The Caledonian Face Test:

A New Test of Face Discrimination

Andrew J Logan[a], Fran Wilkinson[b], Hugh R Wilson[b], Gael E Gordon[c] and Gunter Loffler[c]

[a] Bradford School of Optometry and Vision Science, University of Bradford, UK.

[b] Centre for Vision Research, York University, Toronto, Ontario, Canada.

[c] Department of Life Sciences, Glasgow Caledonian University, UK.

Word Count: 7028.

Correspondence: Andrew J Logan; A.Logan@bradford.ac.uk

Fran Wilkinson; franw@yorku.ca

Hugh R Wilson; hrwilson@yorku.ca

Gael E Gordon; G.Gordon@gcu.ac.uk

Gunter Loffler; G.Loffler@gcu.ac.uk

Abstract

This study aimed to develop a clinical test of face perception which is applicable to a wide range of patients and can capture normal variability. The Caledonian face test utilises synthetic faces which combine simplicity with sufficient realism to permit individual identification. Face discrimination thresholds (i.e. minimum difference between faces required for accurate discrimination) were determined in an "odd-one-out" task. The difference between faces was controlled by an adaptive QUEST procedure. A broad range of face discrimination sensitivity was determined from a group (*N*=52) of young adults (mean 5.75%; *SD* 1.18; range 3.33-8.84%). The test is fast (3-4 minutes), repeatable (test-re-test r^2=0.795) and demonstrates a significant inversion effect. The potential to identify impairments of face discrimination was evaluated by testing LM who reported a lifelong difficulty with face perception. While LM's impairment for two established face tests was close to the criterion for significance (Z-scores of -2.20 and -2.27) for the Caledonian face test, her Z-score was -7.26, implying a more than three-fold higher sensitivity. The new face test provides a quantifiable and repeatable assessment of face discrimination ability. The enhanced sensitivity suggests that the Caledonian face test may be capable of detecting more subtle impairments of face perception than available tests.

Keywords: Clinical tests, face perception, face discrimination, psychophysics, synthetic faces, prosopagnosia.

1. Introduction

Faces are amongst the most complex and important stimuli that the visual system processes. Accurate interpretation of face information is important for social interactions. The biological salience of faces is reflected in the complex neuroanatomy that animals have evolved to process them. Several interconnected regions have been implicated in various aspects of face processing in primates, including the occipital face area (OFA) (Gauthier et al., 2000), fusiform face area (FFA) (Kanwisher, McDermott, & Chun, 1997), inferior temporal cortex (Desimone, Albright, Gross, & Bruce, 1984), superior temporal sulcus (STS) (Allison, Puce & McCarthy, 2000) and the amygdala (Breiter et al., 1996). Individuals with impairments of face recognition, for example due to developmental prosopagnosia (Yardley et al., 2008), experience embarrassment, anxiety, depression and even career limitations.

In the past, most case reports of impaired face perception concerned patients with acquired prosopagnosia; an inability to recognise familiar faces associated with specific patterns of brain damage (Barton, 2008). Acquired prosopagnosia is a relatively rare condition (Sorger et al., 2007). Developmental prosopagnosia, on the other hand, is considerably more common; current estimates place the prevalence of developmental prosopagnosia at 2-3% (Bowles et al., 2009, Kennerknecht et al., 2006). In most manifestations similar to the acquired form, patients with developmental prosopagnosia are typically free from any major structural brain abnormality (Susilo & Duchaine, 2013) but some differences in activity have been reported (Furl et al., 2011).

Impairments of face perception are also associated with, and result from, conditions other than prosopagnosia. It is well established that children with complex neurodevelopmental disorders (e.g. autism spectrum disorder (Weigelt, Koldewyn & Kanwisher, 2012), cerebral visual impairment (Dutton et al., 1996)) are at risk of impaired face perception. Face processing deficits have also been identified in patients with Alzheimer's disease (Roudier et al., 1998), Parkinson's disease (Sprengelmeyer et al., 2003) and schizophrenia (Kohler et al., 2009). Further, face recognition ability is diminished by conditions affecting visual acuity (VA) and contrast sensitivity (Bullimore, Bailey & Wacker, 1991, Lott et al., 2005, McCulloch et al., 2011). For example, Tejeria et al. (2002) found that 97% of patients with age-related macular degeneration experienced difficulties with face recognition. Similar impairments of face recognition have recently been identified in patients with glaucoma (Glen, Smith & Crabb, 2013). These impairments are associated with reduced quality of life and have been identified by patients as a high priority for improvement (Tejeria et al., 2002).
The primary aim of this work was to develop a novel, clinically-applicable test of face discrimination. Such a test could be utilised to identify patients at risk of impaired face

perception, quantify face perception difficulties, evaluate quality of life and thereby inform patients and their care-givers about rehabilitative and adaptive strategies. To be suitable for use in a clinical environment, the test must be adaptable for a wide range of patients and fast to administer. Moreover, a test should be able to differentiate between low-level (e.g. ocular) and high-level (e.g. cortical) causes of impaired face perception. This will be critical in patient management: dispensing a low vision aid will benefit certain patients but will not ameliorate prosopagnosia.

A number of tests of face perception are currently available. The vast majority employ face photographs as stimuli. Some tests (Benton, 1983, Warrington, 1984) have been criticised (Duchaine & Nakayama, 2006, Duchaine & Weidenfeld, 2003, Kress & Daum, 2003) for using images that are poorly controlled and include non-face information such as clothing, image artefacts and postures. In other tests, face photographs can be identified by highly distinctive features (e.g. the eyebrows) (Duchaine & Nakayama, 2004). It has been shown that patients with an established diagnosis of prosopagnosia can utilise alternative recognition strategies to score within the normal range of face recognition ability (Duchaine, 2000, Duchaine & Nakayama, 2004, Nunn, Postma & Pearson, 2001). This reduces test sensitivity to impairments of face perception. To address this problem, other tests employ cropped face photographs (Duchaine & Nakayama, 2006). While this has the desired effect of eliminating extraneous information it also removes key face features, such as the head-shape and hairline. It is well established that these external features make a significant contribution to unfamiliar face recognition (Bruce et al., 1999, Haig, 1986, Veres-Injac & Persike, 2009). It has been suggested that tests which utilise cropped face photographs may encourage participants to adopt atypical processing strategies (Burton, White & McNeill, 2010).

A number of face tests require participants to memorise and subsequently identify a series of unfamiliar faces (Duchaine & Nakayama, 2006, Warrington, 1984). Performance in such tests is influenced by the ability to process faces and to memorise them (Weigelt, Koldewyn & Kanwisher, 2012). While this may be an advantage when testing patients with prosopagnosia (often considered a disorder of face 'memory', with or without a face perception impairment) or autism spectrum disorder, these tests are unable to differentiate between a specific visual impairment of face perception and a more general memory deficit (e.g. Alzheimer's disease). In addition, some tests require participants to name face photographs of famous people (Rizzo, Venneri & Papagno, 2002). Performance on these tests is critically dependent upon familiarity with specific faces. The Cambridge face perception test, on the other hand, was specifically designed to assess face perception ability independently of any memory requirement (Duchaine, Germine & Nakayama, 2007).

Participants are asked to order six face photographs in terms of their similarity to a simultaneously-visible target face.

Restricted testing ranges are a further limitation of currently available face tests. This becomes problematic (Bird et al., 2003, Burton, White & McNeill, 2010, Nunn, Postma & Pearson, 2001, Russell, Duchaine & Nakayama, 2009) if face discrimination ability of typical participants exceeds the difficulty of a test. If typical participants score similarly, this prevents description of the normal range of face discrimination ability. For example, the original version of the Cambridge face memory test was unable to distinguish between control participants and "super-recognisers"; individuals known to have exceptional face recognition ability (Russell, Duchaine & Nakayama, 2009). It should be noted that this limitation has been overcome in an extended version of CFMT. Tests which cannot differentiate between patients without gross impairments of face perception may miss more subtle deficits. At the other extreme, if patients with significant deficits of face discrimination score close to chance level on existing tests, (Bowles et al., 2009, Duchaine & Nakayama, 2006, Nunn, Postma & Pearson, 2001) this precludes comparison of the relative severity of individual cases.

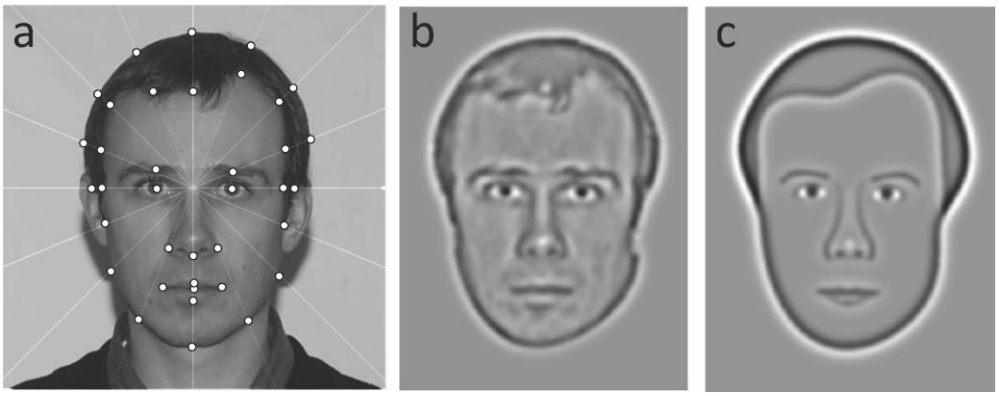
Finally, as most existing tests of face perception are based on face photographs, the sheer complexity of these images makes quantification of face discrimination ability difficult. To address this, the new face test described here utilises synthetic faces which combine simplicity with sufficient realism to permit accurate individual identification (Wilson, Loffler & Wilkinson, 2002). A comparable BOLD fMRI signal in the fusiform face area suggests that the brain processes these face images in a similar way to face photographs (Loffler et al., 2005b). Synthetic faces have previously been utilised to investigate cases of prosopagnosia (Lee et al., 2010).

This study aims to develop a new test of unfamiliar face discrimination; the Caledonian face test. The test stimuli include all the major face features (head-shape, hairline, eyes, nose, mouth, eyebrows). Non-face cues (e.g. clothing, postures, photographic imperfections) have been excluded. The Caledonian face test utilises a simultaneous presentation design which ensures that performance is independent of non-face factors (e.g. memory, familiarity). One key advantage of the synthetic face metric is that the range of possible face differences is unrestricted. The difference between synthetic faces can be adjusted to facilitate testing around threshold in any participant, irrespective of their absolute level of face discrimination ability. This removes ceiling or floor effects. The Caledonian face test provides a rapid and direct quantification of sensitivity to face information. The present study aimed to assess the validity of the face test and gather normative data.

