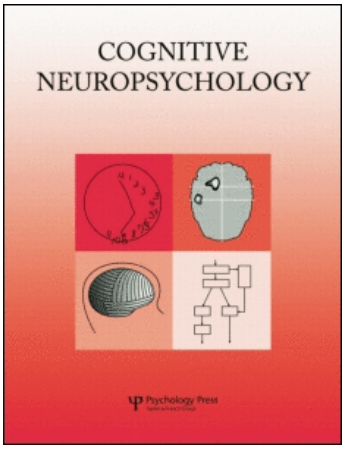

Cognitive Neuropsychology
Publication details, including instructions for authors and subscription information:
http://www.informaworld.com/smpp/title~content=t713659042

Diagnosing prosopagnosia: Effects of ageing, sex, and participant-stimulus ethnic match on the Cambridge Face Memory Test and Cambridge Face Perception Test

Devin C. Bowles [a]; Elinor McKone [a]; Amy Dawel [a]; Bradley Duchaine [b]; Romina Palermo [ac]; Laura Schmalzl [c]; Davide Rivolta [c]; C. Ellie Wilson [c]; Galit Yovel [d]
[a] Australian National University, Canberra, Australia [b] Institute of Cognitive Neuroscience, University College London, London, UK [c] Macquarie Centre for Cognitive Science (MACCS), Macquarie University, Sydney, Australia [d] Tel Aviv University, Tel Aviv, Israel

First published on: 16 November 2009

To cite this Article Bowles, Devin C., McKone, Elinor, Dawel, Amy, Duchaine, Bradley, Palermo, Romina, Schmalzl, Laura, Rivolta, Davide, Wilson, C. Ellie and Yovel, Galit(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: 5, 423 — 455, First published on: 16 November 2009 (iFirst)
To link to this Article: DOI: 10.1080/02643290903343149
URL: http://dx.doi.org/10.1080/02643290903343149

COGNITIVE NEUROPSYCHOLOGY, 2009, 26 (5), 423–455

Psychology Press
Taylor & Francis Group

Diagnosing prosopagnosia: Effects of ageing, sex, and participant–stimulus ethnic match on the Cambridge Face Memory Test and Cambridge Face Perception Test

Devin C. Bowles[1], Elinor McKone[1], Amy Dawel[1], Bradley Duchaine[2], Romina Palermo[1,3], Laura Schmalzl[3], Davide Rivolta[3], C. Ellie Wilson[3], and Galit Yovel[4]

[1]Australian National University, Canberra, Australia; [2]Institute of Cognitive Neuroscience, University College London, London, UK; [3]Macquarie Centre for Cognitive Science (MACCS), Macquarie University, Sydney, Australia; [4]Tel Aviv University, Tel Aviv, Israel

The Cambridge Face Memory Test (CFMT) and Cambridge Face Perception Test (CFPT) have provided the first theoretically strong clinical tests for prosopagnosia based on novel rather than famous faces. Here, we assess the extent to which norms for these tasks must take into account ageing, sex, and testing country. Data were from Australians aged 18 to 88 years ($N = 240$ for CFMT; 128 for CFPT) and young adult Israelis ($N = 49$ for CFMT). Participants were unselected for face recognition ability; most were university educated. The diagnosis cut-off for prosopagnosia (2 *SD*s poorer than mean) was affected by age, participant–stimulus ethnic match (within Caucasians), and sex for middle-aged and older adults on the CFPT. We also report internal reliability, correlation between face memory and face perception, correlations with intelligence-related measures, correlation with self-report, distribution shape for the CFMT, and prevalence of developmental prosopagnosia.

Keywords: Prosopagnosia; Face recognition; Ageing; Gender; Other-race effects; Aging.

Prosopagnosia refers to the inability to recognize and discriminate faces. It has a well-known acquired form, where an individual who could previously recognize faces loses this ability after stroke or other brain injury. It also has a more recently discovered developmental or congenital form, which can run in families, in which individuals with no known history of brain injury

Correspondence should be addressed to Elinor McKone, Department of Psychology, Australian National University, Canberra, ACT 0200, Australia (E-mail: elinor.mckone@anu.edu.au)

Supported by the following grants: Australian Research Council DP0450636 and DP0984558 to E.M., Economic and Social Research Council RES–061–23–0400 to B.D., Australian Research Council Special Research Centre for Cognitive Science and Cognitive Neuropsychology S0001507 for work at Macquarie Centre for Cognitive Science (MACCS), and Marie Curie IRG 046448 to G.Y. We thank Kate Crookes for assistance in manuscript preparation and Prof Evelyn Teng of the University of Southern California for providing instructions on administering and scoring the MAT test.

http://www.psypress.com/cogneuropsychology DOI:10.1080/02643290903343149

appear never to properly develop the ability to tell faces apart (e.g., Behrmann, Avidan, Marotta, & Kimchi, 2005; Duchaine, Germine, & Nakayama, 2007a; Schmalzl, Palermo, & Coltheart, 2008). There are no previous data on prevalence rates of developmental prosopagnosia from formal behavioural testing, but an estimate from self-report questionnaires with follow-up semistructured interview suggested a surprisingly high prevalence of 2.47% in the general population ($N = 687$; Kennerknecht et al., 2006). In combination with the potential for prosopagnosia to be a cause of social anxiety (Yardley, McDermott, Pisarski, Duchaine, & Nakayama, 2008), or to be mistaken for other conditions that also manifest poor face processing (e.g., autism spectrum disorder), it is clear that both clinical neuropsychologists and researchers need access to a quick, reliable method of diagnosing prosopagnosia.

Creating a simple test to diagnose prosopagnosia is not a straightforward task, due to a mixture of practical and theoretical difficulties. Tests of famous face recognition are theoretically strong, in that these measure an ability similar to everyday face recognition—namely, the ability to identify a person, in a new image, from many hundreds or thousands of possibilities. However, famous face tests are beset by practical difficulties: People famous in one testing location can be unknown in another; people famous to one age group can be unknown to another; people famous at one time can be merely history even 10 years later; and participants vary in the extent to which they have engaged with sources of famous people such as popular culture (e.g., film stars) or politics. Thus, although famous face tests should always be included in a battery to diagnose prosopagnosia, the only feasible possibility for widely applicable off-the-shelf neuropsychological tests are tasks that assess memory or perception for novel faces.

It is only recently that theoretically valid tests involving novel faces have been developed. Two traditional neuropsychological tests have been shown to be invalid and in particular to frequently falsely diagnose individuals with prosopagnosia as normal (Duchaine & Nakayama, 2004, 2006a). The Warrington Recognition Memory for Faces test (Warrington, 1984) allows recognition of "faces" based on clothing and distinctive hairstyles in the images. The Benton Facial Recognition test (Benton, Sivan, Hamsher, Varney, & Spreen, 1983) is somewhat better in that it removes clothing and hair from the stimuli and precludes the most obvious form of image matching by using different face viewpoints for learning images (front) and probe images (three-quarter). However, because this test uses *simultaneous* presentation of learning and probe images with *unlimited presentation duration*, unusual strategies can still support apparently "normal" accuracy of face recognition (e.g., back-and-forth matching based purely on eyebrows, Duchaine & Weidenfeld, 2003), and 73% of a group of 19 prosopagnosics performed at normal accuracy levels on this task (Duchaine & Nakayama, 2004, 2006a).

To overcome these difficulties, the Cambridge Face Memory Test (CFMT; Duchaine & Nakayama, 2006a) and Cambridge Face Perception Test (CFPT; Duchaine et al., 2007a) were developed. The CFMT requires recognition of 6 learned faces in three stages: recognition of the same images; recognition of the same faces in different images (different viewpoint and/or lighting); and recognition of the same faces in different images covered with heavy visual noise (see Figure 1A). The CFPT requires participants, on each trial, to order a series of faces for similarity to a target face, where the comparison stimuli comprise the target face morphed towards several different faces by varying degrees (Figure 1B). Both the CFMT and CFPT have been shown to be able to reliably diagnose prosopagnosia—for example, people who perform very poorly on famous faces tests often also perform poorly on the CFMT and CFPT—and the tasks also show another theoretically expected laboratory phenomenon—namely, a large face inversion effect in normal individuals (Duchaine et al., 2007a; Duchaine & Nakayama, 2006a; Duchaine, Yovel, & Nakayama, 2007b). As a result, the tests, particularly the CFMT, have become widely used by

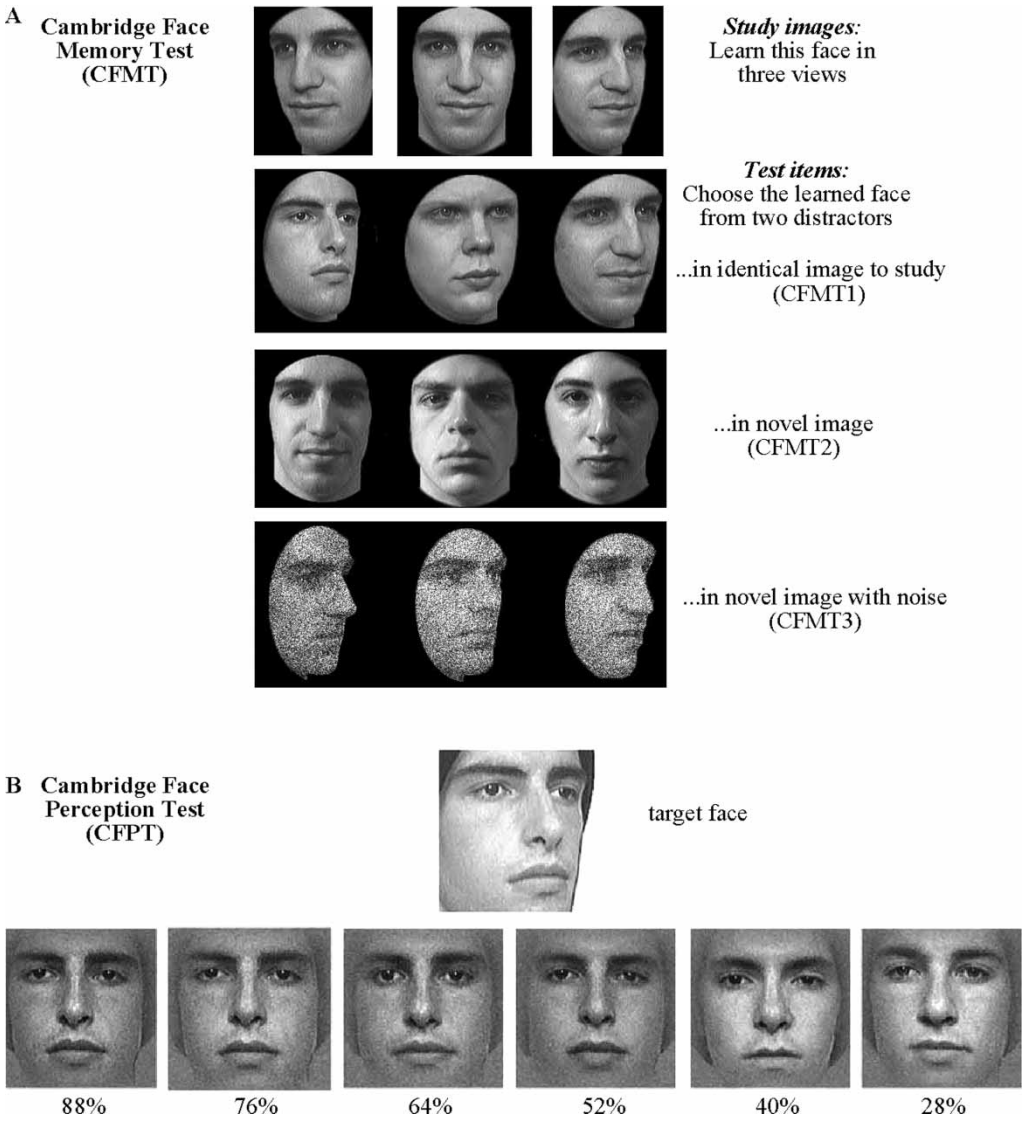

Figure 1. *Tasks. A. Examples of study images and test items from the Cambridge Face Memory Test. B. Images from an item in the Cambridge Face Perception Test. Numbers under each morph image (the frontal shots) indicate the percentage of the target face (shown at the top in three-quarter view) in the morph. The six frontal shots are initially ordered randomly, and participants sort them based on similarity to the target image.*

face recognition and prosopagnosia researchers since their publication (e.g., Bate, Haslam, Tree, & Hodgson, 2008; DeGutis, Bentin, Robertson, & D'Esposito, 2007; Herzmann, Danthiir, Schacht, Sommer, & Wilhelm, 2008; Iaria, Bogod, Fox, & Barton, 2009).

Are the original CFMT and CFPT norms appropriate for all participants?

Published control sample data on the CFMT and CFPT come primarily from young adults from Boston in the United States of America (with

smaller sample sizes available for middle-aged adults with mixed USA and UK origin). However, broad use of the CFMT and CFPT—that is, in other age groups and in other countries—requires access to task norms from appropriate control samples. The present article addresses the question of to what extent the original norms can be used broadly and to what extent it is necessary to obtain control data from groups very closely matched on variables such as age, sex, and/or exposure to exactly the same "diet" of face types as the potential prosopagnosic.

The first important issue concerns ageing. The CFMT was originally created with young adult participants, and some data have since been published for adults with a mean age up to 46 years. As shown in Table 1, results suggest no age-related decline in performance from 20 to 46 (Duchaine & Nakayama, 2006a; Duchaine, Yovel, Butterworth, & Nakayama, 2006; Duchaine et al., 2007b; Garrido, Duchaine, & Nakayama, 2008). There has been no systematic investigation of older participants, and clearly this is needed. Studies in the cognitive psychology literature often show that face memory declines with age (e.g., Adams-Price, 1992; Bastin & Van der Linden, 2003; Lamont, Stewart-Williams, & Podd, 2005; Lindholm, 2005). In addition, the CFMT uses young adult face stimuli, and older participants have poor memory for young adult faces even when they can be better at older adult faces (Anastasi & Rhodes, 2005; Lamont et al., 2005). Age-related decline would also be expected on the CFPT: Face perception declines with age on other tasks (e.g., Grady, McIntosh, Horwitz, & Rapoport, 2000; Lott, Haegerstrom-Portnoy, Schneck, & Brabyn, 2005) including specifically cross-view matching of morphs as required by the CFPT (Habak, Wilkinson, & Wilson, 2008); the CFPT again uses young adult faces; the CFPT is a speeded task, and reaction time slows with age; and the CFPT requires use of a computer mouse, and some older participants might be less skilled at mouse usage than younger participants. As shown in Table 1, previous studies have found a very small increase in mean CFPT error rate with age, but have tested only two groups, with mean ages of 33.4 years and 46.5 years (Duchaine et al., 2007a; Garrido et al., 2008).

