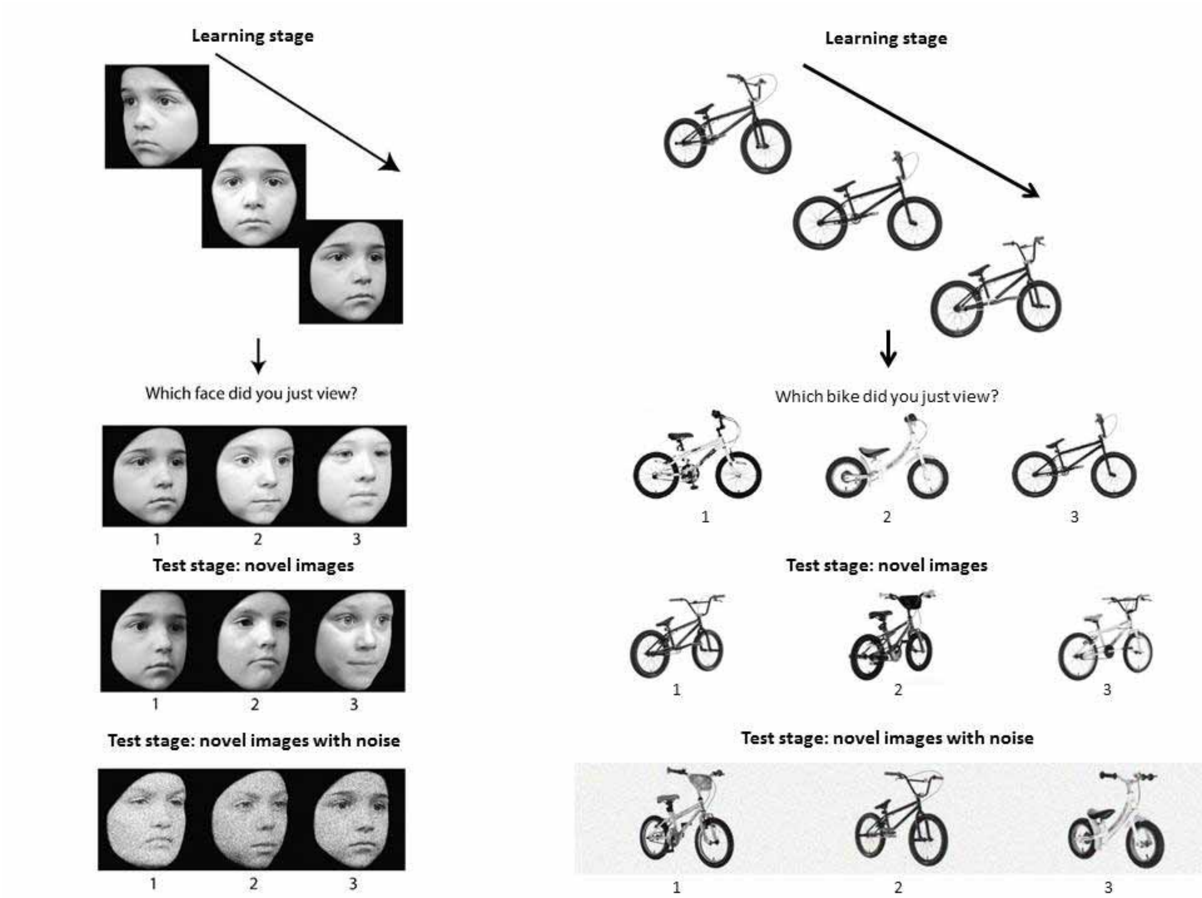
* Indicates cut-off score is below chance levels of performance

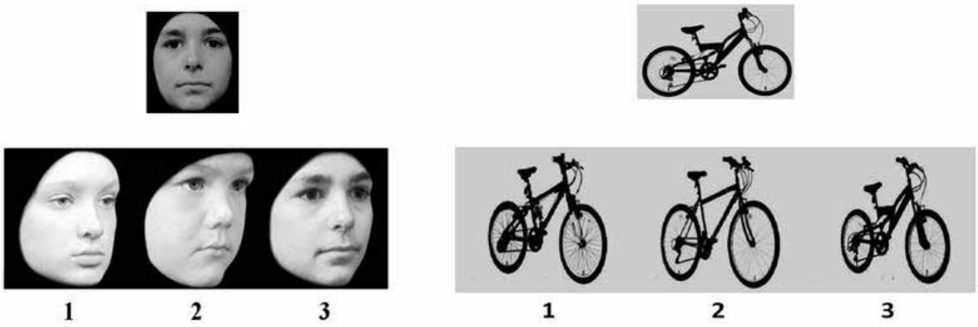
Table 8: Profile of individual cases who showed poor face matching ability