2. Methods

2.1 Synthetic Faces

Synthetic faces (Wilson, Loffler & Wilkinson, 2002) capture the major geometrical face information from grey-scale face photographs with neutral expressions. A polar coordinate grid was superimposed on the face photograph, centred on the bridge of the subject's nose (figure 1a). The external contour of the subject's head was interpolated from 16 equally spaced measurements; the hairline from 9 further points. The internal features were defined by 14 additional measurements. While the position of all features was idiosyncratic, the shape of the eyes and eyebrows was generic. Individuating information was contained within variations in horizontal and vertical eye position, in addition to the height of the eyebrows, defined relative to the centre of the eyes. The mouth and nose shapes were derived from generic forms that were altered in terms of length and width based on individual face measurements.



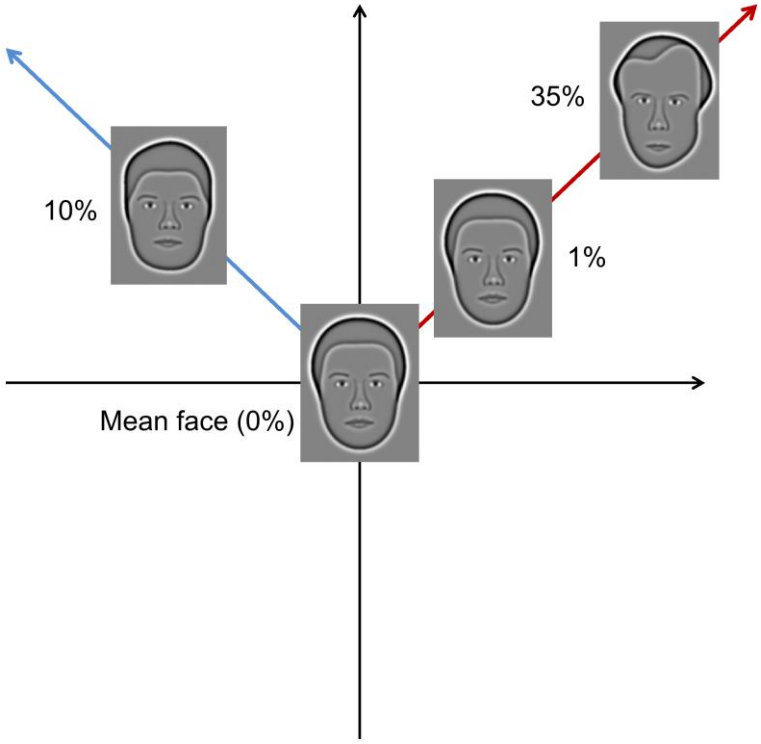
The authors suggest 1 column width for figure 1. Greyscale.

Figure 1. Synthetic face construction. (a) Greyscale photograph superimposed with polar coordinate grid centred on the bridge of the nose. The head-shape was measured at 16 locations (white dots) around the external contour, angularly positioned at equal intervals of 22.5°. The polar co-ordinates of 14 of the measured points were used to define 7 radial frequencies (RFs) to describe the subject's head-shape. RF patterns (Wilkinson, Wilson & Habak, 1998) are circular contours with sinusoidally-modulated radii that can be used to describe a range of natural shapes including fruits and head shapes (Wilkinson, Wilson & Habak, 1998). A further 9 points were utilised to define 4 RFs which captured the shape of the subject's hairline. All RFs were defined relative to the mean head radius of all synthetic faces of the subject's gender. The location and shape of the internal face features were also digitised. In sum, the face is described by 37 measurements. (b) Photograph filtered with a 2.0 octave bandwidth DOG filter with peak spatial frequency of 10 c/face width. (c) Corresponding synthetic face

In sum, each synthetic face is defined by 37 parameters and represented by a 37-dimensional vector. The images were subsequently band-pass filtered (circular DOG filter

with a bandwidth of 2.0 octaves) at the optimal spatial frequency for face identification (10 cycles/face-width; figure 1b) (Näsänen, 1999). The resulting faces accentuate geometric information in the most important frequency band while omitting high spatial frequency cues (e.g. hair texture, skin wrinkles) which contribute little to face identification (Goffaux & Rossion, 2006).

A mean face was produced by averaging each of the 37 dimensions of all synthetic faces of the same gender (see figure 2). All faces were expressed relative to the head size of the gender-appropriate mean face which served as the origin of a multi-dimensional face space. The identity of an individual face is represented by the direction of deviation from the mean face (colored arrows in figure 2). Within this framework, the difference between a given face and the mean face was calculated as the length of the face vector. Synthetic faces can be morphed to have any defined geometric difference from the mean face. This value quantifies mathematical differences between face stimuli. Previous studies have shown that this correlates closely with discrimination sensitivity (Wilson, Loffler & Wilkinson, 2002). The red arrow in figure 2 shows the mean face and two individual face images of the same identity that differ from the mean by 1% and 35% respectively. The Caledonian face test measures the minimum face difference required for reliable face discrimination (i.e. a face discrimination threshold).



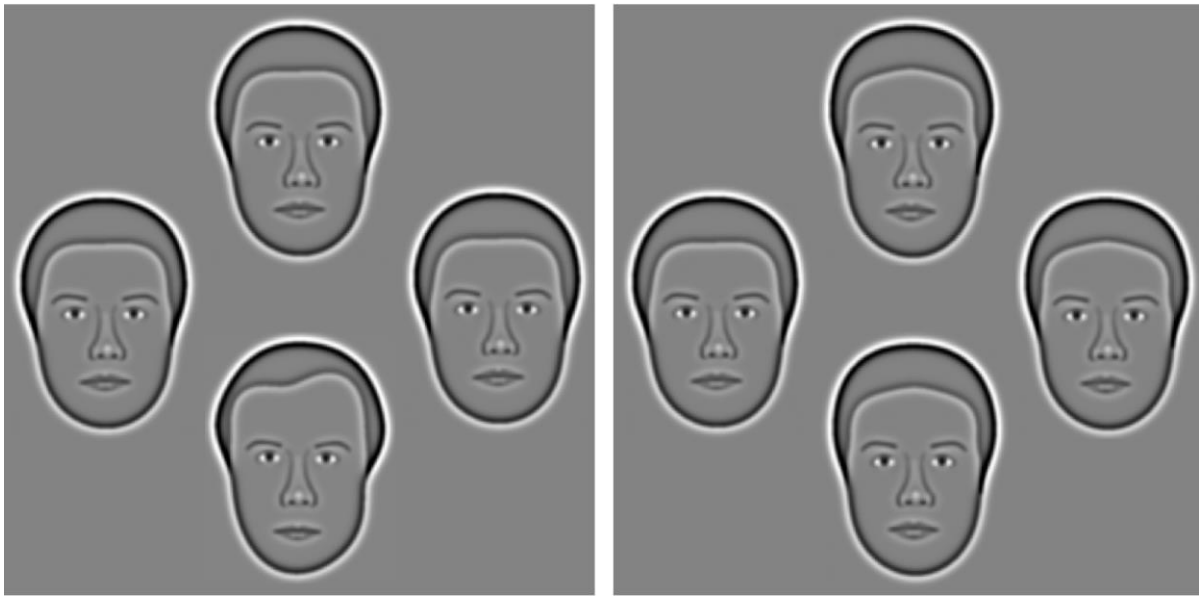
The authors suggest 1.5 column width for figure 2. Color online only.

Figure 2. Simplified synthetic face space. The mean face is located at the origin of a 37-dimensional framework. For clarity, this has been reduced to two dimensions (indicated by the black arrows). Different identities are represented by individual face vectors with discrete directions (colored arrows). Note that the synthetic faces on the two face vectors shown above depict two different identities. All faces which lie on the same face vector belong to the same identity. The difference between the mean face and a given face is quantified by the length of the face vector. This is expressed as a percentage of the mean head size. The difference between the mean face and a given face can be reduced to an imperceptible level (e.g. 1% face difference, see figure above) or amplified to such a degree that typical participants can easily identify that two faces are different (e.g. 35% face difference, see figure above). Note that the scaling on the red and blue axes is different. This unlimited range permits testing around threshold in all participants, without ceiling or floor effects.

2.2 The Caledonian face test

The Caledonian face test is a computer-based odd-one-out task. Participants are shown four faces in a diamond configuration (figure 3). Three of the faces are identical (distracters) while one face (target) is morphed to differ from the distracters by a specified amount. Participants are asked to respond by indicating the odd-one-out via computer mouse click and to guess when uncertain. Viewing time is unlimited. The mean face, which features in every trial, is randomly assigned as the target face in 50% of trials. The identity of the other face is randomly selected from a large database (40 male, 40 female). This allows the test

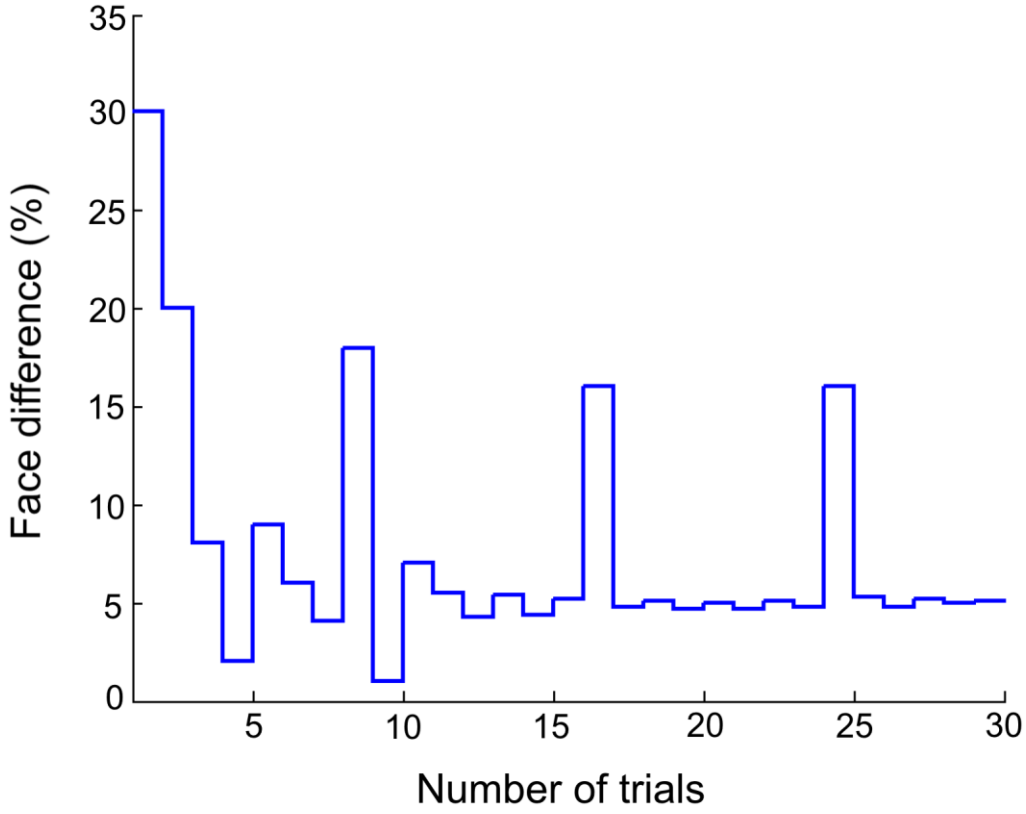
to be administered repeatedly without introducing familiarity or learning effects. Face gender is randomly selected for each trial. The gender of the mean face is matched to that of the non-mean face.