A second issue concerns sex differences. Predictions here are less clear. On face memory tasks, several studies report that women outperform men, but one found this only for individuals with lower than average IQ (Herlitz & Yonker, 2002), and others have found it only for women's faces (Lewin & Herlitz, 2002; McKelvie, Standing, St Jean, & Law, 1993) while the CFMT uses only men's faces. In the only previous report for the CFMT, young adult women showed a small but nonsignificant advantage over men (Duchaine & Nakayama, 2006a). Turning to the CFPT, the two previous studies containing norm data did not analyse sex effects; we are also not

Table 1. *All information regarding age effects available from previous studies*

		Age (years)		*Performance*		
Task	*Article*	*M* (*SD*)	*N*	*Mean*	*SD*	*Cut-off*
CFMT	Duchaine & Nakayama (2006a)	20.2 (1.8)	50	57.9	7.9	42.1
	Garrido, Duchaine, & Nakayama (2008)	33.3 (6.4)	13	58.9	7.4	44.1
	Duchaine, Yovel, & Nakayama (2007b)[a]	45.1	20	59.6	7.6	44.4
	Duchaine, Yovel, Butterworth, & Nakayama (2006)	46.5 (7.7)	9	62.8	6.8	49.2
CFPT	Garrido, Duchaine, & Nakayama (2008)	33.4 (5.5)	11	35.5	14.6	64.7
	Duchaine, Germine, & Nakayama (2007a)	46.5	21	36.7	12.2	61.1

Note: CFMT = Cambridge Face Memory Test. CFPT = Cambridge Face Perception Test. CFMT performance in groups with mean ages between 20 and 46 years of age, and CFPT performance in groups with mean ages between 33 and 46 years of age (young adult samples from Boston USA, adults older than 30 years from USA and UK).

[a]The same control data were also used in Duchaine, Germine, and Nakayama (2007a; a typographical error in that article indicated the sample size to be 8 rather than 20).

aware of any studies using other face *perception* tasks that have tested sex effects.

A third issue concerns ethnic match between the face stimuli and the participants. In the "other-race effect", memory and perceptual discrimination is poorer for faces not of the participant's race than for own-race faces. This means that the CFMT and CFPT—which use Caucasian faces—will of course not be appropriate to diagnose prosopagnosia in, say, an Asian individual living in Korea, or a Masai individual in Kenya. Less well recognized, however, is that there may very well also be effects of ethnicity even within Caucasian faces. Consistent with this idea, memory for Turkish and German participants looking at Turkish and German faces shows a pattern of poorer memory for the other group (Sporer, 1999); also, Caucasian South Africans show poorer memory for US Caucasian faces than for South African Caucasian faces (Chiroro, Tredoux, Radaelli, & Meissner, 2008).

Previous young adult norms for the CFMT and CFPT come from participants with ethnic background well matched to the face stimuli: specifically, all from the Boston area in the USA. However, migration patterns from Europe differ for those from Australia and the USA, and so the typical Caucasian face in Australia is physically rather different from the American faces used in the CFMT and CFPT (see details in Results and Discussion). Theoretically, this could produce poorer CFMT and CFPT performance in Australians than in Americans, via one of two possible mechanisms. First, the participant's "face-space" could be less than ideally tuned for relatively unfamiliar Caucasian subtypes (cf. Valentine, 1991). Second, social outgroup categorization could occur, and Bernstein, Young, and Hugenberg (2007) have shown that merely categorizing someone as being in another personality group, or as attending another university (all faces and participants Caucasian) leads to poorer memory than that for the same faces described as ingroup members. These ideas raise the possibility that country-specific norms might be needed to ensure that individuals are not misdiagnosed as having prosopagnosia when they do not.

The present study reports CFMT and CFPT data obtained from the general population of educated individuals unselected for face recognition ability. The primary data set was from Caucasian Australians (and a few Caucasian New Zealanders)[1] and included ages across the adult lifespan, particularly adults older than those reported in most previous tests (i.e., older than 50 years). A further data set, used only for addressing effects of country of origin, was from young adult Israelis. Our interest was in describing the *variance* of scores in each group as much as the mean performance: This is important because a diagnosis of prosopagnosia normally is made when an individual's performance falls a certain number of standard deviations worse than the mean (typically, 2 *SD*s). Our initial research questions concerned the effect of ageing, sex, and country of origin (Australia vs. Israel vs. America) on diagnosis criteria. Our data also allowed us to provide some limited discussion of the effects of intelligence-related measures, to perform reliability analyses for the tasks, to ask whether simple participant self-reports accurately reflect real ability, to assess the extent to which face perception and face memory are independent abilities, and to examine prevalence rates of likely prosopagnosia in the general population using objective tests.

Method

Data were collated from two different laboratories in Australia and one in Israel and came from several research projects: Data had originally been collected as part of studies investigating other theoretical questions within the normal population, or obtained from participants tested as controls for specific individual prosopagnosics. Most participants had been tested on the CFMT; fewer were also tested on the CFPT.

[1] Ethnic background of Caucasian New Zealanders is similar to that of Caucasian Australians.

All studies involved participants performing additional cognitive tests in addition to those reported here.

Participants

Australian samples. These participants were all Caucasian and living in Australia (95.2%) or visiting Australia from New Zealand (4.8%). Most had been born in these countries, and all had spent the majority of their lifetime living in them. We excluded any participants with known history of major brain injury, or other major disorders likely to affect face recognition (e.g., 1 participant with schizophrenia).

Testing laboratories were located in Canberra ($N = 192$) or Sydney ($N = 49$). All participants were unselected as regards face recognition ability. Known relatives of previously diagnosed prosopagnosics were excluded. The extent to which the participants represented a random sample of the community varies across the different original research projects, labelled here as Studies 1–4. Study 1 (Bowles & McKone, Australian National University, ANU, Canberra), which tested young adults (18–30 years, $N = 37$), 55–64-year-olds ($N = 25$), 65–74-year-olds ($N = 30$), and 74–88-year-olds ($N = 15$), used general community recruitment from churches, sporting clubs, bridge clubs, and so on and recruited undergraduate students (from the ANU) only to the extent necessary to match the relatively high education level of the older participant groups. Study 2 (Dawel & McKone, ANU), which tested 80 young adults (18–32 years), recruited either from the general undergraduate population of the ANU or from a highly academically selected group of ANU "summer scholars" (see Results for more details). Study 3 (Palermo, Schmalzl, Rivolta, & Wilson, Macquarie Centre for Cognitive Science, MACCS, Sydney) recruited 49 participants of various ages on a word-of-mouth basis as controls for projects involving testing known prosopagnosics. Study 4 (Bowles & McKone, ANU) tested a handful of extra participants in the 43–51 years age range ($N = 5$) using the general procedures of Study 1 but without testing all the tasks. Overall, education level of the total sample was higher than that in the general Australian population: The great majority of the participants (86%) had at least some post-secondary-school education.

ANU participants were paid $10 (those tested in 2006) or $12 (2007). MACCS participants either volunteered for no financial return or were paid $15 per hour. All participants were tested individually in quiet rooms.

Israeli sample. For Study 5, all participants ($N = 49$) were Caucasian and living in Israel. Most had been born in Israel. All were undergraduate psychology students, aged 19–31 years (mean age = 22), with a minimum score of 660 on the Israeli Scholastic Aptitude Test (top 10%). They participated for course credit.

CFMT and CFPT

The CFMT and CFPT were administered on Macintosh computers following the standard instructions (Duchaine et al., 2007a; Duchaine & Nakayama, 2006a; Duchaine et al., 2007b). Studies 1, 2, and 4 used a CRT-screen eMac with a 16-inch monitor running Mac OS X with screen resolution 1152 × 864, refresh rate 80 Hz, and contrast and brightness maximized. Participants sat a comfortable distance from the screen (approximately 60 cm) in Study 1; in Study 2, viewing distance was 90 cm (with chin rest). Study 3 used a 15-inch Macintosh Power Book G4 running OSX, and participants were placed at a distance of approximately 50 cm from the computer screen. Study 5 used a 17-inch CRT monitor connected to a PC running Windows XP, with a viewing distance of 45 cm controlled by a chin rest. (Note that face recognition is insensitive to changes in face stimulus size across the range created by these small differences in viewing distance; Loftus & Harley, 2005; McKone, 2009.)

Both tasks were scored in the standard way. For the CFMT, scores represent number of trials on which the learned face was correctly chosen, out of 18 for Section 1 "same images" (chance = 6), out of 30 for Section 2 "novel images" (chance = 10),

out of 24 for Section 3 "novel images with noise" (chance = 8), and out of 72 for the total summed score (chance = 24). All CFMT scores reported here (CFMT1, CFMT2, CFMT3, CFMTtotal) are for the upright version of the test. On the CFMT, a higher score equals better performance.

For the CFPT, participants sort the 6 test faces from most like to least like the target face. For each trial, the final sorted order is scored by summing the deviations from the correct order (e.g., if a face is five places away from its proper place, it contributes 5 to the score). Each of 8 target faces is used on two trials, one with all faces upright (CFPTupright), the other with all faces inverted (CFPTinverted). In each orientation, a perfect score is zero, and chance is 93.3 errors. On the CFPT, a higher score equals poorer performance.

Self-rated face recognition ability

In Studies 1 and 2, participants were asked to rate their ability to recognize faces in everyday life "compared to the average person", when they were close enough to see people clearly (0 = much worse than average; 5 = average; 10 = much better than average). The experimenter specified that this meant how well the participant recognized *faces* as familiar, and did not mean how well the participant remembered people's *names*. Older participants in Study 1 were also asked to rate their face recognition ability at present compared to when they were 30 years (0 = much worse now than when they were 30; 5 = equal to when they were 30; 10 = much better now than when they were 30).

Visual acuity at computer-viewing distances

All participants were tested wearing their usual optical correction. Participants in Study 1 were asked to confirm that they could focus without seeing blur both on a finger held up at 30 cm and on the computer screen. No participant reported any difficulty with focus at these distances. In Studies 2 and 4, normal visual acuity (i.e., the equivalent of 20/20 vision) was verified at 50 cm and 3 m using Snellen charts. In Study 3, some participants ($N = 5$) were tested on a contrast sensitivity test (Functional Acuity Contrast Test, FACT—Vision Sciences Research Corporation, 2002) and had normal performance.

Mental Alternation Test (MAT)

Participants in Study 1 (i.e., including the majority of participants older than 55 years) were tested on the MAT, which provides a quick and fairly reliable method of screening participants for dementia (Jones, Teng, Folstein, & Harrison, 1993). In the MAT, participants produce aloud alternating numbers and letters of the alphabet, starting from 1 and A: that is, "1, A, 2, B, 3, C, 4, . . . ". Progression from the last letter or number to the next number or letter is counted as one alternation, and participants have 30 s to produce as many alternations as possible. Participants were informed of the 30-s time limit and were encouraged to work quickly but accurately. Consistent with previous findings (Sliwinski, 1997), MAT performance decreased with age, most noticeably between the young adult group and the 55–64-year-old group. Mean number of alternations was: young adults = 31.1; 55–64-year-olds = 26.8; 65–74-year-olds = 25.3; 75–88-year-olds = 25.1. Scores below 15 are indicative of probable cognitive impairment (e.g., arising from age-related dementia, although note low scores can also arise from other factors such as stuttering), as demonstrated via comparison to more complete tests such as the Mini-Mental State Examination and Trailmaking Test Part B (Billick, Siedenburg, Burgert, & Bruni-Solhkhah, 2001; Jones et al., 1993; Salib & McCarthy, 2002). Proportion of MAT-tested participants falling below a score of 15 was: young adults = 2.8% ($N = 1$); 55–64-year-olds = 12.0%; 65–74-year-olds = 10.3%; 75–88-year-olds = 13.3%.

Computer mouse test

The CFPT relies on skilled computer mouse usage. Participants in Study 1 (i.e., including most of the participants older than 55 years) were asked to rate their mouse skill from 1–5. They were not tested on the CFPT if they rated themselves as 1 (had never used a computer

mouse before) or 2 (had little experience with mouse use and were apprehensive of it). Remaining Study 1 participants were tested on the CFPT and were also given an objective mouse test that measured time taken to sequentially click 11 numbered boxes scattered around the screen. As expected, mean completion time increased with age (young adults = 13.9 s; 55–64-year-olds = 18.3 s; 65–74-year-olds = 21.5 s; 75–88-year-olds = 23.3 s).

Results and discussion

Our first set of analyses—examining the effects of age, sex, and country of origin (and also intelligence-related variables, which are of relevance to interpreting the country of origin effects)—were derived from "control" or "norm" participants. Importantly, these excluded 5 individuals whose scores indicated that in fact they had probable prosopagnosia (see later section *"Prevalence of prosopagnosia in an unselected educated Australian sample"* for detailed rationale).

Throughout, the CFMT scores represent accuracy, and thus cut-off scores for prosopagnosia diagnosis are 2 standard deviations lower than the mean. The CFPT scores represent error rate, and so the cut-off scores are 2 standard deviations higher than the mean. Also note that, throughout, CFMT and CFPT scores are presented on their original scales (i.e., number correct out of 72 for CFMT; number of errors up to chance score of 93.3 for CFPT), rather than having been converted to percentages.

No effect of the different Australian sample groups

Before proceeding to the analyses of interest, it was important to demonstrate that there were no differences between the three main Australian sample groups. This was done in two ways. First, scores for the CFMT were available for two sets of young adults, from Study 1 and Study 2, and we confirmed that these two groups did not differ in CFMT scores, on either the mean (CFMTtotal score: Study 1 = 54.6, Study 2 = 55.3), $t(113) = 0.286$, $p > .7$, or the standard deviation (Study 1 = 9.4; Study 2 = 8.3, Levene's test for equality of variances, $F < 1$). Second, to allow comparison of all four Australian samples, Figure 2 shows a scatterplot of age against CFMTtotal and CFPTupright scores with individuals (excluding the identified prosopagnosics) colour coded by the study in which they were tested. Taking into account the general age trend, the individuals from the three different samples overlapped nicely: In particular, it was not the case that the participants from Sydney (Study 3) had higher or lower performance than the participants from Canberra (Studies 1, 2, 4).