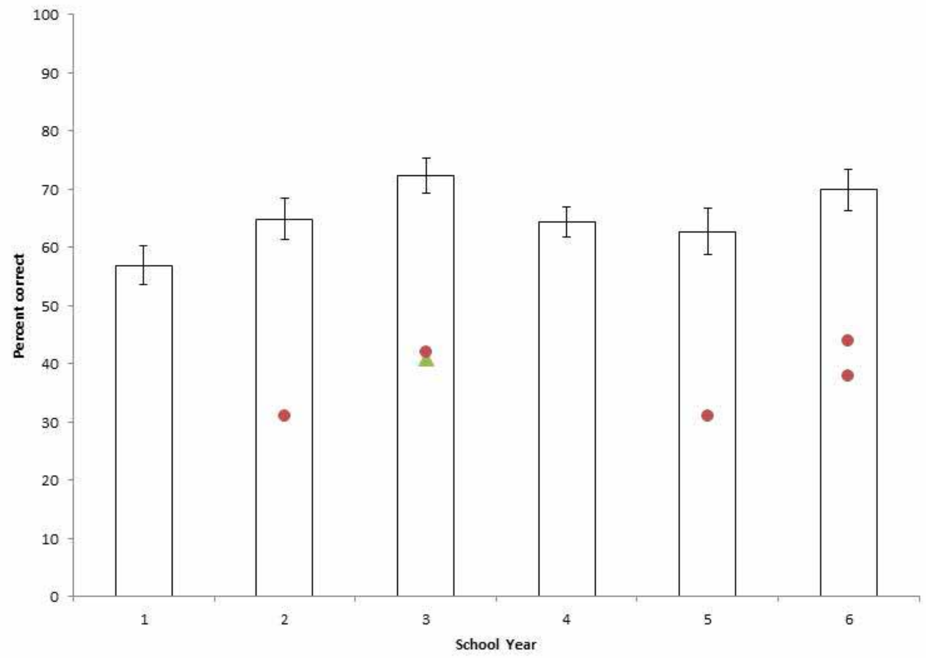
Participant	School Year	Age (years)	Faces z	Bicycles z	Parental rating	Familiar faces test z
P2_15	2	7	-2.02	-.82	3	-2.12
M4_24	4	9	-2.19	-1.17	4	-1.14
H4_06	4	9	-2.04	-.75	5	-.18
P5_05	5	9	-2.75	-2.52	3	-.99
B5_12	5	10	-2.01	-.65	4	.94
B6_11	6	11	-2.19	-1.06	3	-.76
H2_09	2	8	-1.93	-.52	3	.67
B2_45	2	7	-1.78	-.19	4	-1.40
B2_40	2	7	-1.71	-1.45	4	.44
B3_33	3	8	-1.94	0.58	4	
P3_01	3	7	-1.94	-1.70	4	-1.19
M3_09	3	8	-1.9	1.14	4	-0.98
M3_06	3	7	-1.74	-.74	4	-.63
P3_13	3	8	1.70	-0.42	4	0.57
B4_06	4	8	-1.89	-1.73	5	

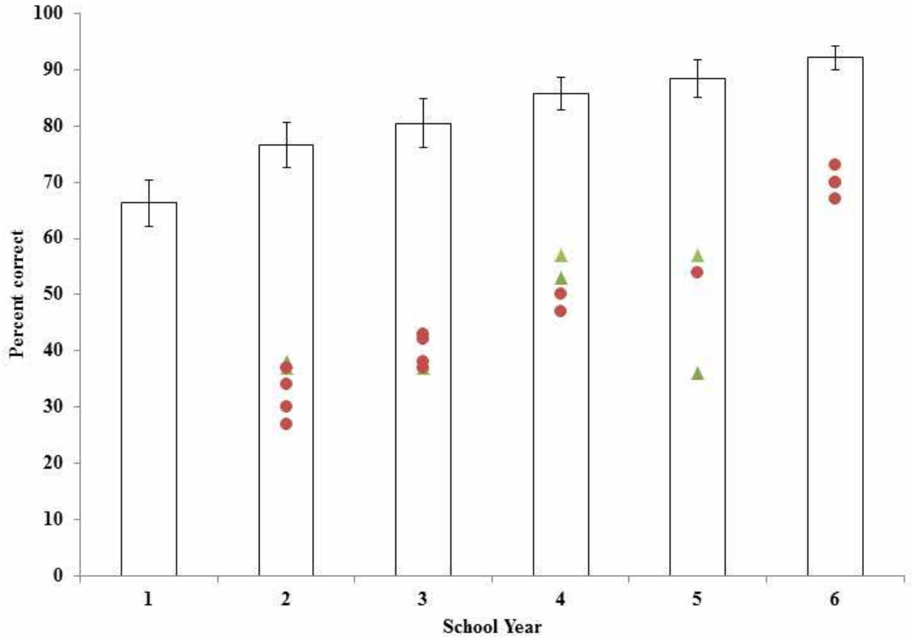
H4_13	4	8	-1.74	-2.10	4	-1.51
H5_06	5	10	-1.89	-1.73	4	.35
H6_04	6	10	-1.97	1.21	3	.82
H6_07	6	11	-1.97	0.45	4	.18
B6_12	6	10	-1.74	-0.04	4	-1.23

Red indicates participant scored > 2 *SD*s below the mean. Orange > 1.7 *SD*s below the mean.


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


1
2
3
1
2
3


100
90
80
70
60
50
40
30
20
10
0
Percent correct
1
2
3
4
5
6
School Year


Percent correct
100
90
80
70
60
50
40
30
20
10
0
1
2
3
4
5
6
School Year