The authors suggest either 1.5 or 2 column width for figure 3. Greyscale.

Figure 3. The Caledonian face test. Observers were presented with 4 faces arranged in a diamond configuration and were asked to indicate the 'odd' face that differed from the others. Left: supra-threshold trial for most participants (target face differs from mean face by 10%). The target (odd one) is at the bottom. Right: difficult trial, approximately at threshold for a typical participant (5%). Target is to the left.

The magnitude of the difference between the faces on each trial is controlled by a QUEST adaptive procedure (Watson & Pelli, 1983). This highly efficient algorithm adjusts the task difficulty to concentrate testing around the participant's face discrimination threshold (see figure 4). QUEST utilises a maximum likelihood procedure to produce a threshold estimate after each trial based on all responses made from the beginning of the test run.



The authors suggest 1 – 1.5 column width for figure 4. Color online only.

Figure 4. Example test output. The graph shows the difference between the target and reference face as a function of trial number. The Caledonian face test varies the face difference over the test run in relation to the participant's performance. Typically, a correct response results in a decrease of face difference on a subsequent trial and vice versa. Note that the changes in face difference become smaller as the test progresses. To maintain participant engagement, dummy trials (face difference set to 3 times current threshold estimate) were included. Following established protocols for the measurement of visual acuity, a dummy trial was introduced every 7th trial (Bach, 1996). Dummy trials are indicated by the periodic peaks in the magnitude of face difference. The face discrimination threshold (5.1% in this example) is defined as the best estimate of threshold at the conclusion of the test.

2.3 Apparatus

The study was carried out with binocular viewing under an ambient illumination of 75 cd/m^2. Participants were seated at 1.2m from a LaCie "electron 22 blue II" monitor (1024 X 768 at 85 Hz) of 61 cd m^{-2} mean luminance which was controlled by an Apple Mac Mini computer. At the test distance, faces subtended 5.5° on average. The color look up table was defined to maximise contrast linearity of the monitor. The Caledonian face test was written in Matlab (www.mathworks.com) and includes routines from the Psychtoolbox extension (Brainard, 1997, Pelli, 1997).

2.4 Participants

In total, 80 naïve young adults (mean age = 26.4 years old, *SD* 13.84; range 20-52; 40 males and 40 females) took part in the study. Participants gave informed consent in accordance with the Declaration of Helsinki, as approved by the Life Sciences Ethics Committee of Glasgow Caledonian University. No reimbursement was offered for participation. Participants were unselected for face discrimination ability. All participants were in good health with no history of ocular disease, amblyopia (greater than one line difference in VA between the eyes) or strabismus. Participants with a refractive error of greater than ±6.00 DS or 2.50 DC were excluded.

All participants had normal or corrected-to-normal vision. Optimal refractive correction was determined for each participant and, where required, provided by trial lenses mounted in a trial frame. Distance VA was measured with a Bailey-Lovie LogMAR chart at 3m (Bailey & Lovie, 1976). Contrast sensitivity was assessed with a Pelli-Robson test chart (Pelli & Robson, 1988). Both charts were displayed at the luminance recommended by the manufacturers. Participants were required to have a best-achievable binocular VA of at least +0.10 LogMAR and no significant deficit in contrast sensitivity.

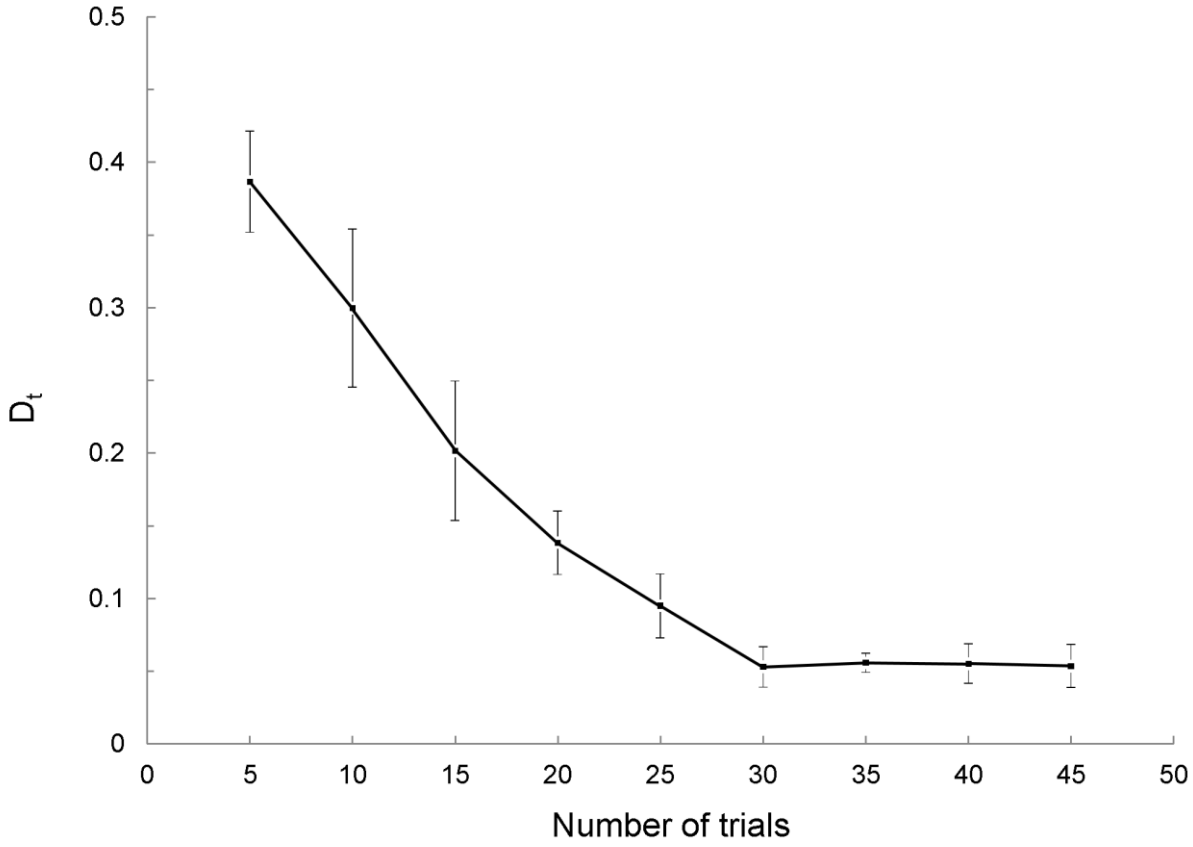
2.5 Statistical analysis

All statistical analyses utilised a one-factor, repeated measures ANOVA, unless otherwise specified. Where Mauchly's test indicated a violation of the sphericity assumption, the Greenhouse-Geisser correction was utilised.

3. Results

3.1 Number of Trials

The number of trials which provides the most efficient measurement of face discrimination thresholds was empirically determined. 10 participants (5 male) completed a 50 trial run of the Caledonian face test. The face discrimination threshold at the end of 50 trials was considered to be the participant's '*true'* threshold. Threshold estimates were then compared to this *'true'* threshold after every five trials. Figure 5 plots the deviation of each of these estimates from the observers' '*true'* threshold (D_t) as a function of the number of trials.



Authors suggest 1.5 column width for figure 5. Greyscale.

Figure 5. Effect of the number of trials on threshold estimates. The following equation was used to calculate the average deviation from the *'true'* threshold, determined after 50 trials:

$$D_t = \left(\sum_{i=1}^{N} \frac{\sqrt{(th_{t,i} - th_{50,i})^2}}{th_{50,i}} \right) \div N$$

D_t expresses the average, unsigned difference between the threshold estimate (*th*) made after *t* trials and the true threshold (th_{50}), measured after 50 trials, as a fraction of the true threshold for participant *i*. Data are averaged across 10 participants (*N*) and error bars denote the standard error of the mean. Threshold estimates improve with the number of trials up to 30 after which they plateau. At 30 trials, $D_t = 0.054$. This indicates that the threshold estimate after 30 trials is within 5.4% of the true threshold. This corresponds to a deviation that is approximately 20 times below threshold and, as such, imperceptible.

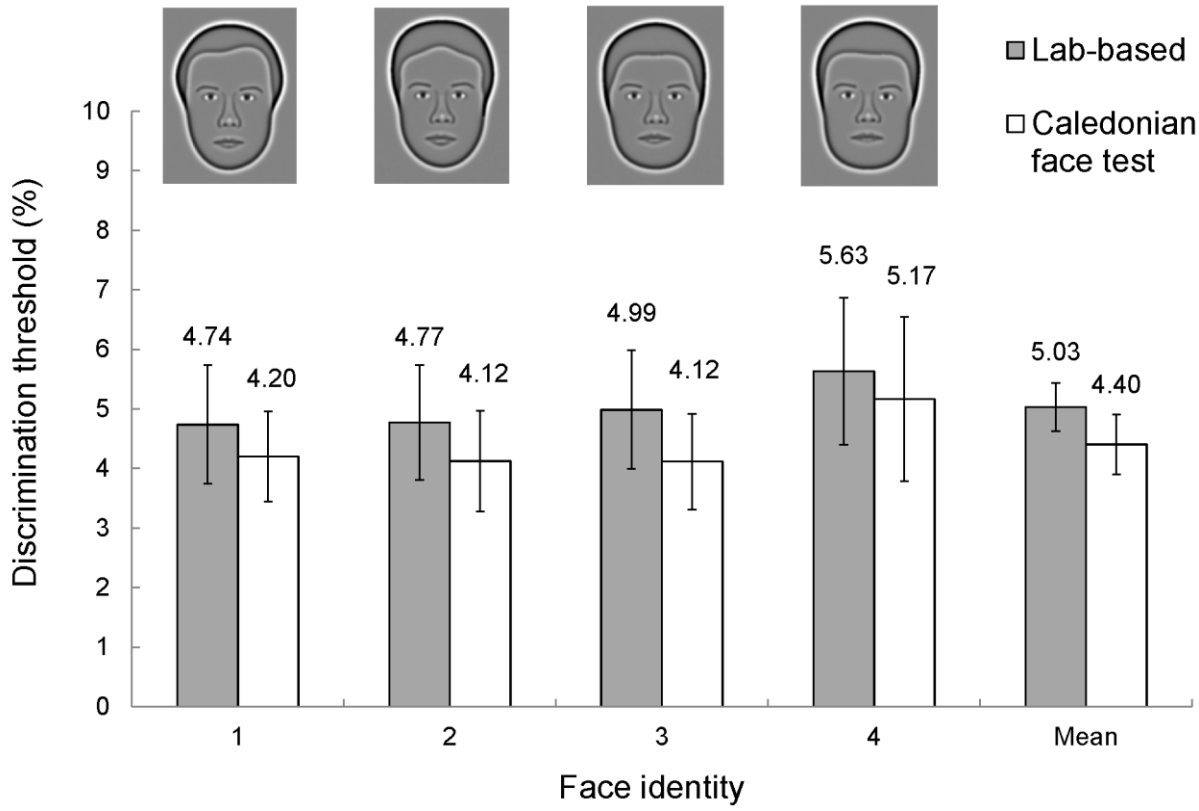
Threshold deviations (D_t) depended strongly on the number of trials ($F_{3.246,\ 29.216} = 20.327$; $p<0.001$). Figure 5 shows that threshold deviations decreased monotonically with the number of trials, until a plateau was reached at approximately 30 trials. Pairwise comparisons with Bonferroni correction revealed that there was no significant difference between the '*true'* threshold (50 trials) and that measured after 30 ($p=0.99$) trials.