Ageing: Australian sample

Cambridge Face Memory Test. Figure 3A shows a scatterplot of age against CFMTtotal score, including all participants (aged 18–88 years). Table 2 gives group means for five age groups in which we had reasonable numbers of participants: young adults (18–35 years), early middle age (36–49 years), and then decades for participants in their 50s, 60s, and 70s (note the 8 participants in their 80s do not appear in the table). In Figure 3A and Table 2, participants are included regardless of Mental Alternation Test (MAT) scores. We did not initially remove participants with low MAT scores because age-related decline in general cognitive functioning is a typical feature of the normal population against which prosopagnosia must be diagnosed. However, where clinicians and researchers are testing an older potential prosopagnosic who clearly does not suffer from dementia, norms derived from a population excluding individuals with MAT scores less than 15 may be considered more useful. This information is provided in Table 3. Results show no evidence that scores for diagnosing prosopagnosia were affected by whether an MAT exclusion was (Table 3) or was not (Table 2) applied (if anything, in fact, the tendency was for slightly poorer performance when participants with poor MAT scores were excluded than in the full sample, i.e., a trend in the reverse direction to that which would be predicted if MAT mattered). Therefore, reported statistical

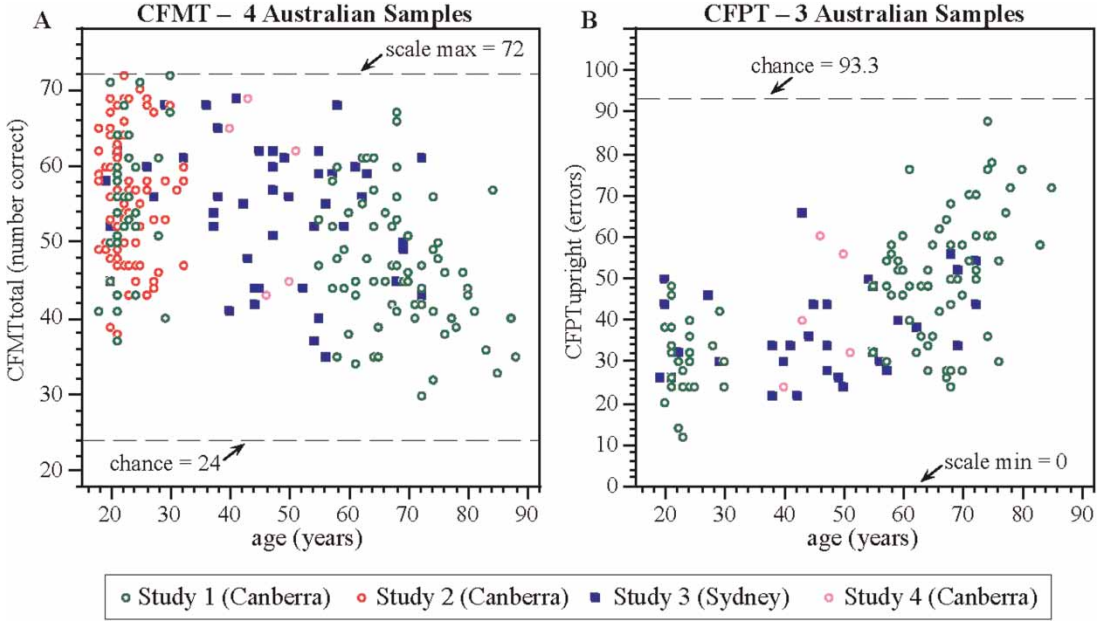

Figure 2. *(A) Cambridge Face Memory Test (CFMT) and (B) Cambridge Face Perception Test (CFPT) scores from our Australian samples, showing no apparent effect of the study in which participants were tested; in particular, taking into account the age trend, it was not the case that participants from Sydney (Study 3) performed either better or worse than participants from Canberra (Studies 1, 2, 4). We recommend readers view this figure in colour: please see the online issue of the Journal.*

analysis was conducted using all participants (excluding the prosopagnosics).[2] All tests were two-tailed.

Very noticeable age-related decline was apparent on the CFMT. Considering CMFTtotal scores, analysis of variance (ANOVA) showed a strong difference in mean performance across the five age groups, $F(4, 223) = 9.749$, $MSE = 73.354$, $p < .001$, and there was no change in variance: Levene's test for equality of variances, $F(4, 223) = 1.052$, $p > .3$. To address when age-related decline begins, we compared the young adult group to successively older comparison groups. The early middle-age group (36–49 years) showed no change in mean performance relative to young adults, $t(143) = 0.114$, $p > .9$. However, participants in their 50s (50–59 years) already showed significantly poorer mean performance than young adults, $t(146) = 2.456$, $p < .02$, suggesting that age-related decline begins at approximately 50 years of age.

We also fitted age-related curves to the CFMTtotal scores from all individuals, as shown in Figure 3A. This procedure had two aims. First, it provided further information regarding the starting point of age-related decline. Second, it allowed us to calculate the best estimate of prosopagnosia cut-offs: Given that there was no change in variance of CFMTtotal scores across age groups (see Levene's test above), the most reliable method of estimating the diagnosis cut-off comes from providing a formula for the estimated mean at any given age together with an *average* measure of the standard deviation derived from the entire sample.

Fit results were as follows. A linear-only fit explained significant variance ($R^2 = .151$, $p < .001$, $df = 234$) but, importantly, a second-order

[2] Statistical results with the MAT exclusion applied did not differ in any important ways from those reported.

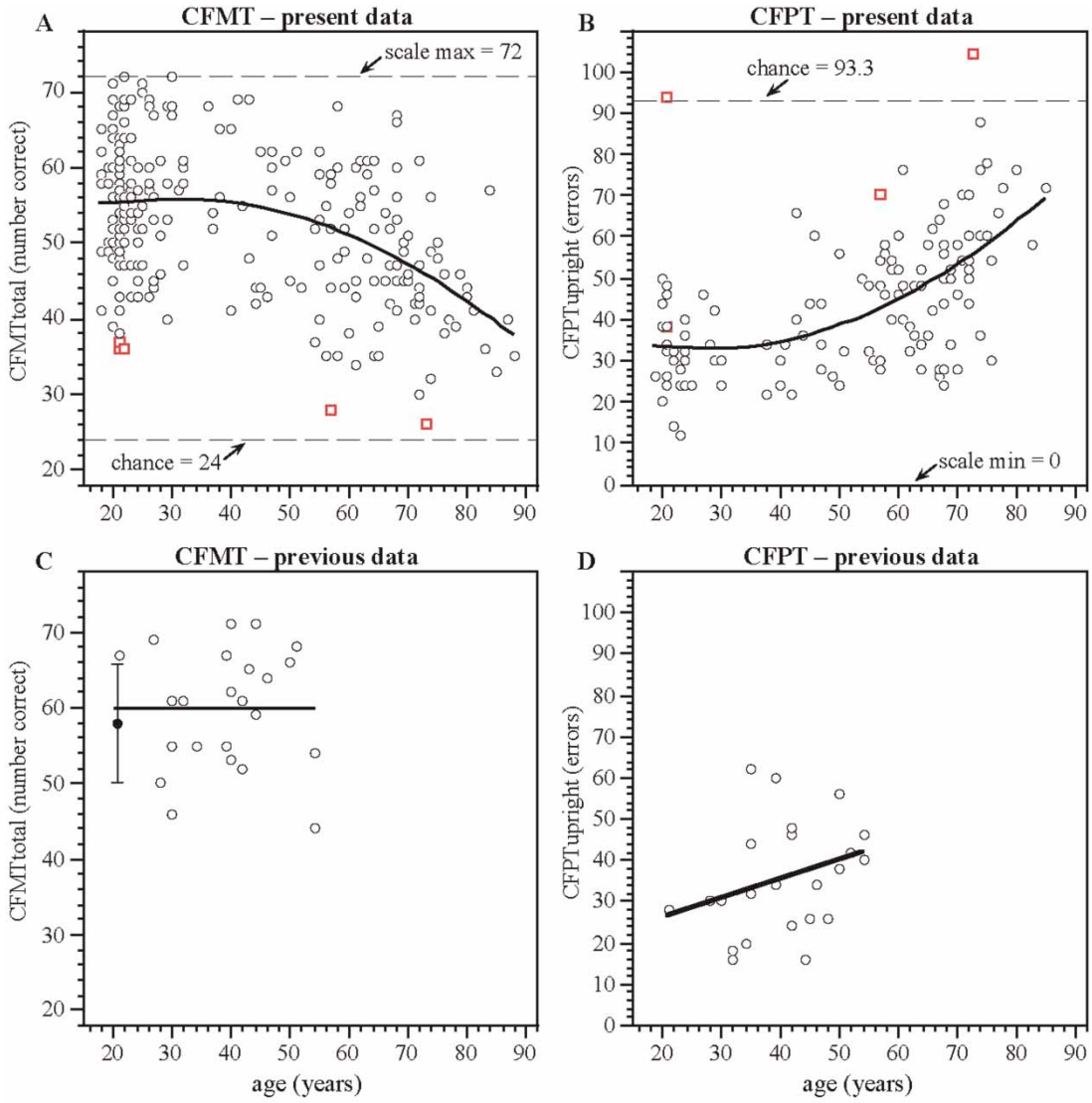

Figure 3. *(A) and (B). Scatterplots of exact age against total correct for the Cambridge Face Memory Test (CFMT, upright faces), and Cambridge Face Perception Test (CFPT) errors for upright face trials, in the present data set. Most individuals are shown as black circles; the 5 individuals diagnosed with prosopagnosia are shown as squares (only 4 of these were tested on the CFPT). The curve fits show the best fitting second-order polynomial relating age to the relevant performance measure; the individuals with prosopagnosia were excluded from the data used for fitting. (C) CFMT (upright faces) for all individuals (*N = 22*) under 55 years previously reported as part of norm groups (in Duchaine, Yovel, Butterworth, & Nakayama, 2006; Duchaine, Yovel, & Nakayama, 2007b; Garrido et al., 2008), plus the mean and standard deviation (shown as filled circle and error bar) for 50 young adults with mean age 20.2 years (from Duchaine & Nakayama, 2006a). (D) CFPT upright face trials for all individuals (*N = 23*) under 55 years previously reported as part of norm groups (in Duchaine, Germine, & Nakayama, 2007a; Garrido et al., 2008). In C and D, curve fits are linear because a second-order polynomial did not improve the fit. We recommend readers view this figure in colour: please see the online issue of the Journal.*

Table 2. *Ageing effects: Control scores for educated Australians on CFMT and CFPT with no Mental Alternation Test exclusion applied*

	Age group (years)														
	Young adult (18–35)			*Early middle age (36–49)*			*50–59*			*60–69*			*70–79*		
	N	*M*	*SD*	*N*	*M*	*SD*	*N*	*M*	*SD*	*N*	*M*	*SD*	*N*	*M*	*SD*
CFMT1	124	17.7	0.7	21	17.9	0.5	24	17.8	0.5	37	17.2	1.5	22	16.2	1.8
CFMT2	124	22.5	4.9	21	22.5	4.9	24	19.8	4.8	37	20.5	4.8	22	17.3	4.1
CFMT3	124	15.2	4	21	15.3	4.9	24	13.1	4.7	37	12.8	4.2	22	10.9	3.5
CFMTtotal	124	55.4	8.5	21	55.6	9.3	24	50.7	9	37	50.5	8.7	22	44.4	7
CFPTupright	36	32.4	9.4	15	36.3	13	19	43.1	11.3	31	45.7	13.4	21	57.4	15.5
CFPTinverted	36	61.4	13.2	15	66	9.9	19	68.2	13.7	31	70.2	12.1	21	79.7	12.9
CFPT inversion effect	36	29	13.2	15	29.7	14.7	19	25.2	14.3	31	24.5	15.2	21	22.3	16.2

Note: CFMT = Cambridge Face Memory Test; CFPT = Cambridge Face Perception Test. Control scores as a function of age group, showing upright-face results for the CFMT (CFMT1 = same images stage; CFMT2 = novel images stage; CFMT3 = novel images with noise stage; CFMTtotal = total summed CFMT score), plus CFPT results for upright faces, inverted faces, and the inversion effect on the CFPT (i.e., inverted minus upright). Scale range for accuracy on CFMTtotal is from chance = 24 to scale maximum = 72; scale range for errors on CFPTupright is from scale minimum = 0 to chance = 93.3. For the CFPT, a higher number equals poorer performance, so cut-offs for prosopagnosia diagnosis would be calculated as 2 standard deviations above the mean. Cut-offs for individual age groups are not provided in this table because the most accurate way to calculate the cut-off for a given age is to use the second-order polynomial fit in conjunction with the standard deviations of the residuals of that fit (see main text for details).

polynomial function significantly improved the fit ($R^2 = .184$, significance of R^2-*change* $p < .005$, $df = 233$). A third-order polynomial produced no further improvement ($R^2 = .186$; significance of R^2-*change* $p > .5$, $df = 232$). The second-order polynomial fit is shown in Figure 3A. The function describing the relationship between age and CFMT was:

$$\text{CFMT}; = -.0056353\,\text{age}^2 + .35142\,\text{age} + 50.288$$

Examining this function visually in Figure 3A suggests that CFMT total remained stable across early middle age, but began to decline noticeably at approximately 50 years of age. This is consistent with our earlier group-based analysis. The lack of age-related decline before late middle age is also consistent with results of previous studies, which have presented data for groups up to mean ages of 46.5 years (see Table 1; Figure 3C also provides a scatterplot of all unique participants in the studies summarized in Table 1 who were aged under 55, given that some individuals had been included as part of norm samples in more than one of the original articles).

From the best fit curve, we then calculated the residual for each participant (i.e., the difference between their score and the estimated score for their age in years). The standard deviation of these residuals was:

$$\text{CFMT } SD \text{ of residuals} = 8.4785$$

This value provides the most reliable estimate of the standard deviation of CFMTtotal scores and the most reliable method of determining a cut-off score for prosopagnosia diagnosis. To do so, one takes the predicted value for a given age from the age-function of best fit, then calculates 2 standard deviations below this predicted value as the diagnosis cut-off. Below, we provide an example of calculating the cut-off, and also the z score, for an imaginary participant with an age of

Table 3. *Ageing effects: Control scores for educated Australians on CFMT and CFPT with Mental Alternation Test exclusion applied*

	Age group (years)											
	Young adult (18–35)			*50–59*			*60–69*			*70–79*		
	N	*M*	*SD*	*N*	*M*	*SD*	*N*	*M*	*SD*	*N*	*M*	*SD*
CFMT1	35	17.8	0.5	10	17.8	0.6	25	17	1.7	19	15.9	1.8
CFMT2	35	22	5.3	10	18.5	4.4	25	19.6	4.9	19	17.3	4
CFMT3	35	14.9	4.6	10	12.7	3.7	25	12.2	3.9	19	10.4	3.1
CFMTtotal	35	54.8	9.5	10	49	7.3	25	48.8	8.8	19	43.6	6.4
CFPTupright	29	31.4	9.1	10	47.8	9.6	23	44.1	14.1	18	58.2	16.5
CFPTinverted	29	59.8	13.1	10	71.2	16.1	23	71.4	12.8	18	80	13.9
CFPT inversion effect	29	28.4	13.3	10	23.4	14.6	23	27.3	14.7	18	21.8	17.1

Note: CFMT = Cambridge Face Memory Test; CFPT = Cambridge Face Perception Test. Participants for whom the Mental Alternation Test was not measured do not appear in this table.