3.2 Comparison with Lab-based Measurements

We compared discrimination thresholds measured by the new test with those measured for the same stimuli by an established lab-based procedure which has been used extensively (Loffler et al., 2005a, Or & Wilson, 2013, Wilson, Loffler & Wilkinson, 2002). This lab-based procedure (based on the method of constant stimuli) requires approximately 25 minutes to measure face discrimination sensitivity. This time requirement precludes the application of this method in a clinical setting. The Caledonian face test, however, provides a measurement of face discrimination sensitivity in approximately 3-4 minutes. This experiment aimed to investigate if the measurement of face discrimination sensitivity provided by the new face test is comparable to that provided by the established lab-based procedure

Participants were asked to discriminate between the mean face and a face which differed from the mean face by a specified amount. A target face was shown for 110ms, followed by a low-level noise mask and then a uniform grey screen, each for 200ms. Following the offset of the grey screen, two faces were presented. One face (target) was the face shown previously. The other face (distracter) differed from the target face by a specific amount. The observer was asked to indicate the target face via computer mouse click.

10 new participants (5 male) completed both the lab-based procedure and the Caledonian face test. Testing order was balanced. Face discrimination thresholds did not differ significantly between participants (univariate ANOVA $F_{9, 70}$ =1.457; p=0.086). Figure 6 presents a comparison of mean face discrimination thresholds measured by the two tests.



Authors suggest 1.5 column width for figure 6. Greyscale.

Figure 6. Mean discrimination thresholds for the lab-based procedure (dark bars) and Caledonian face test (light bars) across four face identities (shown by icons at the top of the figure at 10% face difference which is approximately two times threshold for typical observers). For the lab-based procedure, face discrimination accuracy was measured at 6 face differences, repeated 20 times each. The resulting data were fit by a Quick function (Quick, 1974) using a maximum likelihood procedure. Face discrimination thresholds were defined as the face difference that resulted in 75% accuracy. Within a single block, observers were presented with 4 different, randomly interleaved, face identities. The interleaved design randomised both the order of presentation of the individual identities and the increments of face difference. For the Caledonian face test, in order to allow direct comparison between the two tests for individual face identities, the test was modified to comprise four independent, but interleaved, staircases (each 30 trials) which measured face discrimination thresholds separately for the same four identities (light bars) tested by the lab-based procedure. The rightmost bars are the mean thresholds across observers and face identities. Error bars denote 95% confidence intervals.

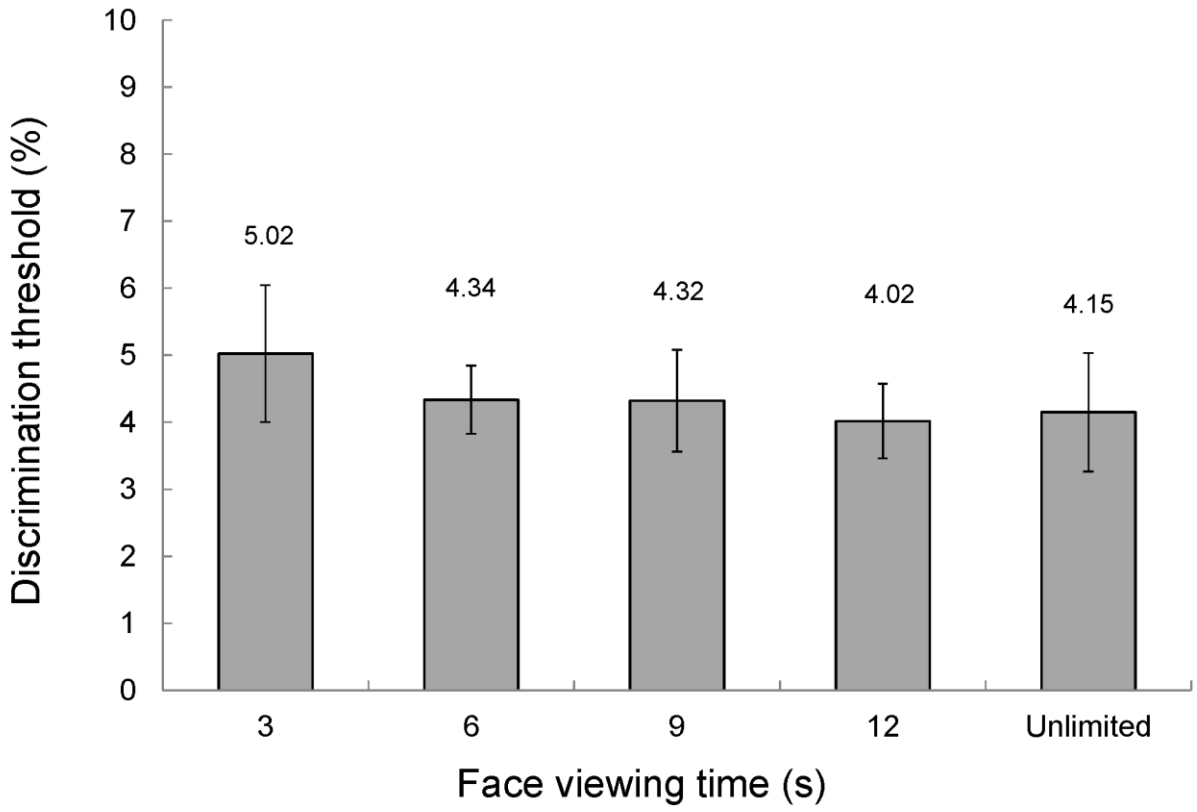
There was no significant effect of face identity on discrimination thresholds for either the lab-based procedure ($F_{3, 27}$ =2.942; p=0.074) or Caledonian face test ($F_{3, 27}$ =3.185; p=0.064). Accordingly, mean discrimination thresholds were calculated (figure 6 bars labelled ‘Mean’). Although it appears that the new face test slightly overestimates face discrimination

sensitivity, compared to the lab-based procedure, there was no significant effect of test type on face discrimination thresholds (two-factor [test type, identity], repeated measures ANOVA; $F_{1, 9} = 1.744$; p=0.219). This result demonstrates that, compared to the lab-based procedure, the Caledonian face test offers a significantly more efficient, but equally valid, measurement of face discrimination sensitivity. An additional experiment found no significant effect of test face gender on face discrimination thresholds measured by the new face test (paired samples t-test; $t(9)$=0.612; p=0.555).

3.3 Face Viewing Time

A common criticism of face tests which are based on a simultaneous presentation design, such as the Caledonian face test, is that participants are afforded the opportunity to match individual faces based on differences in local features (e.g. inter-ocular separation, lip thickness) (Bowles et al., 2009, Duchaine & Weidenfeld, 2003, Kress & Daum, 2003). This slow, piece-wise approach may not be representative of the rapid, holistic strategy which is utilised to recognise faces encountered in the real-world (Richler & Gauthier, 2014). An established method of overcoming this limitation is to restrict face viewing time (Duchaine & Weidenfeld, 2003, Nunn, Postma & Pearson, 2001).

We measured the effect of face viewing time on discrimination thresholds for the Caledonian face test (figure 7).



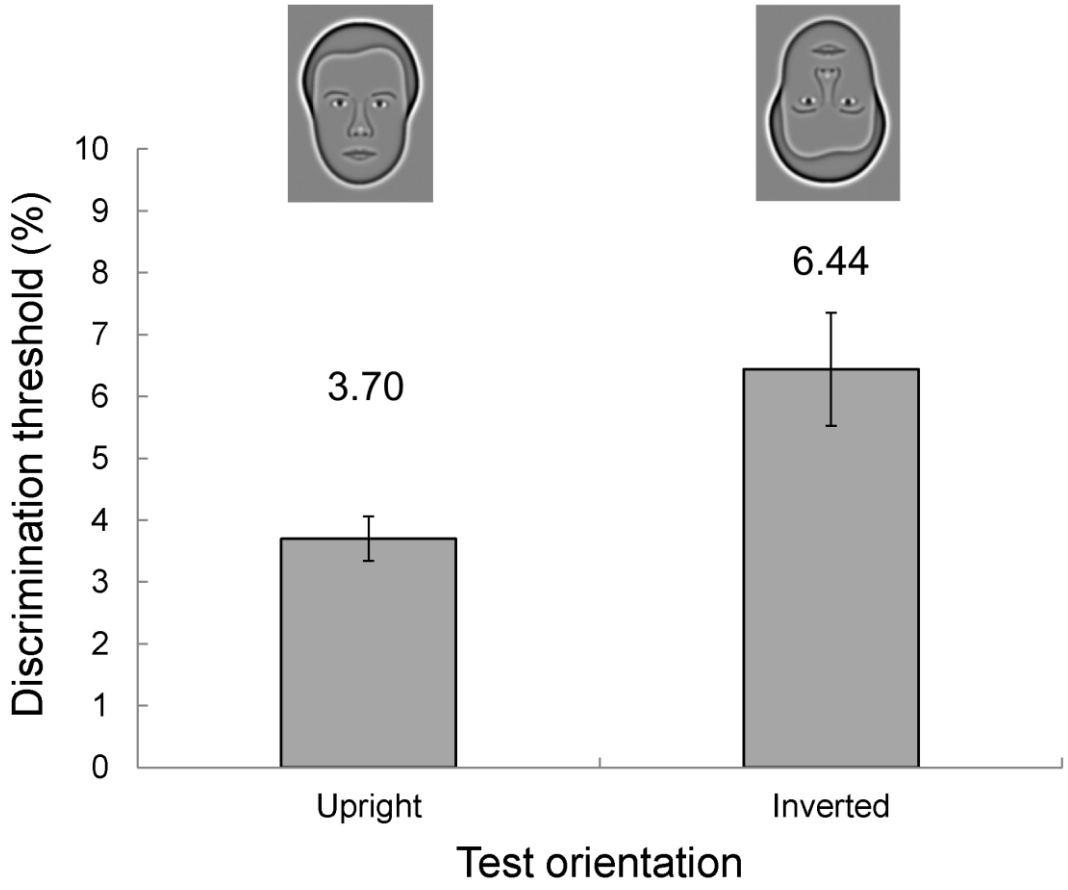
Authors suggest 1 column width for figure 7. Greyscale.