54 years and a CFMT score of 37:

$$\text{age} = 54\text{years; actual CFMT score} = 37$$

$$\text{estimated CFMTscore} = -.005635 \times 54^2 + .3514 \times 54 + 50.29 = 52.832$$

$$\text{diagnosis cut-off} = \text{estimated CFMT score} - 2 \times SD = 52.832 - 2 \times 8.4785 = 35.875$$

$$\text{z score} = (\text{actual CFMT score} - \text{estimated CFMT score})/SD = (37.000 - 52.832)/8.4785 = -1.87$$

NB. To obtain *z* score correct to 2 decimal places, it is necessary to begin with 5 significant figures in parameter estimates. This participant would not be diagnosed as prosopagnosic.

Finally, we used the curve fit and associated standard deviation to make concrete the importance of taking into account age effects when making prosopagnosia diagnoses. Table 4 indicates the number and percentage of the entire sample who would be diagnosed as having prosopagnosia based on (a) performance 2 standard deviations poorer than their own age, and (b) performance 2 standard deviations poorer than a 21-year-old. Results indicate that, were researchers and clinicians to ignore the ageing effects and inappropriately use the young adult norms for all age groups, this would have resulted in 11 older individuals in our sample falsely diagnosed as having prosopagnosia (out of a total *N* of 112 individuals aged 35–88 years tested on the CFMT), with apparent prosopagnosia rates reaching as high as 22% of the population in the 70–88 years age range.

Cambridge Face Perception Test. The effects of ageing on the CFPT were even more dramatic than those on the CFMT. Tables 2 (no MAT exclusion applied) and 3 (MAT exclusion applied) give CFPT results by age group, for upright faces, inverted faces, and the inversion effect (i.e., difference between upright and inverted). Figure 3B shows the full scatterplot of all participants against age, for CFPTupright. Note upright faces is the relevant orientation for diagnosing prosopagnosia.

Comparing the five age groups, ANOVA revealed a highly significant effect of age on mean performance, $F(4, 117) = 15.111$, $MSE = 153.295$, $p < .001$, and there was no change in variance, Levene's test: $F(4, 117) = 1.763$, $p > .1$. The *t* tests contrasting young adults to successively older age groups revealed no significant change in mean performance between young adults and early middle age (35–49), $t(49) = 1.199$, $p > .2$. As with the CFMT, however, even participants in their 50s (50–59) performed

Table 4. *Ageing effects on rates of "prosopagnosia" diagnosis*

	CFMT total				*CFPT upright*			
	Compared to own age		*Compared to 21 yrs*		*Compared to own age*		*Compared to 21 yrs*	
Age group (years)	*N*	*%*	*N*	*%*	*N*	*%*	*N*	*%*
18–35	3	2.4	3	2.4	1	2.7	1	2.7
36–49	0	0	0	0	1	6.7	2	13.3
50–59	1	4.0	4	16.0	1	5.0	2	10.0
60–69	0	0	4	10.8	1	3.2	7	22.6
70+	1	3.2	7	22.6	2	8.0	15	60.0

Note: CFMT = Cambridge Face Memory Test; CFPT = Cambridge Face Perception Test. For CMFT and CFPT, table shows number of individuals (and percentage relative to number tested in each age group) defined to have prosopagnosia (performance more than 2 *SD*s poorer than mean) based on their own age norm versus a 21-year-old norm.

significantly more poorly (i.e., made a higher number of CFPT errors) than young adults, $t(53) = 3.733$, $p < .001$.

When curves were fitted to the scatterplot of age against CFPTupright (Figure 3B), a linear-only fit explained significant variance ($R^2 = .334$, $p < .001$, $df = 123$) but a second-order polynomial fit was significantly better ($R^2 = .379$, significance of R^2-*change* $p < .005$, $df = 122$), and a third-order polynomial produced no further improvement ($R^2 = .380$, $p > .6$, $df = 121$). The formula for the best fitting second-order polynomial was:

$$\text{CFPTupright} = 0.010953\,\text{age}^2 - .59044\,\text{age} + 40.260$$

and the standard deviation of the residuals was:

$$\text{CFPTupright } SD \text{ of residuals} = 12.143.$$

Examining this function visually (Figure 3B) indicated that age-related decline in performance (i.e., increase in the CFPT error score) had begun by age 50. Indeed, there are reasons to suggest that perhaps decline might begin even earlier: In addition to Figure 3B's suggestion of decline in early middle age, a scatterplot of CFPT data from previous studies (Figure 3D; also see Table 1) shows the same tendency. Combining the scores from participants in Figures 3B and 3D to increase sample size indicated not only a significant increase in CFPT scores across the 30–55 age range ($r = .306$, $N = 47$, $p = .036$), but also some suggestion of an increase even across the 30–50 age range ($r = .257$, $N = 38$, $p = .119$). Note that earlier decline on the CFPT than the CFMT is theoretically feasible due to the fact that the CFPT is a speeded task, and reaction time is slower even in 40-year-olds than in 20-year-olds (e.g., Anstey, Dear, Christensen, & Jorm, 2005).

Regarding diagnosis cut-offs, three findings emerged. First, the effects of ignoring age and inappropriately using young adult norms for diagnosis would be even more detrimental for the CFPT than for the CFMT. As Table 4 shows, using 21-year-old norms for all participants would result in 27 extra false diagnoses of prosopagnosia (out of a total *N* of 89 participants in the 35–88 years age range), and apparent prosopagnosia rates would reach as high as 60% in the 70–88 years age group.

Second, at least in this Australian sample, the CFPT is unsuitable for diagnosis in individuals over the age of approximately 80 years. The cut-off score for diagnosis in an 85-year-old reaches chance (cut-off = 93.5, chance = 93.3). Although individual scores may be poorer than chance, such scores cannot be taken to be a meaningful indication of face perception ability. Also, the cut-off score at 80 years (87.4) is close enough to chance to mean there is little room for

a participant to produce a score reliably worse than cut-off. Note, however, that at 75 years the cut-off (81.9) is clearly better than chance meaning that the test is still useful at this age.

Third, it seems that for the CFPT (much more so than for the CFMT), estimates of the standard deviation of particular groups, and thus the diagnosis cut-off for that group, are quite unreliable with *N*s of 15–36 in the control sample. This is shown in both Table 2 (present Australian sample) and Table 1 (previous studies), where standard deviation for the CFPT varies quite noticeably and apparently randomly across groups. This information is of practical importance because currently, in the absence of large published norm samples covering all ages, it is not uncommon for researchers testing a suspected prosopagnosic to test an age-matched control group of only 10–30 participants against which diagnosis is made, particularly where the target individual is not a young adult. Our results thus argue that, for the CFPT, this procedure is likely to produce quite unreliable diagnoses, unless the participant is clearly very far from the cut-off (e.g., $z < -3$).

We now turn briefly to the CFPT results for inverted faces (CFPTinverted). As with upright faces, performance progressively worsened with age in a nonlinear fashion (see Table 2). Regression revealed both a linear effect of age ($R^2 = .198$, $df = 123$, $p < .001$) and an improvement in fit with adding a quadratic component ($R^2 = .227$; R^2-*change* $= .028$, $df = 122$, $p = .036$). The equation for the best fitting second-order polynomial was:

$$\text{CFPTinverted} = .00799246\,\text{age}^2 - .44298\,\text{age} + 67.506$$

with a standard deviation of the residuals of:

$$\text{CFPTinverted } SD \text{ of residuals} = 12.351.$$

Finally, the size of the *inversion effect*—the amount by which inverted performance was poorer than upright—did not change with age: When participants older than 75 years were excluded (due to mean inverted performance approaching chance), there was no linear ($p = .116$) or nonlinear ($p = .908$) effect of age. By taking the mean and standard deviation of the scores for all participants aged 75 or younger ($N = 118$), we then obtained:

CFPT inversion effect: $M = 26.432$; $SD = 14.414$

This confirms the theoretically expected presence of a large inversion effect on the CFPT and moreover argues that *holistic processing* (as is associated with large inversion effects for faces, e.g., for review see Robbins & McKone, 2007) does not weaken with age.

Ageing: Implications for diagnosis. Both CFMT and CFPT were strongly affected by ageing. For the CFMT, results imply that young adult norms can be used for diagnosing prosopagnosia only up until approximately 50 years of age, and that age-specific norms are required from 50 years onwards. For the CFPT, results clearly indicate a need to use age-specific norms beyond 50 years; however, they also suggest some caution about using young adult norms even for early-middle-aged participants (e.g., 35–49), given the suggestion of some small decline even over this age range. This implies that, for example, if only young adult norms were available, and a 45-year-old participant produced a *z* score relative to these norms only just outside the normal range (e.g., $z = -2.1$), then a diagnosis of prosopagnosic-level performance should be considered with caution.

In other findings, we also found that (a) the CFPT cannot be used to diagnose prosopagnosia in Australian individuals of 80 years or older due to floor effects, (b) caution should be applied to interpreting CFPT scores within several points either side of the notional cut-off value if a norm sample for a group is relatively small (e.g., in the 15–30 participants range), and (c) the size of the inversion effect on the CFPT does not change with age.

Finally, given that we found no change in variance with age on either the CFMT or the CFPT, our results argue that for both tasks, if resources

are not available to test a large control sample (e.g., 50+) of a single age group, then a reliable estimate of the standard deviation (and thus diagnosis cut-offs) can alternatively be determined from a large sample distributed across age using the curve-fitting procedures we have outlined here.

Sex: Australian sample

Before analysing the effects of sex, we tested whether sex interacted with ageing. Breaking the data into the five age groups used in the previous section, two-way ANOVAs revealed no interaction between sex and age group for either the CFMT, $F(4, 218) < 1$, $MSE = 72.606$, $p > .9$, or the CFPT, $F(4, 112) = 1.183$, $MSE = 149.75$, $p > .3$.

Cambridge Face Memory Test. In the one previous study reporting sex effects on norms for the CFMT, Duchaine and Nakayama (2006a) found that, in their sample of 50 young adults, women showed a small 2.5-point advantage over men in mean performance (male $M = 56.4$, female $M = 58.9$). This difference was not statistically significant. Also, given that the women showed a slightly larger standard deviation (male $SD = 7.3$, female $SD = 8.3$), the cut-off score for prosopagnosia was very similar for each sex (male cut-off = 41.8, female cut-off = 42.3).

In our sample of 124 Australian young adults (51 male and 73 female), an almost identical pattern emerged. Young adult women showed a 2.7-point advantage over men in mean CFMTtotal performance (male $M = 53.8$, female $M = 56.5$), a difference that only approached significance, $t(122) = 1.414$, $p = .081$. In combination with a slightly larger standard deviation for females (male $SD = 8.0$, female $SD = 8.9$; Levene's test for differences in variances $F < 1$, $p > .4$), there was again very little sex difference in diagnosis cut-off for prosopagnosia (male cut-off = 37.8, female cut-off = 38.7) for young adults.

Figure 4A shows a scatterplot with separate second-order polynomial curve fit for males and females. The curve fit results for males ($N = 96$)

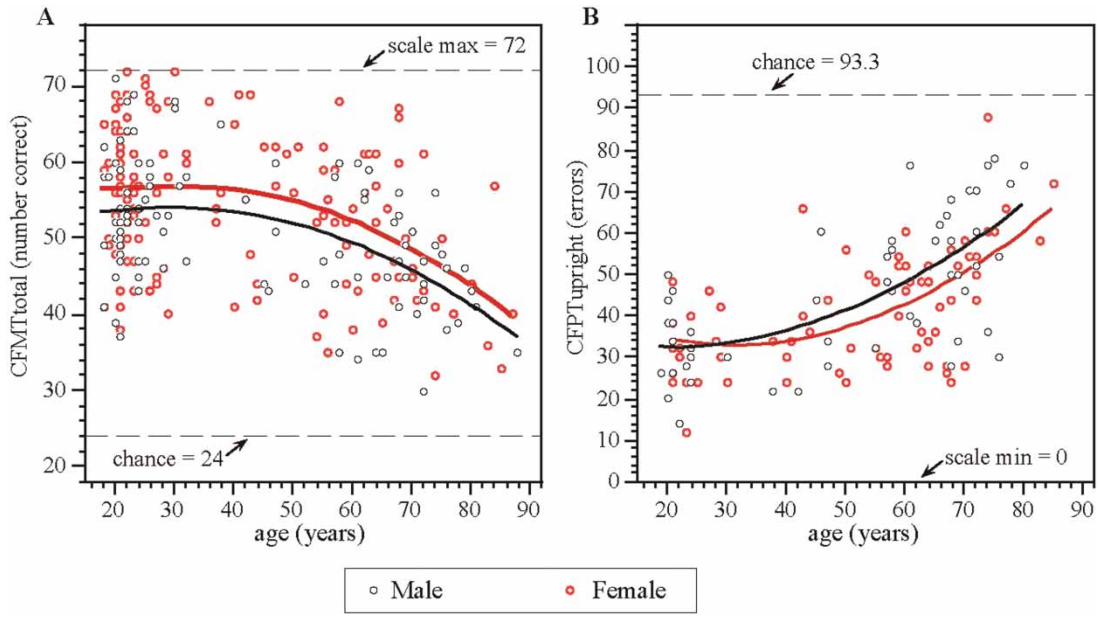

Figure 4. *(A) Cambridge Face Memory Test (CFMT) and (B) Cambridge Face Perception Test (CFPT) scores from our Australian sample with separate curve fits for males (black) and females (red). Data from the 5 prosopagnosics were not used for the fits. We recommend readers view this figure in colour: please see the online issue of the Journal.*

were:

$$\text{CFMT males} = -.0049785\,\text{age}^2 + .29594\,\text{age} + 49.511;$$
$$SD\ \text{of residuals} = 7.7219$$

and for females ($N = 140$) were:

$$\text{CFMT females} = -.0053762\,\text{age}^2 + .32825\,\text{age} + 51.807;$$
$$SD\ \text{of residuals} = 9.0342.$$

Results indicate that the slightly better mean performance in females than in males was maintained across all ages. Unsurprisingly, then, a significant female advantage was obtained when power was increased by combining participants of all ages (total $N = 236$) in a regression testing the independent effects of sex, taking out the effects of age, on CFMTtotal performance, beta = .161, $p < .01$ (note that it was important to use age as a covariate because there was a large ageing effect and uneven age distribution across sexes).

The fits also confirm that, while the *mean* is affected by sex, the *cut-off* is not, even at older ages. For example, estimated CFMT score for a 70-year-old female was 48.4 and for a 70-year-old male was 45.8 (giving a female advantage of 2.6) but, given that the standard deviation of the fit residuals was slightly larger in women (9.0) than in men (7.7), diagnosis cut-off for a 70-year-old was the same for both sexes (female = 30.4, male = 30.4).

Overall, results showed a small (approximately 2.6-point) sex difference in mean CFMT performance, favouring women, consistent with Duchaine and Nakayama's (2006a) earlier finding (and in contrast to results suggesting sex differences in face memory occur only for female faces; e.g., Lewin & Herlitz, 2002). Critically, however, there was no important sex difference in diagnosis cut-off for prosopagnosia on the CFMT.

Cambridge Face Perception Test. No previous studies have investigated sex effects for the CFPT. Here, we report sex effects with data combined across Australian participants of all ages (sample sizes were too small for reliable analysis of sex effects in any one age group). Across the full sample ($N = 125$), regression revealed an effect of sex, taking out the effect of age, that was close to significant, beta = .141, $p = .055$. Given that this sex effect was in fact closer to significance than was the CMFT sex difference with the same number of participants ($N = 124$, see previous subsection), and that it trended in the same direction as that for the CFMT (i.e., a female advantage), it seems quite probable that a small female advantage on mean performance for the CFPT is genuine and would be significant with a larger sample size.

Figure 4B shows a scatterplot with separate second-order polynomial curve fit for males and females. The curve fit results for males ($N = 52$) were:

$$\text{CFPTupright males} = .0090988\,\text{age}^2 - .33712\,\text{age} + 35.029;$$
$$SD\ \text{of residuals} = 12.763$$

and for females ($N = 73$) were:

$$\text{CFPTupright females} = .0110408\,\text{age}^2 - .66915\,\text{age} + 42.541;$$
$$SD\ \text{of residuals} = 11.720.$$

The important results here were that (a) the mean female advantage tended to become more apparent in older than younger participants (although recall that the interaction between sex and age was nowhere near statistically significant); and that (b) in contrast to the CFMT, the standard deviation of the residuals was slightly smaller in women (11.7) than in men (12.8). In combination, these two findings meant that, although diagnosis cut-off was not meaningfully influenced by sex in young adults (for age = 21, male cut-off = 57.5, female cut-off = 56.8), important differences appeared at older age groups, with a 4.7-point female advantage in cut-off at 40 years (male cut-off = 61.6, female cut-off = 56.9) increasing to an 8.3-point female advantage

in cut-off at 70 years (male cut-off = 81.5, female cut-off = 73.2).

Overall, our data suggested that there may well be a small mean advantage to females on the CFPT, and that diagnosis cut-off needs to take sex into account in middle-aged and older participants.

Sex: Implications for diagnosis. For mean task performance, results demonstrate a small but consistent female advantage on the CFMT and suggest a similar small female advantage being present on the CFPT. Regarding the cut-off for prosopagnosia diagnosis, the results differ for the two tasks.

On the CFMT, women had better mean performance but also tended to have slightly higher variability. Thus, although the results demonstrate a small significant sex difference in face memory, they imply that, where the aim is to diagnose prosopagnosia on the basis of a cut-off of 2 standard deviations poorer than the mean, sex can safely be ignored. That is, potential prosopagnosics can be CFMT compared to norms derived from mixed-sex samples, and there is no need to develop control samples matched to the sex of the individual in question.

For the CFPT, in contrast, women's small mean advantage over men was combined with slightly lower variability. This produced noticeable sex differences in prosopagnosia cut-off in middle-aged and older participants. Thus, for the CFPT, it appears that, in participants who are older than approximately 35 years, it is necessary to use control samples matched to the sex of a potential prosopagnosic.

Measures related to general intelligence: Australian sample

No published studies have addressed whether CMFT and CFPT performances are associated with general intelligence or interrelated constructs such as academic ability or education level. Even regarding other tests of face recognition, there are no published results for the correlation between face processing and full-scale IQ. Herlitz and Yonker (2002) found that estimated IQ (partial Wechsler Adult Intelligence Scale–Revised, WAIS–R, score) was independent of face memory for novel faces in women ($r = .001$). There was a correlation in men ($r = .42$), which appeared to come primarily from lower-than-average-IQ groups (i.e., 60–80 IQ and 81–100 IQ men performed more poorly than all other IQ/sex groups). Dobson and Rust (1994) reported that face memory was substantially poorer in teenagers with mental retardation (mean IQ = 64.5) than in age-matched controls. A conference abstract has reported no association between face recognition and normal-range intelligence assessed via the Cattell Culture Fair Test (Jeffery & Anderson, 2004). Overall, these results suggest that, although low IQ might be associated with poor face memory (in men), in the upper half of the IQ distribution (and across the full IQ range in women), face memory might be independent of general intelligence.

In the present study, data were collected on a number of measures that bear some relationship to general intelligence. Although these are limited by not measuring IQ directly, and the fact that our sample was more educated than average, they provide the first evidence of any sort regarding the relationship between intelligence-related measures and specifically the CFMT or CFPT. They are also important for our theoretical interpretation of country-of-origin effects in the next section.

Cambridge Face Memory Test. Evidence from four different measures argued that there is no association between intelligence and CFMT performance within our generally well-educated sample. Regarding *number of years of education*,[3] multiple regression of CMFTtotal against years of education, age, and sex showed no independent effect of years of education, beta = −.005, $p > .9$ ($N = 231$). Similarly, multiple regression of

[3] Of course, education level in our sample may be only rather loosely related to intelligence, partly because many of the young adults were only beginning university degrees, and partly because for the older groups (particularly the over-65s) university education in Australia was less widely available than today, particularly for women.

CMFTtotal against *Mental Alternation Test (MAT)* score, age, and sex showed no independent effect of MAT, beta = −.028, $p > .7$ ($N = 104$). And, multiple regression of CMFTtotal against a *verbal memory* score (memory for a list of unrelated words, tested for participants in Study 1; see Bowles, 2007, for method), age, and sex showed no independent effect of verbal memory, beta = .110, $p > .3$ ($N = 103$).

A final analysis compared a highly academically selected subset of young adults to the rest of the young adults. The former group were part of an ANU summer scholarship programme, in which scholars visit the ANU over the summer following third-year undergraduate or Honours years to undertake two months of research in an academic's laboratory. Entry to the programme is highly competitive: ANU is the highest ranked university for research in Australia/New Zealand; students from all universities across Australia and New Zealand are eligible to apply; and successful applicants usually have undergraduate grades comprising primarily High Distinctions and Distinctions and either already hold or go on to complete a First Class Honours degree. There are thus good reasons to presume that the mean IQ of this group would be noticeably higher than that of the rest of the young adults. Table 5 shows mean, standard deviation, and prosopagnosia cut-off values for the summer scholars and the rest of the young adults (excluding 3 enrolled in PhD programmes). As can be seen, there were no differences in mean performance, $t(113) = 0.455$, $p > .6$, or in variance (Levene's test $F = 1.182$, $p > .2$), or in the resulting score below which prosopagnosia would be diagnosed (37.7 for summer scholars, 37.9 for the rest of the young adults.

Table 5. *Effects of intelligence-related variables*

	N	*Mean*	*SD*	*Cut-off*
Summer scholars	33	55.73	9.04	37.65
Rest of young adults	82	54.91	8.52	37.87

Note: CFMTtotal performance for a highly academically selected subgroup of the young adults ("summer scholars") and for the rest of the Australian young adults. CFMT = Cambridge Face Memory Test. Cut-off: −2 standard deviations.

Cambridge Face Perception Test. Results suggested that CFPT performance may possibly have some relationship to general intelligence, even in an educated sample, although the evidence for this was rather weak. For CFPTupright, a multiple regression including verbal memory, age, and sex as predictors found a significant effect of verbal memory, beta = −.198, $p < .05$ ($N = 87$). The effect was in the direction that individuals with higher verbal memory scores (associated with higher IQ) obtained lower (i.e., better) CFPT scores. However, similar multiple regressions produced age- and sex-independent contributions of years of education and MAT that were not even close to significant: education, beta = −.055, $p > .4$ ($N = 123$); MAT, beta = −.101, $p = .261$ ($N = 88$).

Note the summer scholars were not tested on the CFPT.

Intelligence: Implications for diagnosis. It will be important in future studies to properly assess the relationship between full-scale IQ and CFMT/CFPT performance (and, more generally, between IQ and other measures of face processing). Our present analysis here is limited by the facts that the predictor variables investigated were only loose correlates of IQ; we could address only mean effects of these variables, rather than effects specifically on prosopagnosia cut-off; and our sample was more educated than the general population.

Still, the present results are consistent with those of previous studies (Herlitz & Yonker, 2002; Jeffery & Anderson, 2004) in suggesting that, within the upper half of the IQ distribution, face memory, which includes the CFMT, is independent of general intelligence (although again note that we cannot rule out a relationship being present in the lower half of the IQ range). This is information is of some practical importance, given that probably a relatively high proportion of developmental prosopagnosics reported in the literature or currently being studied are of higher than average education/intelligence: It is this

group that is more likely to seek out treatment, to become aware of the symptoms of prosopagnosia through reading newspaper articles or watching television science programmes, or to be discovered accidentally through mass testing of face tasks on undergraduate psychology students.

Turning to the CFPT, it is harder to rule out a relationship with intelligence even in an educated sample, given that we found a significant association between verbal memory and the CFPT. Theoretically, an association with intelligence could be more apparent in the CPFT than in the CFMT because the former, unlike the latter, is a time-limited task, and higher intelligence is associated with faster general processing speed (even within the upper half of the IQ range). In terms of diagnosing prosopagnosia, the possible intelligence effect is a finding that clinicians and researchers might wish to keep in mind if, for example, a client performs normally on the CFMT but would be suggested to be prosopagnosic from CFPT results.

Country of origin (ethnic match): Australian and Israeli samples

No previous studies have investigated country-of-origin effects on the CFMT or CFPT. In the only previous study that is directly relevant—in that it used faces from the same database as that for the CFMT and CFPT—Gilchrist and McKone (2003) found that Australian participants' memory was more strongly focused on local features for the Boston-area faces than on those for Australian faces. Possibly this occurred because it is local features that stand out most obviously as unusual in other-ethnicity faces.

Cambridge Face Memory Test. To examine the effects of country of origin, we started by comparing our Australian sample to the published USA norms for the CFMT. We considered only our young adult group because previous studies of USA-based (Boston-area) samples do not include groups matched in age to our three older Australian groups. Table 6 shows results. Mean CFMTtotal scores were significantly poorer for our Australian young adults than for the American young adults of Duchaine and Nakayama (2006a), $t(167) = 2.41$, $p < .02$. The 2.7-point difference in means, in combination with a somewhat larger variance for the Australian sample ($SD = 8.6$) than for the USA sample ($SD = 7.9$), produced a 4-point difference in diagnosis cut-offs (38 vs. 42). This difference is of strong practical importance: A total of 2.6% of the Australian sample fell below the Australian CFMT cut-off, but inappropriately comparing Australian individuals to the Duchaine and Nakayama (2006a) norms gives 8.6% below the cut-off, implying 7 false diagnoses of prosopagnosia out of 119 participants.

Results from our Israeli sample (Study 5) were quite different. In this case, both mean and diagnosis cut-off scores for CFMTtotal were almost identical to the original USA norms (see Table 6).

Cambridge Face Perception Test. There was an insufficient number of participants available in

Table 6. *Effects of participant–stimulus ethnic match*

Country	*N*	*Age (years) M* (range)	*Sex (% female)*	*M* (/72)	*SD*	*Cut-off*
USA (Boston)	50	20.3	50	57.9	7.9	42
Israel	49	22.0 (18–31)	69	57.6	8.4	41
Australia/New Zealand	117	23.0 (18–32)	59	55.2	8.6	38
Germany	153	24.0 (18–35)	53	52	8.5	35

Note: Young adult CFMTtotal performance as a function of country of origin, including Australia and New Zealand (present study), Israel (present study), Boston area in USA (Duchaine & Nakayama, 2006a), Germany (Grit Herzmann, personal communication, June 10, 2009, based on data from study of Herzmann et al., 2008). Performance is better in the countries where participant–stimulus ethnic match is stronger (USA, Israel) than where it is weaker (Australia, Germany). CFMT = Cambridge Face Memory Test. Cut-off: −2 standard deviations.

overlapping age groups to make it worthwhile examining country-of-origin effects for the CFPT.

Country of origin: Theoretical interpretation. Before we can conclude from our CFMT results that it is necessary to test country-matched control participants for potential prosopagnosics, we must rule out other interpretations of the difference in CFMT performance between our present sample and previous studies. The only obvious alternative is that the USA sample was likely to have been particularly high in intelligence. Duchaine and Nakayama's (2006a) young adults control participants were mostly students from the Boston area and included a high proportion of students from Harvard University. However, we argued in the previous section that intelligence in a generally well-educated sample does not affect CFMT performance. Moreover, there is no reason to expect that our highly academically Australian selected "summer scholars" (who would fall in the academic top 1%) would be noticeably less intelligent than Harvard undergraduates, yet the diagnosis cut-off for this group remains 4 points poorer than that for the USA sample.

Overall, the much more likely interpretation of the group differences is that they are, indeed, due to country of origin, and specifically that the relatively poor performance in Australians arises from mismatch in ethnic background between the face stimuli (Boston-area students) and either the participant themselves or the typical "diet" of Caucasian face subtypes to which the participant has been exposed over the course of his or her lifetime. Several pieces of information support this interpretation.

First, demographic data indicate that there are quite different patterns of ancestry between Australians and Boston-area students: These can be summarized by saying that typical Caucasian Australians are more likely to be British or Northern European in appearance, while typical Caucasian Boston students are relatively more likely to be Jewish or Southern European in appearance. In Federal Government censuses, the proportion of responses implying British Isles ancestry (English, Scottish, Irish, Welsh) is high in Australia (71% in Canberra, 55% in Sydney, from 2006 Australian census) and much lower in Massachusetts (33%, from 2000 US census). The reverse is true for Italian ancestry (11.8% Massachusetts, 2.6% Canberra, 3.4% Sydney). And, the proportion of individuals who are Jewish is dramatically higher amongst Harvard undergraduates (approximately 34% of Caucasians; data for 2000s taken from Hillel, the Journal of Blacks in Higher Education, The Harvard Crimson, and the Harvard University Factbook), than amongst Australians (<0.5% of Caucasians, 2006 census; also note proportion of people with Jewish background remains very low even amongst Australian university students).

Second, there was anecdotal evidence that at least some Australian participants perceived the mismatch in face ethnicity. Spontaneous comments from participants included that the CFMT/CFPT faces looked "Mediterranean"; others described them as "Middle Eastern", and 1 participant then went so far as to describe them as looking "like terrorists".

Third, in Study 2, a separate test was included using Australian-photographed faces (the "Canberra faces task", not reported here). Some participants in that study commented spontaneously that the Canberra faces appeared "a lot more normal" than the faces in the CFMT. This difference in perception can only be attributed to the ethnic background of the faces, given that there were no major differences in the photograph format (e.g., hair was removed from both).

Fourth, these ideas predict that participants from a country with a high Jewish (or southern European) population should show good CFMT performance, while participants from a Northern European country should, like Australians, show poor performance—and Table 6 indicates that this is exactly what happens. Caucasian Israeli participants (present Study 5) perform as well as the original USA participants. And Caucasian Germans (data kindly provided by Grit Herzmann, from study of Herzmann et al., 2008) perform very poorly—in fact, even more poorly than Australians.