Figure 7. The effect of face viewing time on face discrimination thresholds. 7 participants (4 male) completed the Caledonian face test with 5 different presentation time limitations (3s, 6s, 9s, 12s and unlimited) administered in a balanced order. Following the allocated time period, the synthetic faces were replaced by an image of low-level noise which remained on the screen until a response was made. This was intended to remove any residual visual transient.. Responses made before the end of the specified duration were accepted and resulted in immediate progression to the next trial. Participants were encouraged to make full use of the available face viewing time and only to respond early when certain. Data are averaged across observers and error bars represent 95% confidence intervals.

There was no significant effect of face viewing time (3s, 6s, 9s, 12s and unlimited) on discrimination thresholds ($F_{4,24}$ = 1.627; p=0.20). Although it seemed that thresholds increased when the faces were shown for only 3 seconds, this difference was not significant (pairwise comparisons with Bonferroni correction; p>0.05). This result suggests that the Caledonian face test does not encourage typical participants to employ a slow, feature-based matching strategy, even when viewing time is unlimited.

3.4 The Inversion Effect

Compared to other objects, faces are significantly more sensitive to image inversion (Rossion, 2008, Yin, 1969). This disproportionate inversion effect is generally considered to be evidence that face processing mechanisms are distinct from those utilised for other objects (Farah et al., 1995, Leder & Carbon, 2006, Robbins & McKone, 2007). The existence of an inversion effect for the Caledonian face test would suggest that the test engages typical face processing mechanisms instead of the use of non-face specific strategies (e.g. identifying local differences in curvature, shape discrimination) (Duchaine & Nakayama, 2006).



Authors suggest 1 column width for figure 8. Greyscale.

Figure 8. Mean face discrimination thresholds for upright and inverted faces. 8 new participants (4 male) completed both an upright and inverted version of the test. The order of testing was balanced. There was no effect of participant on face discrimination thresholds (univariate ANOVA $F_{7,8}$ =0.138; p=0.986). Bars are mean discrimination thresholds and error bars denote 95% confidence intervals.

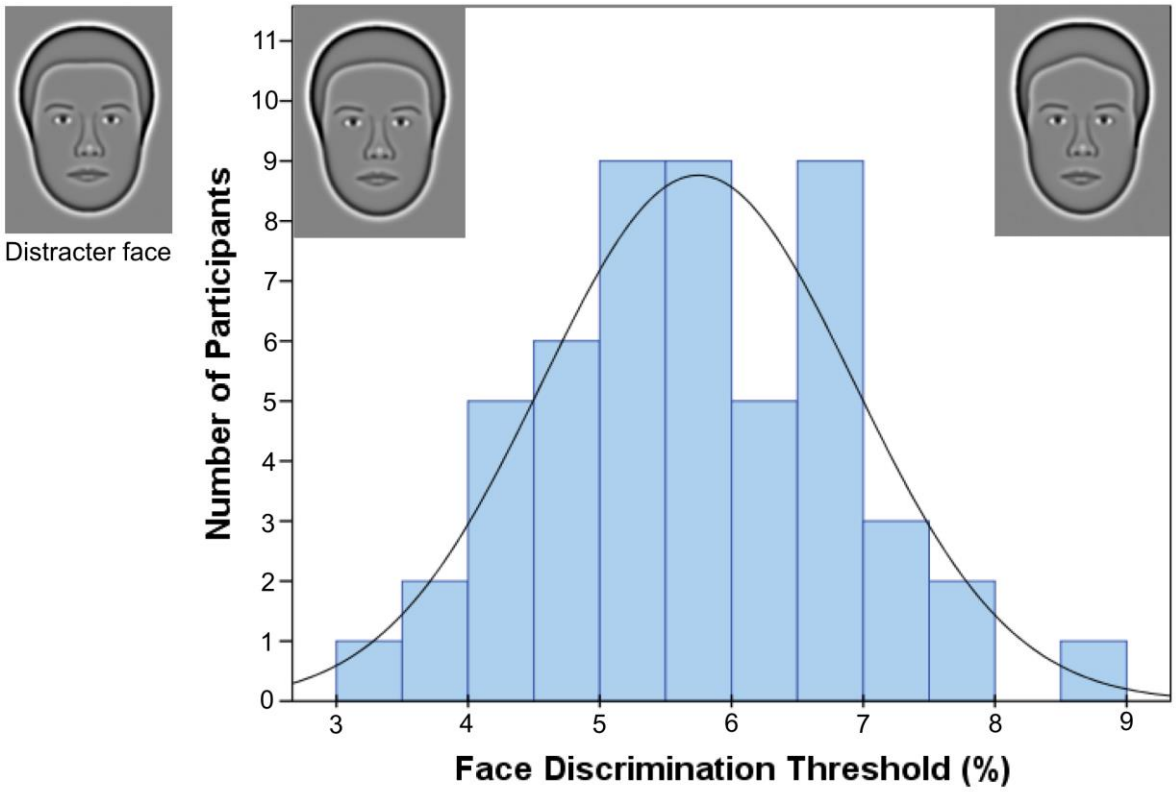
Discrimination thresholds (Figure 8) for inverted faces (6.17%, *SD*=1.231) were significantly higher than those for upright (3.82%, *SD*=0.484) faces (paired samples t-test; $t(7)=4.888$; $p<0.005$). Thresholds were increased, on average, by a factor of 1.74 as a result of test inversion. The finding of a robust inversion effect supports the proposal that the Caledonian face test engages face processing mechanisms.

In order to consider the magnitude of the inversion effect as measured with the new face test, we compared it with those reported for other face tests. Idiosyncratic test scoring systems make it difficult to directly compare inversion effects measured by different tests. Therefore, the analysis presented here considered the relative cost of face inversion and should be less influenced by individual test designs. For the Caledonian face test, inversion elevated mean face discrimination thresholds from 3.70% to 6.44%. This equates to an inversion effect of 1.74, i.e. observers are 74% poorer with inverted than upright faces. To our knowledge, there are no data available on the magnitude of the inversion effect for the Glasgow face matching test. The mean scores for the upright and inverted versions of the Cambridge Face Memory Test (scored as % correct recognitions) are 80.4% and 58.4% respectively, resulting in an inversion effect of 1.38 (Duchaine and Nakayama, 2006). The Cambridge Face Perception Test is scored as the number of errors made. The mean number of errors are 36.7 and 65 for the upright and inverted versions of the test respectively (Duchaine et al. 2007), giving an inversion effect of 1.77. Thus, the cost of inversion as measured with the new test is broadly in line with those found in previous studies.

3.5 Normative Data

None of the participants who had completed the experiments above were included in the normative data. This resulted in 52 participants (26 male) who completed one practice run of the Caledonian face test before data collection. Face viewing time was unlimited. The mean age was 29.7 years old (*SD* 15.13; range 20-52). Mean binocular VA for this group was found to be -0.11 LogMAR (*SD* 0.073; range -0.26 to +0.06). Mean contrast sensitivity was 1.93 log units (*SD* 0.136; range 1.65 to 2.10).

The distribution of face discrimination thresholds measured by the new face test is given in figure 9. Face discrimination thresholds ranged from 3.33-8.84%. The mean threshold was 5.75% (*SD* 1.18). The distribution of discrimination thresholds did not deviate significantly from a normal distribution (Shapiro-Wilk = 0.986, *df* = 52, *p*=0.807). The broad (*SD* = 20% of mean value) distribution suggests that the Caledonian face test is sensitive to small differences in face discrimination ability between normal participants.



Authors suggest 1.5-2 column width for figure 9. Color online only.

Figure 9. Histogram of face discrimination thresholds measured by the Caledonian face test. The fitted Gaussian indicates a normal distribution. The mean of 52 observers was 5.75% with a similar median (5.64%) which is in agreement with the finding that the data are not significantly skewed (skew = 0.152, *SE* = 0.330, *z* = 0.461, *p*>0.05). The icons above the histogram correspond to the face discrimination thresholds of the most (left; 3.3%) and least (right; 8.8%) sensitive participant. These observers could just discriminate between the icon face and the mean face (distractor; left inset).

Using the typical convention (Bowles et al., 2009), the normal range of face discrimination sensitivity was defined as the mean discrimination threshold ± 2 *SD*. This produced a normal range of face discrimination thresholds from 3.39-8.11% face difference.

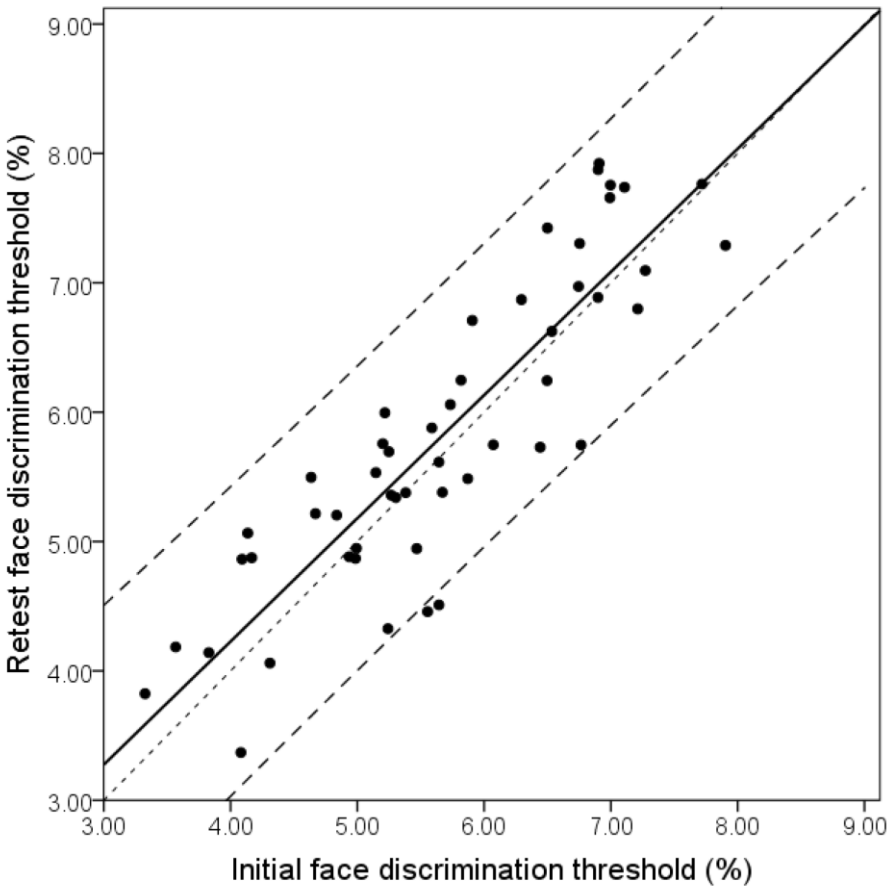
There was no significant effect of sex on face discrimination thresholds measured by the new face test (male = 5.81%, female = 5.68%; univariate ANOVA $F_{1, 50} = 0.146$; $p=0.704$). Further, there was no significant correlation between participant age and face discrimination thresholds (20-52 years; $r=0.0019$, $N=52$, $p=0.896$).

There was no effect of VA on face discrimination thresholds ($r=0.059$, $N=52$, $p=0.677$). The regression ($r^2=0.003$) was not significant ($F_{1,38} = 0.175$; $p=0.677$). Similarly, there was no

relationship between contrast sensitivity and face discrimination thresholds (r=-0.24, N=52, p=0.086; regression: r^2 =0.058; $F_{1,50}$ = 3.066; p=0.086). In sum, performance on the Caledonian face test is independent of normal variations in low-level aspects of vision (VA, CS) in a group of young adults.

3.6 Test-Retest Repeatability

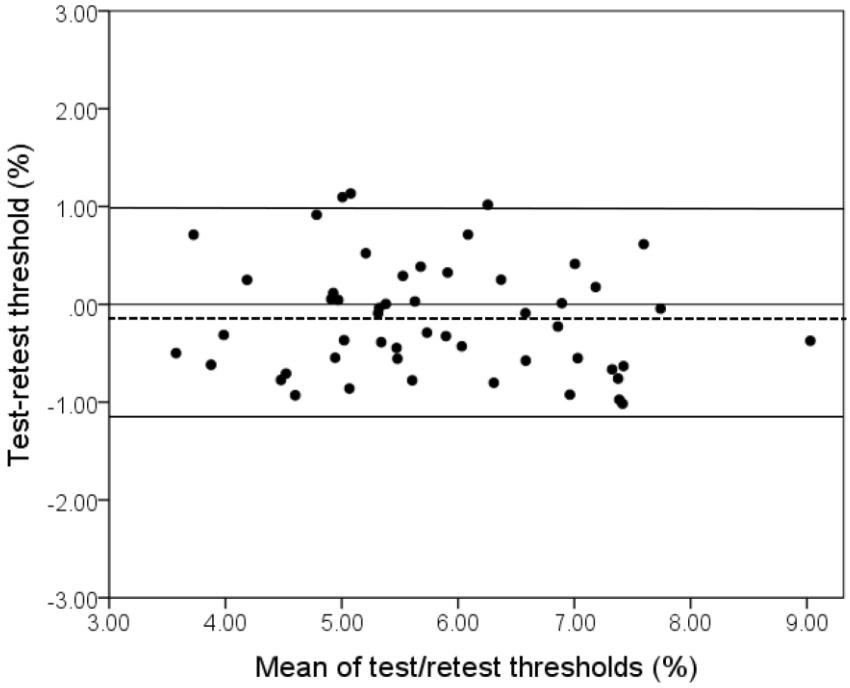
Participants were asked to complete the Caledonian face test again after approximately 30 minutes. Figure 10 presents a scatterplot of initial and retest thresholds. There was a positive correlation (r=0.892, N=52, p<0.001) between test and retest. The regression (r^2=0.795) was significant ($F_{1,50}$ = 193.825, p<0.001). The regression equation had a slope = 0.952 (t_{50} = 13.922, p<0.001). This value is close to the ideal slope of 1 which indicates that the measurement of face discrimination sensitivity made by the new face test is highly repeatable.



Authors suggest one column width for figure 10. Greyscale.

Figure 10. The relationship between initial and retest face discrimination thresholds for the Caledonian face test. The solid line indicates the line of best fit. 95% confidence intervals are represented by the two long-dashed lines. The central short-dashed line indicates the line of y=x.

We further analysed test-retest variability with a Bland-Altman plot (figure 11) (Bland & Altman, 1986).



Authors suggest one column width for figure 11. Greyscale.

Figure 11. Bland-Altman analysis of face discrimination thresholds. This method is utilised to assess the degree of agreement between test and retest thresholds by plotting the difference between the two measurements as a function of the mean of the test and retest thresholds. The mean difference between test and retest discrimination thresholds was -0.144% (dashed line). The bold solid lines indicate 95% confidence intervals. The 95% confidence limits span ±1.13%, i.e. 1/5 of the average threshold. The proximity of the mean difference to zero suggests that the new face test is highly repeatable.

This revealed (figure 11) that there is no significant change in the test-retest variability with mean face discrimination threshold. This suggests that the repeatability of the Caledonian face test is unaffected by individual differences in face discrimination sensitivity. Furthermore, there is no evidence of a learning effect which would be manifested by a significant positive mean difference between test and retest discrimination thresholds.

3.7 Identifying Impairments of Face Perception

The testing of a patient who experiences difficulty with face perception will determine the new face test's suitability to identify face processing impairments. Patient LM, a 36-year-old bookkeeper (right-handed) reported a specific difficulty recognising and discriminating

between individual faces. The symptoms have been present for as long as the patient can remember but have not been medically investigated. Although LM was born slightly prematurely, she is in good health and has no known neurological disorder or history of trauma.

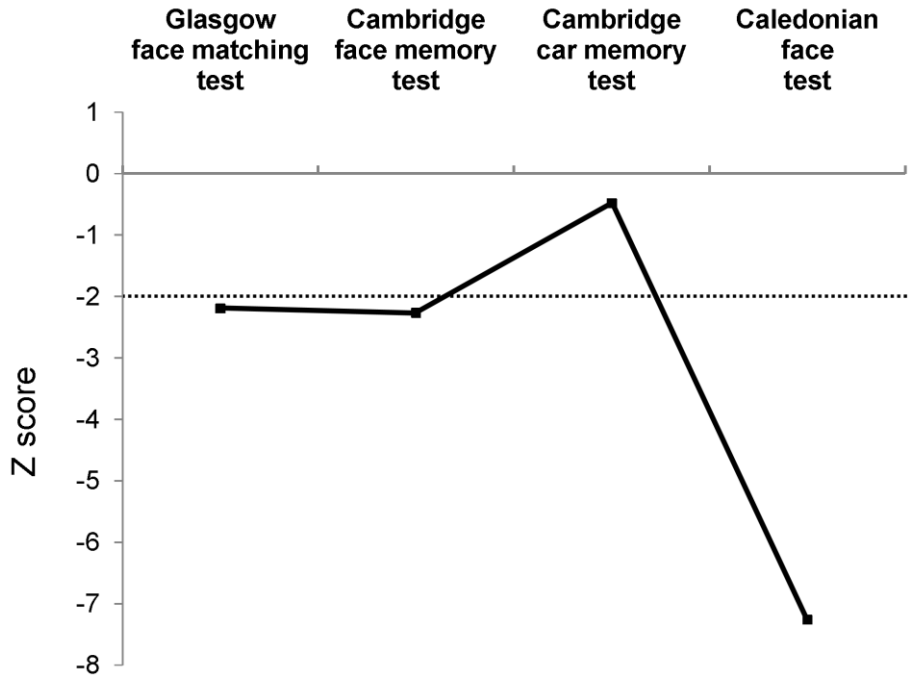
LM's experiences with face perception are consistent with those reported by patients with developmental prosopagnosia (Yardley et al., 2008). For example, cosmetic changes in appearance (e.g. a new hairstyle, make-up) impair LM's ability to identify familiar people. Similarly, the patient reported several instances of failing to recognise social acquaintances outside their usual context (e.g. a chance meeting in the street). LM consciously attends to alternative recognition cues (e.g. voices, clothing, gait and hairstyles) and experiences particular difficulty in situations where these are unavailable. For example, collecting her children from school is difficult because all children wear the same uniform.

Prior to testing, LM underwent a full eye examination which was unremarkable. VA (-0.12 Log MAR) and contrast sensitivity (2.10 log units) were in the normal range.

LM was asked to undertake two established tests of face perception. Firstly, LM's unfamiliar face discrimination sensitivity was assessed with the short version of the Glasgow face matching test (GFMT) (Burton, White & McNeill, 2010). The GFMT asks participants to indicate if two simultaneously-presented face photographs depict the same individual or two different people. LM scored 24 out of 40 (60%) on this test which is considerably lower than the average test score for control participants (mean = 81.3% (*SD* 9.7)). This score placed LM below the 5th percentile of the distribution of test scores for typical participants. LM then completed the Cambridge face memory test (CFMT; Duchaine and Nakayama (2006)), a test of face recognition with a memory requirement. LM correctly identified 40 out of a possible 72 faces. This is substantially poorer than the scores of control participants (mean = 57.92 (*SD* 7.91)). LM's score on the CFMT fell marginally below the established cut-off criterion (42/72) necessary for a diagnosis of impaired face recognition. To investigate the face-specificity of her impairment, LM was asked to undertake the Cambridge car memory test: a test of non-face object recognition which is identical in construction to the CFMT (Dennett et al., 2012). In stark contrast to her performance on the face tests, LM's car recognition score (47/72) was within the normal range (mean = 50.44 (*SD* 7.15)).

LM's face discrimination threshold (14.34%) measured by the Caledonian face test was approximately 2.5 times larger than the mean threshold for typical participants. This threshold lies far outside the normal range of face discrimination thresholds for typical adults calculated above (3.39-8.11%, see figure 9).

In order to directly compare LM's performance on all components of the test battery, Z-scores (multiple of the test's standard deviation) were calculated for each of the tests outlined above (figure 12).



Authors suggest 1 column width for figure 12. Greyscale.

Figure 12. LM's Z-scores for the full test battery (Glasgow face matching test, Cambridge face memory test, Cambridge car memory test and Caledonian face test). All tests were administered on the same computer monitor described above at a distance of 0.8m. Normative data for the tests was extracted from the original validation articles (Burton, White & McNeill, 2010, Dennett et al., 2012, Duchaine & Nakayama, 2006). Z-scores provide a quantification of the difference between a participant's test score and the mean test score as a multiple of the test's standard deviation and allow performance to be compared for different tests. The dashed line indicates the typical cut-off for impairment: a Z-score that is 2 standard deviations poorer than the mean score for typical participants. Specific Z-scores: Glasgow face matching test, z = -2.20; Cambridge face memory test, z = -2.27; Cambridge car memory test, z = -0.48, Caledonian face test, z = -7.26.