Overall, results make a strong case that CFMT scores are better where there is stronger ethnic match between the face stimuli and participants (USA, Israel) and poorer where there is weaker

ethnic match (Australia, Germany). Possible mechanisms for this effect include (a) social outgroup categorization, which is known to lower memory performance for faces (Bernstein et al., 2007), and/or (b) poorer tuning of face-space dimensions for less-familiar Caucasian subtypes (cf. Valentine, 1991). Importantly, note that to explain our own findings it is not necessary to propose that poor face-space tuning or outgroup categorization would apply for every Australian participant on every trial; our findings of a lower mean and larger standard deviation in Australians than in Americans can be accounted for even if these problems apply only in some cases.

Country of origin: Implications for diagnosis. The practical implication of the present findings is that ethnic background of the lifetime "diet" of faces to which participants have been exposed cannot be ignored when using the CFMT to diagnose prosopagnosia, even within Caucasians. Our results show that potential prosopagnosics must be compared to norms derived from controls who have experienced a similar mix of faces as the prosopagnosic, and they imply that is important to develop CFMT norms for different countries (or regions) based on those countries' particular patterns of heritage from different places in Europe.

Reliability of CFMT and CFPT and relationship between tasks

Two important questions relevant to diagnosis using the CFMT and CFPT are reliability of the individual measures and the extent to which both tasks address the same underlying ability. Our data provide some useful new information on these topics.

Regarding reliability of the CFMT, Duchaine and Nakayama (2006a) correlated accuracy on Stage 2 of the test (novel images) with accuracy on Stage 3 (novel images in noise; Stage 1 was left out of this analysis because normal participants perform at ceiling). Their correlation, in 50 young adults, was $r = .74$. The present results agree: Correlation between Stage 2 and Stage 3 was $r = .751$, $p < .001$, $N = 124$ in the young adults, and $r = .734$, $p < .001$, $N = 236$ in the entire sample. (The prosopagnosics were excluded from this analysis to allow closest comparison with Duchaine & Nakayama, who tested only nonprosopagnosic individuals.) Regarding the standard internal consistency measure Cronbach's alpha, Herzmann et al. (2008) have provided the only previous report (in German participants, CFMTtotal alpha = .83). In the present study, internal reliability was very good: alpha = .887 based on all 72 trials and all age groups ($N = 224$); alpha = .877 based on the 54 trials from Stages 2 and 3 combined and individuals from all age groups ($N = 224$); and alpha = .875 based on all 72 trials and young adults only ($N = 126$).

Turning to the CFPT, no previous studies have provided any type of reliability analysis. Our results showed that, for upright faces, internal reliability was lower than that for the CFMT but still quite good: Cronbach's alpha = .743 based on the entire sample ($N = 126$); and alpha = .732 for young adults only ($N = 37$). For inverted faces, reliability was noticeably lower: alpha = .500 ($N = 126$). Presumably, the lower reliability for inverted faces reflects the fact that participants usually find it difficult to perceive identity-related information in inverted faces, which would result in rather random similarity ordering (except for test faces very obviously different from the target face).[4]

The extent to which the CFMT and CFPT assess the same underlying ability can be addressed by considering the correlation between the two tasks, in the light of the reliability measures. Correlation between

[4] Supporting evidence for this interpretation was obtained. For each individual target face in the CFPT, we calculated the mean error score (i.e., averaging across participants) and correlated this with the "default" error score for that face—namely, the score that would be obtained if participants made no changes at all to the initial left–right ordering of the faces on each trial. If CFPT participants' perception of facial identity is reliable, we would expect the correlation to be zero; that is, starting-point ordering should not matter. Results showed the expected independence of starting and final order for upright faces ($r = -.19$), but there was some evidence of a relationship for inverted faces ($r = .43$), suggesting that participants to some extent tended to leave the inverted faces in their original order.

CMFTtotal and CFPTupright was $r = -.607$, $N = 124$, $p < .001$ in the entire sample, and $r = -.619$, $N = 36$, $p < .001$ in young adults. (Note that the correlations are negative because CFMT uses accuracy while CFPT uses errors.) This correlation was high relative to the upper bound correlation (the product of the square root of the reliability scores) of $r = .807$. Thus, results imply that the CFMT and CFPT assess abilities that are largely overlapping: However, the fact that the correlation is not at maximum carries the theoretical implication that face perception and face memory are partially dissociable skills.

Reliability and interrelationship: Implications for diagnosis. Internal reliability of the CFMT for upright faces meets standard requirements for clinical tests (Cronbach's alpha $> .85$; Aiken, 2003) where the intention is to use an individual participant's score, as is the case when diagnosing prosopagnosia. Internal reliability of the CFPT for upright faces did not achieve this criterion level (possibly because there are only eight trials for upright faces), although the test is reliable enough to be useful when reporting scores averaged across groups of participants and is also still of value in diagnosing extreme single cases (i.e., although the interpretation of a z score very close to the cut-off is open to question, an extreme z score such as -3 would clearly be indicative of prosopagnosia; also note that a small z score such as -1.1 would be clearly indicative of normal performance). External validity of the CFMT and CFPT was supported by the strong correlation between the two tasks (and also, of course, by the large inversion effects shown by each task and by the ability of the tasks to diagnose known cases of prosopagnosia; Duchaine et al., 2007a; Duchaine & Nakayama, 2006a).

Overall, the finding that the reliability of the CFMT (.89) was higher than that of the CFPT (.74) is consistent with our earlier observations of somewhat unstable standard deviation estimates for the CFPT with small-to-medium *N*s in a group. Also, the correlation between tasks indicated that the two tasks do not tap identical cognitive functions (unsurprisingly, given that one tests face perception, and the other face memory). We thus suggest that there are clear advantages to conducting both tests, partly because reliability for either is not perfect, and partly to test for a specific diagnosis of *prosopamnesia*—that is, the ability to perceive identity-related information in novel faces but not to store or consciously retrieve a memory record of facial information (while showing normal memory for other material; cf. Tippett, Miller, & Farah, 2000; Williams, Berberovic, & Mattingley, 2007).

Ability of our norms to diagnose cases of suspected prosopagnosia

It has previously been shown that, in US samples, the CFMT and CFPT are both able to diagnose individuals known to be prosopagnosic from other information (e.g., family member recognition and famous face tests; Duchaine et al., 2007a; Duchaine & Nakayama, 2006a). Given that the present results produced rather different norms for Australian participants, we examined CFMT and CFPTupright scores for a group of 7 Australian individuals—not included in our main "unselected" sample—with suspected prosopagnosia. All these individuals had contacted the MACCS centre because they suspected they might have face recognition problems. All but one performed poorly at famous face recognition (Table 7). A total of 6 of the 7 also self-reported face recognition difficulties in response to specific focused questions, most commonly ($N = 5$) trouble following films. All had average or above-average IQ (Raven's matrices). None scored highly (above 32) on autistic traits (scores ranged from 3–30 on the Autistic-Spectrum Quotient of Baron-Cohen, Wheelwright, Skinner, Martin, & Clubley, 2001). All had normal visual acuity at low contrast (FACT test) and normal colour vision (100% on Ishihara plates).

Of these 7 suspected prosopagnosics, Table 7 shows that 4 are clearly diagnosed as prosopagnosic by the CFMT (i.e., SP_F40, SP_M53, SP_F21, and SP_M60 all have CFMTtotal z scores < -2.0), and another 2 have CFMT performance that is noticeably poor ($z < -1.7$)

Table 7. *Diagnosing suspected prosopagnosia*

						CFMT		*CFPT*	
Participant	*Age (years)*	*Famous faces z*	*CFMT1*	*CFMT2*	*CFMT3*	*Total*	*z*	*Upright*	*Up z*
SP_F40	40	−2.43	16	13	8	37	−2.16	68	−2.95
SP_M53	53	−2.46	12	11	7	30	−2.72	36	0.53
SP_F21	21	−2.56	16	9	11	36	−2.26	42	−0.74
SP_M60	60	−1.04	14	10	6	30	−2.49	48	−0.03
SP_F37	37	−2.10	15	17	9	41	−1.72	46	−1.12
SP_F47	47	−4.05	17	13	9	39	−1.81	52	−1.41
SP_F50	50	−2.40	17	16	9	42	−1.39	50	−1.13

Note: CFMT/CFPT scores for Australian individuals suspected to have prosopagnosia (based on interview and/or famous face recognition), relative to Australian norms. All *z* scores are coded such that negative indicates scores poorer than mean. Red indicates < -2 standard deviations. Blue indicates < -1.6 standard deviations. In the participant code: SP = suspected prosopagnosic; F = female; M = male. Sex-specific norms are used for the CFPT (see main text). Norms for the famous face test (MACCS Famous Face Test 2008; Wilson, Palermo, Rivolta, Williams, & Schmalzl, unpublished test) were derived from Australian Caucasians aged 19–72 years ($N = 28$) using the same fit-and-residual procedure as that described in the present article. CFMT = Cambridge Face Memory Test; CFPT = Cambridge Face Perception Test. [To see this Table in colour, please see the online issue of the Journal.]

although not at below-criterion levels. This indicates a hit rate of at least 57% (4/7) for the CFMT, which compares favourably to a 27% hit rate for the Benton test (Duchaine & Nakayama, 2004, 2006a).

Regarding the CFPT, Table 7 reveals an interesting pattern in which, despite a general tendency for the suspected prosopagnosics to score below average for their age (i.e., 6/7 *z* scores are negative), most were only slightly below average, and only 1 individual performed at prosopagnosic levels (SP_F40). Indeed, 5 of the 7 were impaired to clinical levels on famous face recognition and/or the CFMT—both face memory tests—but were well within the normal range on the CFPT, a face perception test. This suggests that it is not uncommon for prosopagnosics to manifest prosopamnesia, meaning that affected individuals are able to represent faces successfully while they are physically present, but are unable to retain faces in memory (or at least to access any stored memory representations consciously, noting that famous face tests and the CFMT test explicit rather than implicit recognition).

Finally, we note that 1 individual in Table 7 (SP_F50) is impaired only on the famous faces task, and not on either the CFMT or the CFPT. The existence of an individual who can be normal on the CFMT and CFPT, while reporting substantial real-world face recognition difficulties and performing very poorly on famous face recognition, deserves further investigation in future studies. It is possible that the CMFT and CFPT do not completely overcome problems of nonface strategies contributing to performance. Alternatively, there might exist a genuine form of prosopagnosia in which an individual can perceive a face (CFPT) and retain it in memory for a few minutes (CFMT), but not retain it permanently in memory (famous faces, real-world recognition).

Prevalence of prosopagnosia in an unselected educated Australian sample

We now return to our main sample of 241 individuals. Because these participants were unselected for face recognition ability, examining participants whose performance was poorer than diagnosis cut-offs (>2 *SD*s) on the CFMT and/or CFPT allows us to gain some information on the likely rate of prosopagnosia in the general population of educated Australians. Further, given that none of the participants had reported suffering a major brain injury (such as stroke) that would plausibly

to lead to acquired prosopagnosia, cases uncovered are presumably developmental in origin.

Table 8 lists the relevant individuals. The first 5 would clearly be diagnosed as probable prosopagnosics (F21, F22, F73, M21, and M57). All these participants had CFMT scores more than 2.1 standard deviations poorer than the mean. Of the 5, 1 (M21) demonstrated prosopamnesia—he scored outside the normal range on the CFMT, but well within the normal range on the CFPT—while the other 3 of the 4 tested on the CFPT were impaired at both face memory and face perception (CFPT z of at least −1.9). In all 5 cases, the poor face recognition could not be attributed to generally poor abilities: The 2 young women were both successful undergraduate students (suggesting higher than average IQ); F21, F73, and M21 were tested on the MAT and a verbal memory test and performed within the normal range relative to age group norms (in fact, F73's verbal memory was noticeably above average), and M57 had normal-range IQ on Raven's matrices.

For 2 of these probable prosopagnosics, we conducted follow-up testing. In both cases, results were consistent with the presence of prosopagnosia. Both were impaired on a famous faces test (F21 and M57, see Table 8). F21 was also interviewed and stated that she could not follow films (and neither could her father), and she reported using nonface strategies for person recognition: For example, she said it took her some time to realize that two similar-looking people on her student dorm floor (two young women with similar body shape and blonde curly hair) were in fact different individuals, after which she put deliberate effort into learning to distinguish them by attending to differences in information such as exact degree of hair curl.

In addition to the 5 clear prosopagnosics, there were a further two cases with somewhat ambiguous results (Table 8). These individuals (M61 and F74) performed moderately more poorly than the cut-off on the CFPT, and more poorly than average but not below cut-off on the CFMT (although very close in the case of M61, with CFMT $z = -1.97$). Given the reasons to believe that diagnosis reliability is greater from the CFMT than the CFPT (see earlier sections), these two cases may be prosopagnosic, but could alternatively be individuals whose identity-related face processing is merely at the poor end of the normal range.

Our results suggest a developmental prosopagnosia prevalence rate of at least 2% including prosopamnesia. Note that we include prosopamnesia because, functionally, "prosopagnosia" is most usefully defined as the inability to *recognize* people, an ability for which face memory as well as face

Table 8. *Rate of prosopagnosia in a sample unselected for face recognition ability*

		CFMT		CFPT						
Participant	*Age (years)*	*Total*	*z*	*Upright*	*Up z*	*Word memory z*	*MAT z*	*Raven's z*	*Famous faces z*	*Likely diagnosis*
F21	21	36	−2.26	94	−5.07[a]	−0.55	−1.10	N/A	−2.51	prosopagnosia
F22	22	36	−2.27	N/A	N/A	N/A	N/A	N/A	N/A	prosopagnosia
F73	73	26	−2.34	104	−4.39	2.16	0.08	N/A	N/A	prosopagnosia
M21	21	37	−2.14	38	−0.44	−0.38	1.21	N/A	N/A	prosopagnosia
M57	57	28	−2.82	70	−1.93	N/A	N/A	−0.68	−3.10	prosopagnosia
M61	61	34	−1.97	76	−2.17	−1.13	0.02	N/A	N/A	prosopagnosia
F74	74	32	−1.58	88	−2.95	0.59	1.58	N/A	N/A	poor at faces?

Note: Individuals in our Australian sample (N=241) who revealed poorer-than-cut-off performance relative to age norms on either CFMT or CFPT. CFMT = Cambridge Face Memory Test; CFPT = Cambridge Face Perception Test.

[a]All z scores (number of standard deviations away from mean) are coded such that a negative z equals performance poorer than the mean. Red indicates < -2 standard deviations. Blue indicates < -1.6 standard deviations. On the CFPT, a negative z thus corresponds to a raw score higher than the mean (i.e., a higher error rate). On the CFPT, sex-specific norms have been used for the individuals aged 57–74 years. [To see this Table in colour, please see the online issue of the Journal.]

perception is essential. If we additionally include the two less clear individuals, the rate goes up to 2.9%. Importantly, our study provides the first estimate of prevalence rate of developmental prosopagnosia that has been based on objective testing of face recognition ability. Although our total sample size is of course small for an accurate estimation of prevalence of rare disorders, it is interesting to note that our results from objective testing (2–2.9%) are quite consistent with Kennerknecht et al.'s (2006) interview-based estimate of 2.47%. In particular, our findings provide support for the idea that developmental prosopagnosia, while rare, is far more common than might have been thought when the disorder was first identified (also see Behrmann & Avidan, 2005; Duchaine & Nakayama, 2006b).

Relationship between "relative to the average person" self-report and objective measures

The decision of a clinician to test for prosopagnosia, and/or the request from a client that such a diagnosis be considered, may be influenced by the client's own judgement of his or her face recognition ability relative to the typical person. The reliability of such simple self-judgements is, therefore, of some interest. In the only previous evidence, a subjective measure of face recognition did not correlate with objective measures of face memory performance (famous face recognition and incidental learning), although it did correlate with perceptual discrimination of "spacing" between facial features (Rotshtein, Geng, Driver, & Dolan, 2007).

We examined the correlations between self-rating and CFMT and CFPT scores in our Australian sample. Table 9 presents mean and standard deviation of self report measures. Table 10 presents correlation results. Given that age can affect metacognitive awareness, we calculated the correlations separately for young adults ($N = 115$ from Studies 1 and 2 for CFMT; $N = 29$ from Study 1 for CFMT), and older adults (55–88 years combined, $N = 69$ for CFMT from Study 1, $N = 59$ for CFPT from Study 1).

For young adults, there was a significant but small correlation ($r = .22$, $p < .05$) between self-rating and CFMTtotal,[5] such that individuals with better objective performance tended, to some limited extent, to be aware of this fact. Self rating also showed small significant correlations with Stages 1 and 2 of the CFMT considered independently, but not with Stage 3. Turning to the CFPT, self-rating and CFPTupright showed a very small relationship in the predicted direction (negative, because higher CFPT scores equal poorer performance), which was far from significant.

For older adults, Table 10 shows no suggestion at all of any relationship between self-report and either CFMT or CFPT performance. Older adults also showed no awareness of age-related decline: When asked to rate their face recognition abilities now compared to when they were 30 years old, the mean rating was 4.9 ($SD = 0.96$; $N = 68$, from Study 1) on the 0–10 scale where 5 represents equal to when they were 30.

Overall, results suggest that even young adults have only a very limited insight into how their face recognition abilities compare to those of other individuals, and that older adults generally have no insight at all. The lack of insight is further supported by information from the 5 prosopagnosics and the 1 very-nearly-prosopagnosic individual we identified within our main Australian sample (see previous section). Of these 6, 5 had been given the self-rating task: Only 2 rated their current face recognition ability as below average (3 on the scale of 0 to 10); the others rated their abilities as average (5) and above average (7 and 7.5), and 1 of these (F21) stated in follow-up interview that she had had no idea her face recognition ability was poorer than that of other individuals until we drew her attention to this fact. Regarding the 6th (M57), anecdotal evidence and gentle probing questions

[5] A very similar correlation of $r = .27$ was found by Germine, Duchaine, and Nakayama (2009), in which 47,471 participants rated their face recognition relative to the average on a 5-point scale and completed an online version of the CFMT that used different faces from the original CFMT.

Table 9. *Self-rating of face recognition ability*

	Age group (years)											
	Young adult			*55–64*			*65–74*			*75+*		
	N	*M*	*SD*	*N*	*M*	*SD*	*N*	*M*	*SD*	*N*	*M*	*SD*
Self-rating	115	6.83	1.56	25	6.84	1.31	29	6.43	1.42	15	7.23	1.69
Self-change	—	—	—	24	5.17	1.17	29	4.74	0.95	15	4.8	0.41

Note: Ratings compared to "the average person" for all age groups (0 = much worse than average; 5 = average; 10 = much better than average) and, for older age groups, self-ratings of face recognition ability now compared to when the participant was 30 years old (0 = much worse now than when I was 30; 5 = equal to when I was 30; 10 = much better at present than when I was 30).

Table 10. *Pearson correlations between self-rated face recognition ability and objective performance on the CFMT and CFPT*

Age group	*CFMT1*	*CFMT2*	*CFMT3*	*CFMTtot*	*CFPT*
Young adults	.205*	.258**	.123	.220*	−.119
Older adults (55–88 yrs)	−.107	.003	−.069	−.054	−.014

Note: CFMT = Cambridge Face Memory Test; CFPT = Cambridge Face Perception Test. Stages 1, 2, and 3 separately, plus total; CFPT (upright faces only). If individuals with higher self-rating demonstrate better objective performance, correlations should be positive for CFMT and negative for CFPT.

*$p < .05$. **$p < .01$, two-tailed.

suggest that he is unaware that his face recognition is not normal.

Self-report: Implications for diagnosis and for research ethics. Anecdotally, a number of different laboratories have observed that targeted questions—such as "Do you have trouble following films?"—can be useful in indicating prosopagnosia that is subsequently confirmed objectively. (We are not aware of any such questionnaires that have been published as yet.) However, the present results argue that simply asking people how good they are at face recognition relative to other people is not useful as a method of either diagnosing prosopagnosia or ruling out such a diagnosis. This finding is of practical importance, for two reasons.

First, a simple comparison-to-the-average is likely to be the idea that a member of the general public would have in mind if, for example, a clinician treating them for social anxiety asks "How good do you think you are at face recognition? Could there be a problem there, do you think?". A related implication for clinical work is that clients with underlying developmental prosopagnosia may well be unaware they have a cognitive deficit[6] and so are likely to attribute resulting social problems they may suffer to inappropriate causes (e.g., "people don't like me, I have an unattractive personality" or "I'm not trying hard enough to recognize people").

Second, for researchers, it raises a potentially important ethical issue—namely, that any testing of the CFMT/CFPT on "control" individuals has the potential to turn up individuals with extremely poor face recognition who are unaware of this fact. We suggest that researchers need to plan in advance for this possibility, for example by using research consent forms that ask participants if they wish to be informed should they perform at prosopagnosic levels.

Distribution shape for the Cambridge Face Memory Test

With CFMT data from a large number of Australian young adults unselected for face recognition ability, a final issue we were able to

[6] This can also occur with congenital perceptual deficits. Even with colour blindness, a phenomenon that is very well known in the general community (unlike prosopagnosia), affected individuals can make it to adulthood without realizing they are colour blind.

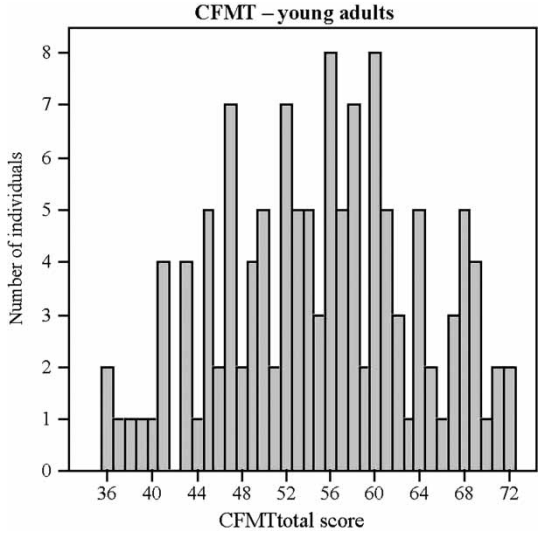

Figure 5. *Frequency distribution of CFMTtotal scores (upright faces) in 126 young adults from Australia/New Zealand. The 3 individuals scoring 36 and 37 are below the cut-off for prosopagnosia. CFMT = Cambridge Face Memory Test.*

address was the shape of the distribution of scores on this task. Figure 5 shows a frequency distribution for the relevant 126 young adults who completed the CFMT (including the 3 young adults discovered to be prosopagnosic, i.e., F21, F22, and M21 in Table 8). Results show that the distribution is normal in shape.[7] Statistical analysis revealed no significant deviation from normality, Kolmogorov–Smirnov = 0.051, $df = 126$, $p = .20$.

A concomitant finding was that, at least within this sample, the scores from the individuals with prosopagnosia did not form a discrete population on the CFMT. The histogram is unimodal rather than bimodal, and the 3 young adults diagnosed as prosopagnosics (F21, F22, M21) performed only incrementally poorer than the next-poorest participants. A possible interpretation is that developmental prosopagnosia reflects merely the low end of continuous variation in face recognition ability and not a discrete population that is different from the population of individuals with normal-range ability. However, we urge caution in drawing this conclusion, for several reasons. First, Yovel and Duchaine (2006) found that, although the 14 prosopagnosics studied were generally only mildly outside the normal range on the CFMT (mean $z = -2.5$), they were often much more dramatically so on famous face tests (mean $z = -6.5$). Second, on the CFPT, 2 of our prosopagnosic individuals tested on that task were dramatically outside the normal range (see Figure 3B). Third, even on the CFMT, some developmental prosopagnosics have been reported who perform far outside the normal range (e.g., 2 individuals with z between −3.5 and −4.0 in Duchaine et al., 2007a).

The finding of a normal distribution also raises the question of whether our prevalence estimate of 2–2.9% of the population with prosopagnosia is meaningful given that, by definition, 2.28% of a normal distribution fall more than 2 standard deviations below the mean. In response, the relevant point is the breadth of the distribution revealed in Figure 5. If face recognition accuracy were closely clustered around one value (i.e., all individuals were very similar), then certainly an individual could be more than 2 standard deviations poorer than the mean while still being quite good in absolute terms and not at all functionally prosopagnosic. However, Figure 3 shows that scores range from perfect to not very much better than chance, meaning that individuals 2 standard deviations poorer than the mean are performing extremely poorly in absolute terms. This is consistent with clinical observation that individuals with CFMT performance more than 2 standard deviations below the mean do indeed have severe real-world problems in face recognition (e.g., Yovel & Duchaine, 2006). Indeed, Duchaine et al. (2007a) found that, in a family of 10 prosopagnosic individuals (7 siblings, their parents, and uncle) with a mean CFMT z score of −2.8, the functional deficits were so severe that nametags were worn at family reunions. Overall, it seems very probable that 2–3% below criterion on the CMFT/CFPT does

[7] For a report on people from the high end of the distribution, see Russell, Duchaine, and Nakayama (2009).

indeed correspond to 2–3% of the population with real functional deficits in face recognition, severe enough to affect social interaction.

CONCLUSIONS

The present article was based on data for the Cambridge Face Memory Test and Cambridge Face Perception Test from individuals unselected for face recognition ability, in the largest such samples to date to be tested on either task (CFMT, $N = 240$ Australians; CFPT, $N = 125$ Australians). This data set allowed us to address a number of disparate issues. We now bring together our findings and implications, first in terms of theoretical questions about face recognition in general, and second in terms of practical issues related to using the CFMT and CFPT in cognitive neuropsychology.

Theoretical implications of our findings

Our findings have implications for a number of different questions of interest to face recognition researchers in general, not only those interested specifically in the CFMT/CFPT or in prosopagnosia. These are as follows (in no particular order).

Other-ethnicity effects in face recognition

The existence of other-race effects on face memory is well established, and there have also been several recent findings of other-age effects (e.g., Susilo, Crookes, McKone, & Turner, 2009). However, the question of whether exact ethnicity of faces within a race influences face recognition has received very little attention. Our present findings support those of three previous studies (Chiroro et al., 2008; Gilchrist & McKone, 2003; Sporer, 1999) to argue that exact ethnicity does matter, with memory being poorer for own-race faces of a different ethnic background than for own-race faces of one's own ethnic background. In the present study, the ethnicity difference to which participants were sensitive was within Caucasians and was probably best described as a difference in facial appearance deriving from perception of the faces as southern European/Middle Eastern ancestry versus British Isles/northern European ancestry. This is perhaps the broadest distinction that can be drawn within Caucasians (and also corresponds approximately to speakers of Romance versus Germanic languages). There have as yet been no tests of whether participants would be sensitive to—and show other-ethnicity effects for—finer variations in within-race appearance: We suspect there would not be other-ethnicity effects, for example, for Scottish versus Danish faces.

Sex differences in face recognition

Our results indicate that there are sex differences in face memory (and quite probably also in face perception), with a small mean advantage to women that can become significant with a sufficiently large sample size (e.g., $N > 200$). Specifically, because the tests used in the present study employed only male faces, we have shown that the female advantage in memory applies even for male faces. This is in disagreement with previous work (with smaller *N*s), which found that women showed better face memory than men but only for female faces (Lewin & Herlitz, 2002; McKelvie et al., 1993). Our evidence that women also show better face memory than men for male faces rules out an interpretation of the previous results in terms of an "own-group bias" (i.e., an idea that, similar to own-race or own-age effects, face memory is better for one's own sex than for the opposite sex). Instead, it argues that women are, on average, better than men at memory for faces per se; the reason for this effect remains unknown, but it could be related to women demonstrating greater social interest than men (Baron-Cohen, 2002), or possibly be part of a wider memory advantage in women (e.g., women also outperform men on verbal memory and memory for object position, although not on some forms of spatial memory; Galea & Kimura, 1993).

Ageing and holistic face processing

We are not aware of any previous studies that have examined the effects of ageing on the so-called

"special" aspect of face processing, variously known as holistic or configural processing. Our results argue that the strength of holistic/configural processing for faces is unaffected by ageing: The effect of face inversion is associated with a lack of holistic processing for inverted faces (e.g., Tanaka & Farah, 1993; Yin, 1969; Young, Hellawell, & Hay, 1987), and the present finding was that the inversion effect on the CFPT does not change with age in the 18–75-year age range. This is a theoretically important observation because it implies that the substantial ageing-related decline that occurs in performance on face memory and face perception tasks does not occur because of any age-related decline in the ability to process faces holistically.

Overlap of face memory and face perception

Our results argue that face memory and face perception are abilities that are strongly overlapping, although still dissociable, especially in cases of atypical development. The CFMT and the CFPT are very different tasks in structure (i.e., the CFMT is far from being the CFPT with merely an added memory component), and yet, across the full range of abilities, performance on the two tasks correlated highly relative to the maximum value predicted from the individual task reliabilities (observed correlation = .61; predicted maximum = .81). This argues that the underlying correlation between normal individuals' face memory and face perception abilities is strong. Interestingly, however, we also uncovered evidence suggesting that it is quite common for developmental prosopagnosics to show very poor face memory in conjunction with quite normal face perception. This implies that the neural mechanisms supporting conscious face memory and face perception are dissociable at least in atypical development.