A test score which is two or more standard deviations poorer than the mean (i.e. z = -2) is generally considered to indicate impaired performance (dashed line in figure 12) (Bowles et al., 2009). The face-specificity of the patient's impairment is supported by her normal performance on the car recognition test. LM was impaired on all three face tests. The Caledonian face test, however, proved to be considerably more sensitive than the 2 standard tests. LM's face impairment for the standard tests was close to the criterion for

significance (Z-scores of -2.20 and -2.27). For the Caledonian face test, the patient's Z-score was -7.26, implying a more than three-fold higher sensitivity. Although these findings are currently limited to this case report, this suggests that the new face test is considerably more sensitive to impairments of face perception than those tests which are currently available. This is likely to be attributable to the extended testing range of the Caledonian face test.

4. Discussion

This study has outlined the construction and validation of a new clinical test of face perception. The adaptive nature of the Caledonian face test provides a highly efficient measurement of face discrimination ability. The Caledonian face test was completed by all participants within 4 minutes. This is significantly shorter than the duration of existing face tests (typically 10 to 15 minutes) (Burton, White & McNeill, 2010, Duchaine & Nakayama, 2006) and is similar to the time required for other clinical tests of vision, such as perimetry (e.g. SITA Fast threshold algorithm on a Humphrey visual field analyser) (Bengtsson & Heijl, 1998).

4.1 Test advantages & limitations

In addition to this increased efficiency, the Caledonian face test has a number of other important advantages. Firstly, the test stimuli contain only face information (e.g. head-shape and hairline); non-face cues (e.g. clothing, jewellery, photographic imperfections) have been excluded. This ensures that the new face test specifically assesses sensitivity to face information. This is supported by the finding of a significant inversion effect; an established hallmark of face-specific processing mechanisms (Robbins & McKone, 2007). The magnitude of the inversion effect is of comparable magnitude to those reported in the literature, suggesting that the new face test engages processing mechanisms which are specialised for face processing.

Unlike existing tests, the identities and order of the faces in the new test are randomly selected on each administration. This approach is intended to minimise the potential of familiarity effects. Participants may also have benefitted from a learning effect regarding the test protocol. To explicitly investigate the possibility of observers becoming familiar with the test and/or any of the faces, we compared test-retest results in 52 participants. Face discrimination thresholds were found to be highly repeatable (test-retest threshold R^2 = 0.795). Bland-Altman analysis revealed that the repeatability of the new face test is unaffected by the face discrimination sensitivity of individual observers. This analysis showed that the repeatability of the threshold measurement provided by the Caledonian face

test is approximately 1% face difference, i.e. 1/5 of the typical threshold. Face differences of this magnitude are imperceptible. Overall, the data show no effect of familiarity or learning. As a result, participants can undertake the new face test multiple times without becoming familiar with either the target faces or their order of presentation. The Caledonian face test has the potential to be utilised as a training tool for patients who have impairments of face perception as part of a rehabilitative strategy. For example, recent studies have indicated that perceptual training paradigms can successfully improve face discrimination ability in patients with prosopagnosia (Bate et al., 2014, DeGutis, Cohan & Nakayama, 2014).

Another key advantage of the Caledonian face test is that the synthetic face stimuli have been filtered for the bandwidth of spatial frequencies that are most important for face identification (circular DOG filter with peak SF of 10 cycles/face-width and a bandwidth of 2.0 octaves). The peak spatial frequency of the filter when presented at 80cm is equivalent to 2 cycles per degree (cpd) and 1.2 LogMAR VA. The face photographs employed in other face tests, on the other hand, contain the full range of spatial frequency information. Employing such broadband stimuli limits the conclusions that can be drawn from a below-average test result: any impairment of face discrimination identified could be due to impaired spatial resolution, a cortical impairment of face processing or a combination of the two. The filtering applied to the synthetic faces largely eliminates this low-level vision confound, as long as VA is above about 1.0 logMAR (equivalent to 6/60 or 20/200).

The present study has demonstrated this empirically: the measured face sensitivity with the new face test is independent of modest differences in VA or contrast sensitivity. Synthetic face discrimination thresholds have been shown to be scale invariant (Wilson, Loffler & Wilkinson, 2002). Consequently, the test stimuli can be enlarged (or viewing distance reduced) to ensure that participants with reduced VA can resolve the individual faces. This will facilitate the application of the Caledonian face test to clinical populations with reduced VA who are at risk of impairments of face perception. For example, deficiencies in face processing have been identified in patients with age-related macular degeneration (Barnes, De l'Aune & Schuchard, 2011, Bullimore, Bailey & Wacker, 1991) and glaucoma (Glen, Smith & Crabb, 2013).

Some existing tests initially familiarise participants within individual faces and assess face recognition accuracy in a subsequent testing phase (Duchaine & Nakayama, 2006). Tests based on this type of recognition paradigm cannot differentiate between specific impairments of face perception and more general memory deficits. The Caledonian face test, on the other hand, utilises a simultaneous presentation design to make a specific assessment of face discrimination sensitivity, which is independent of any memory impairment.

It has been suggested, however, that this simultaneous presentation design encourages participants to adopt a slow and atypical, feature-based matching strategy (Bowles et al., 2009, Duchaine & Weidenfeld, 2003, Kress & Daum, 2003). Specifically, participants are afforded the opportunity to compare local features (e.g. nose length, eyebrow thickness) of individual face images in a piece-wise manner. This task-specific approach may not representative of the rapid, holistic processing strategies which are typically utilised to recognise faces encountered in the real world (Richler & Gauthier, 2014). Consistent with this premise, Nunn, Postma and Pearson (2001) reported that a patient with a symptomatic impairment of face perception used a feature-based matching strategy to score within the normal range of test scores on a face test which uses a simultaneous presentation design.

Restricting face viewing time is an established method of overcoming this limitation (Duchaine & Weidenfeld, 2003, Nunn, Postma & Pearson, 2001). Nunn and colleagues reported that the patient described above performed at chance level when the authors re-administered the test with viewing time restricted to 3 seconds. The present study has demonstrated that reducing face viewing time on the Caledonian face test from unlimited to 3 seconds left face discrimination thresholds unchanged in a normal cohort. Such a short presentation time makes it unlikely that participants scrutinised local differences in individual features to identify the target face.

Therefore, the Caledonian face test can be administered with a restricted face viewing time. This may enhance the sensitivity of the new face test by ensuring that patients with impairments of face perception are unable to use feature-based cues. In addition, time constraints can significantly increase the efficiency of clinical tests which may enhance the appeal of the test to health-care professionals. Further studies, especially on older and younger observers, are required to establish the best compromise between short test time and sufficient viewing time in order to not disproportionately affect test scores in particular populations (e.g. older adults).

As with some other clinical measurements (e.g. visual acuity), the new face test does not have a theoretical minimum or maximum level of performance. Based on the mathematical description of individual faces by 37 parameters, differences between faces can be adjusted to an arbitrary small or large level, thus allowing performance to be measured without floor or ceiling effects. In practice, as with visual acuity, the maximum level of performance of the new face test is limited by the resolution of the display used. With typical displays and viewing distances, that limit is about $1/10^{th}$ of the best performance we have measured in our sample (thresholds ranged from 3.33-8.84%). The actual minimum level of performance that can be measured is restricted by a mathematical algorithm that ultimately results in

unrealistic ‘face’ images (e.g. where the eyes start to fall outside the head contour or the nose starts to overlap with the mouth). That limit is about five times the poorest performance measured in our sample. As our data for patient LM show, this limit is sufficient to allow face discrimination thresholds to be measured in participants with significant impairments of face perception.

It is also important to consider some of the limitations of the Caledonian face test. By their simplified nature, synthetic faces do not contain all of the information present in face photographs. The motivation behind the design of these stimuli was to reduce the enormous complexity inherent in face photographs and to construct a set of face images that can be manipulated in an easy and quantifiable way. To this extent, the faces were designed to concentrate on the most salient geometric aspects of faces (head shape, position of eyes, position and shape of mouth etc.) while ignoring other aspects (e.g. skin and hair texture, surface reflectance). This seems a justifiable simplification as humans can easily recognize a face from a distance of 5m or more despite much information, including that related to skin texture or surface reflectance, being limited or unavailable. This was achieved by applying a band-pass filter to the face images, removing any information contained within high spatial frequencies. Indeed, it has been shown that high spatial frequency information contributes relatively little to face recognition (Fiorentini et al. 1983). Band-pass filtering has another advantage. It provides a level playing field: an ideal clinical test that aims to measure face processing ability should be largely unaffected by optical factors. That is, the optical correction worn by the participant and any potential visual acuity deficit should not affect the test result. As outlined above, the test achieves a degree of independence from high-spatial frequency detail and can therefore be applied to a wide range of both typical participants - irrespective of their correction (optimal correction, habitual correction, near or distance prescription) – and those with compromised visual acuity (e.g. patients with amblyopia, cataract or mild visual impairment). In our view, this advantage outweighs the disadvantage that comes with a simplified stimulus set that ignores some information contained in high spatial frequencies, including surface reflectance and skin/hair texture, which observers may be able to use when recognizing faces from photographs (Russell & Sinha, 2007).

Another simplification of the face set used here concerns the eyes and eyebrows. While the shape of the nose and mouth varies between individual synthetic faces, the shape, but not the location, of the eyes and eyebrows are generic. Introducing individual differences in eye and eyebrow shape would add further complexity to the synthetic faces and could be implemented in the future to enhance the distinctiveness of individual identities. This may be particularly important for the testing of patients with specific impairments of face perception; it has been reported that individuals with prosopagnosia demonstrate a significant reduction

in sensitivity to the eye region (Caldara et al., 2005). Given that the design of our faces limits information provided by the eyes (i.e. faces contain positional eye information but no other features that could be used to discriminate between eyes such as iris colour, shape of eyelids etc.), we would expect this to have a limiting effect on individuals who could use such additional information but not for those who make limited use of eye information (e.g. patients with prosopagnosia). This should result in a smaller difference between typical participants and those with prosopagnosia for our test compared to tests where additional eye information is available. It is of note that our results, although based on a single case of impaired face perception, show the opposite, suggesting that our test is more sensitive to abnormal face perception. Finally, the new face test presents all faces in a frontal view. Variations in face viewing angle may allow the test to better emulate face discrimination tasks performed in the real world. In sum, the Caledonian face test is not intended to replace existing face tests. We propose that the particular design of the faces, and the test in general, has a number of advantages which make it a valuable addition to the battery of tests used to assess face perception ability, with a particular application for clinical settings. Moreover, poor performance on the Caledonian Face Test would be particularly indicative of a specific face perception deficit if the test were used as part of a test battery that also includes the measurement of discrimination sensitivity for non-face objects. Interested parties are invited to contact the corresponding author to request a copy of the test as well as general instructions and information on score interpretation.