Practical implications of our findings

Theoretical development in cognitive neuropsychology—for example, in understanding the pattern of dissociable contributions to prosopagnosia—is only as good as the tests available to researchers. If tests of a particular cognitive function are not internally reliable, or the norms against which a potential prosopagnosic's abilities are judged are inappropriate, then theoretical conclusions may have little validity. With respect to practical issues arising in the use of the CFMT and CFPT, the present article has contributed several novel and important findings. These are as follows.

Norms and ethnic match within Caucasians

For the CFMT, our results show that norms must be derived from countries or regions with a stimulus–participant match in ethnicity similar to that of the potential prosopagnosic. This applies even where participants and faces are matched for race (i.e., all Caucasian). The present article summarizes CFMT norms (for young adults, see Table 6) for the Boston area of the US, for Australia/New Zealand, for Israel, and for Germany. It is possible that ethnicity-specific norms would also be required for the CFPT, although this is currently unknown.

Norms and ageing

Once the appropriate country or region has been selected, our results show that norms must then be derived from an age-appropriate sample. For the CFMT, young adult norms can be used for participants aged up to 50 years, but age-matched samples are required beyond this age. For the CFPT, age-matched samples are certainly required beyond 50 years, and age-related decline quite possibly begins even earlier (e.g., by 35–40 years), perhaps because of the speeded nature of the task. In terms of efficiently collecting norm data across all ages, our data revealed no change in variance of CFMT or CFPT across the full adult age range (18–88 years); this implies that testing a full age-spread of controls and using curve fitting and the standard deviation of the residuals (as we have done in the present article) can be a more time-efficient way of collecting accurate norm data for a particular country/region than attempting to test a large control group for each age decade.

Sample size of control groups

For the CFMT, our results suggest that data from a control group that is age and ethnicity matched to a particular potential prosopagnosic should give a reasonably accurate estimate of the standard deviation (and thus the diagnosis cut-off score) with a sample size of approximately 30 controls. However, control sample sizes of 20–30 are not nearly adequate for the CFPT, given the random differences in standard deviation we observed across successive samples of this size (note that these presumably arise due to the rather lower individual-subject internal reliability for the CFPT than the CFMT; see below).

Norms and sex

Once an appropriate country and age of controls have been determined, then sex can safely be ignored for the CFMT where the aim is to diagnose prosopagnosia based on a score 2 standard deviations poorer than the mean: That is, because women show significantly better mean performance but slightly larger variability than men, CFMT norms can be derived from mixed-sex samples. For the CFPT, in contrast, things are unfortunately more complicated. Our findings suggested that, past young adulthood, a mean advantage to women was combined with slightly smaller variability, meaning that norms for middle-aged and older individuals on the CFPT need to be derived from sex-specific samples.

Norms and intelligence

It is important that further studies properly investigate the effects of intelligence. Our results for intelligence-related measures (e.g., number of years of education) were limited in important ways: We had no direct measure of IQ, and our data were obtained from a sample with atypically high education levels (86% of participants had at least some postsecondary education). Taking our findings together with those of the very few previous studies related to this issue, we have argued that intelligence probably does not affect face memory (CFMT) within the upper half of the intelligence range, although we could not rule out it affecting the CFPT (possibly because this is a speeded task); also, our results cast no light on the possible effects of intelligence on either task in the lower half of the IQ distribution.

Internal reliability for upright faces

Internal reliability scores (Cronbach's alpha) indicate that the CFMT (alpha = .89) is sufficiently reliable to provide an accurate indication of an individual's performance score. Reliability of the CFPT (upright faces alpha = .74) was lower, indicating that, although this task can be used for accurate group-level comparisons, and for diagnosing the more extreme cases of prosopagnosia, it should not be taken in isolation as an accurate indicator of probable prosopagnosia for individuals with *z* scores close to the cut-off value (e.g., *z* score of −2.1 or −1.9).

Internal reliability of CFPT for inverted faces

The internal reliability of the CFPT for inverted faces (alpha = .500) is too poor to provide an accurate indication of an individual's performance for inverted faces, or their upright-inverted difference score, and is best restricted to use in group-level comparisons.

Self-report

Our results argue that, in terms of suggesting, or ruling out, a diagnosis of prosopagnosia, no practical reliance should be based on a simple self-report of how a participant believes their face recognition compares "to the average". Congruently, it is possible for an individual to have developmental prosopagnosia with no awareness that their face recognition is atypical.

Prevalence of developmental prosopagnosia

Finally, we found that prevalence of developmental prosopagnosia, defined by objective performance, in an educated Australian population unselected for face recognition ability, was 2–2.9%.

Manuscript received 15 December 2008
Revised manuscript received 14 July 2009
Revised manuscript accepted 7 September 2009
First published online 16 November 2009

REFERENCES

Adams-Price, C. (1992). Eyewitness memory and aging: Predictors of accuracy in recall and person recognition. *Psychology and Aging*, *7*(4), 602–608.

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

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

Anstey, K., Dear, K., Christensen, H., & Jorm, A. F. (2005). Biomarkers, health, lifestyle, and demographic variables as correlates of reaction time performance in early, middle, and late adulthood. *The Quarterly Journal of Experimental Psychology Section A*, *58*(1), 5–21.

Baron-Cohen, S. (2002). The extreme male brain theory of autism. *Trends in Cognitive Sciences*, *6*(6), 248–254.

Baron-Cohen, S., Wheelwright, S., Skinner, R., Martin, J., & Clubley, E. (2001). The Autism-Spectrum Quotient (AQ): Evidence from Asperger syndrome/high-functioning autism, males and females, scientists and mathematicians. *Journal of Autism and Developmental Disorders*, *31*(1), 5–17.

Bastin, C., & Van der Linden, M. (2003). The contribution of recollection and familiarity to recognition memory: A study of the effects of test format and aging. *Neuropsychology*, *17*(1), 14–24.

Bate, S., Haslam, C., Tree, J. J., & Hodgson, T. L. (2008). Evidence of an eye movement-based memory effect in congenital prosopagnosia. *Cortex*, *44*, 806–819.

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

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

Benton, A. L., Sivan, A. B., Hamsher, K. D. S., Varney, N. R., & Spreen, O. (1983). *Contribution to neuropsychological assessment*. New York: Oxford University Press.

Bernstein, M. J., Young, S. G., & Hugenberg, K. (2007). The cross-category effect: Mere social categorization is sufficient to elicit an own-group bias in face recognition. *Psychological Science*, *8*, 706–712.

Billick, S. B., Siedenburg, E., Burgert, W., & Bruni-Solhkhah, S. M. (2001). Validation of the Mental Alternation Test with the Mini-Mental State Examination in geriatric psychiatric inpatients and normal controls. *Comprehensive Psychiatry*, *42*(3), 202–205.

Bowles, D. C. (2007). *Face processing through the years: Aging and performance on two tests used to diagnose prosopagnosia*. Unpublished honours thesis, Australian National University, Canberra, Australian Capital Territory, Australia.

Chiroro, P. M., Tredoux, C. G., Radaelli, S., & Meissner, C. A. (2008). Recognizing faces across continents: The effect of within-race variations on the own-race bias in face recognition. *Psychonomic Bulletin & Review*, *15*, 1089–1092.

DeGutis, J. M., Bentin, S., Robertson, L. C., & D'Esposito, M. (2007). Functional plasticity in ventral temporal cortex following cognitive rehabilitation of a congenital prosopagnosic. *Journal of Cognitive Neuroscience*, *19*(11), 1790–1802.

Dobson, E., & Rust, J. O. (1994). Memory for objects and faces by the mentally retarded and nonretarded. *Journal of Psychology*, *128*(3), 315.

Duchaine, B. C., Germine, L., & Nakayama, K. (2007a). Family resemblance: Ten family members with prosopagnosia and within-class object agnosia. *Cognitive Neuropsychology*, *24*(4), 419–430.

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

Duchaine, B. C., & Nakayama, K. (2006a). 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. C., & Nakayama, K. (2006b). Developmental prosopagnosia: A window to content-specific face processing. *Current Opinion in Neurobiology*, *16*(2), 166–173.

Duchaine, B. C., & Weidenfeld, A. (2003). An evaluation of two commonly used tests of unfamiliar face recognition. *Neuropsychologia*, *41*, 713–720.

Duchaine, B. C., Yovel, G., Butterworth, E., & Nakayama, K. (2006). Prosopagnosia as an impairment to face-specific mechanisms: Elimination of the alternative hypotheses in a developmental case. *Cognitive Neuropsychology*, *23*(5), 714–747.

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

Galea, L. A. M., & Kimura, D. (1993). Sex differences in route learning. *Personality and Individual Differences*, *14*, 53–65.

Garrido, L., Duchaine, B. C., & Nakayama, K. (2008). Face detection in normal and prosopagnosic individuals. *Journal of Neuropsychology*, *2*(1), 119–140.

Germine, L., Duchaine, B., & Nakayama, K. (2009). Hitting your peak at age 30: Behavioral evidence for extended development of face learning ability [Abstract]. *Journal of Vision*, *9*, 506, 506a.

Gilchrist, A., & McKone, E. (2003). Early maturity of face processing in children: Local and relational distinctiveness effects in 7-year-olds. *Visual Cognition*, *10*(7), 769–793.

Grady, C. L., McIntosh, A. R., Horwitz, B., & Rapoport, S. I. (2000). Age-related changes in the neural correlates of degraded and nondegraded face processing. *Cognitive Neuropsychology*, *17*(1–3), 165–186.

Habak, C., Wilkinson, F., & Wilson, H. R. (2008). Aging disrupts the neural transformations that link facial identity across views. *Vision Research*, *48*(1), 9–15.

Herlitz, A., & Yonker, J. E. (2002). Sex differences in episodic memory: The influence of intelligence. *Journal of Clinical and Experimental Neuropsychology*, *24*(1), 107–114.

Herzmann, G., Danthiir, V., Schacht, A., Sommer, W., & Wilhelm, O. (2008). Towards a comprehensive test battery for face cognition: Assessment of the tasks. *Behavior Research Methods*, *40*(3), 840–857.

Iaria, G., Bogod, N., Fox, C. J., & Barton, J. J. (2009). Developmental topographical disorientation: Case one. *Neuropsychologia*, *47*, 30–40.

Jeffery, L., & Anderson, M. (2004). The contribution of central and modular processes to face perception and individual differences in intelligence [Abstract]. *Australian Journal of Psychology*, *56*(Suppl.), 119.

Jones, B. N., Teng, E. L., Folstein, M. F., & Harrison, K. S. (1993). A new bedside test of cognition for patients with HIV infection. *Annals of Internal Medicine*, *119*(10), 1001–1004.

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

Lamont, A. C., Stewart-Williams, S., & Podd, J. (2005). Face recognition and aging: Effects of target age and memory load. *Memory and Cognition*, *33*(6), 1017–1024.

Lewin, C., & Herlitz, A. (2002). Sex differences in face recognition: Women's faces make the difference. *Brain and Cognition*, *50*(1), 121–128.

Lindholm, T. (2005). Own-age biases in verbal person memory. *Memory*, *13*(1), 21–30.

Loftus, G. R., & Harley, E. M. (2005). Why is it easier to identify someone close than far away? *Psychonomic Bulletin and Review*, *12*(1), 43–65.

Lott, L. A., Haegerstrom-Portnoy, G., Schneck, M. E., & Brabyn, J. A. (2005). Face recognition in the elderly. *Optometry and Vision Science*, *82*(10), 874–881.

McKelvie, S. J., Standing, L., St Jean, D., & Law, J. (1993). Gender differences in recognition memory for faces and cars: Evidence for the interest hypothesis. *Bulletin of the Psychonomic Society*, *31*(5), 447–448.

McKone, E. (2009). Holistic processing for faces operates over a wide range of sizes but is strongest at identification rather than conversational distances. *Vision Research*, *49*, 268–283.

Robbins, R., & McKone, E. (2007). No face-like processing for objects-of-expertise in three behavioural tasks. *Cognition*, *103*, 34–79.

Rotshtein, P., Geng, J. J., Driver, J., & Dolan, R. J. (2007). Role of features and second-order spatial relations in face discrimination, face recognition, and individual face skills: Behavioral and functional magnetic resonance imaging data. *Journal of Cognitive Neuroscience*, *19*, 1435–1452.

Russell, R., Duchaine, B. C., & Nakayama, K. (2009). Super-recognizers: People with extraordinary face recognition ability. *Psychonomic Bulletin and Review*, *16*(2), 252–257.

Salib, E., & McCarthy, J. (2002). Mental Alternation Test (MAT): A rapid and valid screening tool for dementia in primary care. *International Journal of Geriatric Psychiatry*, *12*, 1157–1161.

Schmalzl, L., Palermo, R., & Coltheart, M. (2008). Cognitive heterogeneity in genetically based prosopagnosia: A family study. *Journal of Neuropsychology*, *2*(1), 99–117.

Sliwinski, M. (1997). Aging and counting speed: Evidence for processing-specific slowing. *Psychology and Aging*, *12*(1), 38–49.

Sporer, S. L. (1999, July). *The own-race bias in Germany: Testing the contact hypothesis with Turks and Germans*. Paper presented at the 4th European Conference on Psychology and Law of the U.S. Psychology-Law Association and the European Psychology-Law Association, Dublin.

Susilo, T., Crookes, K., McKone, E., & Turner, H. (2009). The composite task reveals stronger holistic processing in children than adults for child faces. *PLoS ONE*, *4*(7): e6460.

Tanaka, J. W., & Farah, M. J. (1993). Parts and wholes in face recognition. *The Quarterly Journal of Experimental Psychology*, *46A*(2), 225–245.

Tippett, L. J., Miller, L. A., & Farah, M. J. (2000). Prosopamnesia: A selective impairment in face learning. *Cognitive Neuropsychology*, *17*, 241–255.

Valentine, T. (1991). A unified account of the effects of distinctiveness, inversion, and race in face recognition. *The Quarterly Journal of Experimental Psychology*, *43A*(2), 161–204.

Vision Sciences Research Corporation. (2002). *Functional Acuity Contrast Test (FACT)*. Chicago, IL: Stereo Optical Company.

Warrington, E. K. (1984). *Recognition Memory Test*. Windsor, UK: NFER-Nelson.

Williams, M. A., Berberovic, N., & Mattingley, J. B. (2007). Abnormal fMRI adaptation to unfamiliar faces in a case of developmental prosopamnesia. *Current Biology*, *17*, 1259–1264.

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

Yin, R. K. (1969). Looking at upside-down faces. *Journal of Experimental Psychology*, *81*(1), 141–145.

Young, A. W., Hellawell, D., & Hay, D. C. (1987). Configurational information in face perception. *Perception*, *16*, 747–759.

Yovel, G., & Duchaine, B. C. (2006). Specialized face perception mechanisms extract both part and spacing information: Evidence from developmental prosopagnosia. *Journal of Cognitive Neuroscience*, *18*(4), 580–593.