4.2 Normative Data

The Caledonian face test provides a direct quantification of face discrimination sensitivity. In agreement with previous reports (Burton, White & McNeill, 2010), a broad range (*SD* = 20% of mean value) of unfamiliar face discrimination sensitivity was identified in a group of typical adults. This suggests that the new face test is sufficiently sensitive to identify subtle differences in face discrimination ability.

The distribution of face discrimination thresholds measured by the Caledonian face test was utilised to calculate the normal range of face discrimination sensitivity. Using the typical convention (Bowles et al., 2009), a cut-off score for normal sensitivity was defined as the mean discrimination threshold ± 2 *SD*. This produced a normal range of face discrimination sensitivity from 3.39-8.11%. Accordingly, face discrimination thresholds greater than 8.11% should be considered indicative of impaired face discrimination. Participants with thresholds less than 3.39%, on the other hand, can be categorised as being particularly sensitive to differences between individual faces. It has been suggested that information about an individual's location on the spectrum of unfamiliar face discrimination sensitivity may be

particularly useful in determining their suitability for certain careers (e.g. border control officers) (Russell, Duchaine & Nakayama, 2009). Moreover, Russell and colleagues proposed that the sensitivity to face information of eyewitnesses could be taken into account during legal proceedings.

It should be noted that the normal range of face discrimination sensitivity outlined above is only applicable to young adults. The normative data for the Caledonian face test were collected from participants aged between 20 and 52 years old. No effect of age on face discrimination thresholds was identified within this participant group. It has been reported that sensitivity to face information is reduced in older adults (Boutet & Faubert, 2006, Chaby, Narme & George, 2011, Crook & Larrabee, 1992). Moreover, a number of studies have indicated that face processing mechanisms mature slowly and do not reach adult levels until the second decade of life or later (De Heering, Rossion & Maurer, 2012, Karayanidis et al., 2009, Mondloch, Le Grand & Maurer, 2002). It is possible that age will have a significant effect on face discrimination thresholds measured by the new face test when participants are tested across the full lifespan. Accordingly, caution should be exercised in interpreting the face discrimination thresholds for participants outside of the age range tested in the present study. Future work will establish normative ranges of face discrimination thresholds measured by the Caledonian face test for children and older adults. Although our data are limited to a relatively small sample, we found no effect of sex on discrimination thresholds

4.3 Face Impairment

The Caledonian face test was applied to a patient who experiences a specific difficulty with face perception. LM's face discrimination threshold (14.34%) was far outside of the normal range calculated above (3.39-8.11%). This is in line with the patient's performance on two established face tests which also identified an impairment of face perception. The patient's performance on a non-face object (car) recognition test, however, was normal. This is consistent with the premise of a face-specific deficit, although it remains possible that recognition impairments would have been identified for other non-face objects (e.g. houses, tools, greebles) if they had been tested. Nevertheless, the data presented here indicate that the Caledonian face test is capable of identifying symptomatic impairments of face discrimination.

An important finding of the present study is that the Caledonian face test is likely to prove considerably more sensitive to impairments of face discrimination than currently available tests. LM's Z-scores on the Glasgow face matching and Cambridge face memory tests were -2.20 and -2.27 respectively. The patient's Z-score on the new face test, however, was -7.26. While all three tests identified an impairment of face perception, the Caledonian face

test highlighted a substantially more severe deficit. The enhanced sensitivity of the new face test is likely to be attributable to the test's essentially unlimited testing range. Currently available tests of face perception are constrained by restricted testing ranges and comparatively wide standard deviations. The results of the present study support the premise that the new face test provides a more sensitive assessment of face discrimination sensitivity than that offered by currently available tests. This raises the prospect that the Caledonian face test has the potential to identify more subtle deficits of face perception which existing tests may miss.

References

Allison, T., Puce, A., & McCarthy, G. (2000). Social perception from visual cues: role of the STS region. *Trends Cogn. Sci., 4* (7), 267-278.

Bach, M. (1996). The Freiburg Visual Acuity Test-automatic measurement of visual acuity. *Optom. Vis. Sci., 73* (1), 49-53.

Bailey, I.L., & Lovie, J.E. (1976). New design principles for visual acuity letter charts. *Am. J. Optom. Physiol. Opt., 53* (11), 740-745.

Barnes, C.S., De l'Aune, W., & Schuchard, R.A. (2011). A test of face discrimination ability in aging and vision loss. *Optom. Vis. Sci., 88* (2), 188-199.

Barton, J.J. (2008). Structure and function in acquired prosopagnosia: lessons from a series of 10 patients with brain damage. *J. Neuropsychol., 2* (1), 197-225.

Bate, S., Bennetts, R., Mole, J.A., Ainge, J.A., Gregory, N.J., Bobak, A.K., & Bussunt, A. (2014). Rehabilitation of face-processing skills in an adolescent with prosopagnosia: Evaluation of an online perceptual training programme. *Neuropsychol. Rehabil., 25* (5), 733-762.

Bengtsson, B., & Heijl, A. (1998). SITA Fast, a new rapid perimetric threshold test. Description of methods and evaluation in patients with manifest and suspect glaucoma. *Acta Ophthalmol. Scand., 76* (4), 431-437.

Benton, A.L., Hamsher, K.S., Varney, N., Spreen, O. (1983). Contributions to neuropsychological assessment: a clinical manual. (Oxford: Oxford University Press.

Bird, C.M., Papadopoulou, K., Ricciardelli, P., Rossor, M.N., & Cipolotti, L. (2003). Test-retest reliability, practice effects and reliable change indices for the recognition memory test. *Br. J. Clin. Psychol., 42* (4), 407-425.

Boutet, I., & Faubert, J. (2006). Recognition of faces and complex objects in younger and older adults. *Mem. Cognit., 34* (4), 854-864.

Bowles, D.C., McKone, E., Dawel, A., Duchaine, B., Palermo, R., Schmalzl, L., Rivolta, D., Wilson, C.E., & Yovel, G. (2009). Diagnosing prosopagnosia: Effects of ageing, sex, and participant–stimulus ethnic match on the Cambridge Face Memory Test and Cambridge Face Perception Test. *Cogn. Neuropsychol., 26* (5), 423-455.

Brainard, D.H. (1997). The psychophysics toolbox. *Spat. Vis., 10*, 433-436.

Breiter, H.C., Etcoff, N.L., Whalen, P.J., Kennedy, W.A., Rauch, S.L., Buckner, R.L., Strauss, M.M., Hyman, S.E., & Rosen, B.R. (1996). Response and habituation of the human amygdala during visual processing of facial expression. *Neuron, 17* (5), 875-887.

Bruce, V., Henderson, Z., Greenwood, K., Hancock, P.J., Burton, A.M., & Miller, P. (1999). Verification of face identities from images captured on video. *J. Exp. Psychol. Appl., 5* (4), 339.

Bullimore, M., Bailey, I.L., & Wacker, R.T. (1991). Face recognition in age-related maculopathy. *Invest. Ophthalmol. Vis. Sci., 32* (7), 2020-2029.

Burton, A.M., White, D., & McNeill, A. (2010). The Glasgow face matching test. *Behav. Res. Methods, 42* (1), 286-291.

Caldara, R., Schyns, P., Mayer, E., Smith, M.L., Gosselin, F., & Rossion, B. (2005). Does prosopagnosia take the eyes out of face representations? Evidence for a defect in representing diagnostic facial information following brain damage. *J. Cogn. Neurosci., 17* (10), 1652-1666.

Chaby, L., Narme, P., & George, N. (2011). Older adults' configural processing of faces: Role of second-order information. *Psychol. Aging, 26* (1), 71-79.

Crook, T.H., & Larrabee, G.J. (1992). Changes in facial recognition memory across the adult life span. *J. Gerontol., 47* (3), 138-141.

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

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

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

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

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

Duchaine, B.C. (2000). Developmental prosopagnosia with normal configural processing. *Neuroreport, 11* (1), 79-83.

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

Duchaine, B.C., & Weidenfeld, A. (2003). An evaluation of two commonly used tests of unfamiliar face recognition. *Neuropsychologia, 41* (6), 713-720.

Dutton, G., Ballantyne, J., Boyd, G., Bradnam, M., Day, R., McCulloch, D., Mackie, R., Phillips, S., & Saunders, K. (1996). Cortical visual dysfunction in children: a clinical study. *Eye, 10*, 302-309.

Farah, M.J., Wilson, K.D., Drain, H.M., & Tanaka, J.R. (1995). The inverted face inversion effect in prosopagnosia: Evidence for mandatory, face-specific perceptual mechanisms. *Vision Res., 35* (14), 2089-2093.

Furl, N., Garrido, L., Dolan, R.J., Driver, J., & Duchaine, B. (2011). Fusiform gyrus face selectivity relates to individual differences in facial recognition ability. *J. Cogn. Neurosci., 23* (7), 1723-1740.

Gauthier, I., Tarr, M.J., Moylan, J., Skudlarski, P., Gore, J.C., & Anderson, A.W. (2000). The fusiform “face area” is part of a network that processes faces at the individual level. *J. Cogn. Neurosci., 12* (3), 495-504.

Glen, F.C., Smith, N.D., & Crabb, D.P. (2013). Saccadic eye movements and face recognition performance in patients with central glaucomatous visual field defects. *Vision Res., 82*, 42-51.

Goffaux, V., & Rossion, B. (2006). Faces are" spatial"--holistic face perception is supported by low spatial frequencies. *J. Exp. Psychol. Hum. Percept. Perform., 32* (4), 1023-1039.

Haig, N.D. (1986). Exploring recognition with interchanged facial features. *Perception, 15* (3), 235-247.

Karayanidis, F., Kelly, M., Chapman, P., Mayes, A., & Johnston, P. (2009). Facial identity and facial expression matching in 5-12-year-old children and adults. *Infant and Child Development, 18* (5), 404-421.

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

Kohler, C.G., Walker, J.B., Martin, E.A., Healey, K.M., & Moberg, P.J. (2009). Facial emotion perception in schizophrenia: a meta-analytic review. *Schizophr. Bull., 36*, 1009-1019.

Kress, T., & Daum, I. (2003). Developmental prosopagnosia: A review. *Behav. Neurol., 14* (3-4), 109-121.

Leder, H., & Carbon, C.C. (2006). Face-specific configural processing of relational information. *Br. J. Psychol., 97* (1), 19-29.

Lee, Y., Duchaine, B., Wilson, H.R., & Nakayama, K. (2010). Three cases of developmental prosopagnosia from one family: Detailed neuropsychological and psychophysical investigation of face processing. *Cortex, 46* (8), 949-964.

Loffler, G., Gordon, G.E., Wilkinson, F., Goren, D., & Wilson, H.R. (2005a). Configural masking of faces: Evidence for high-level interactions in face perception. *Vision Res., 45* (17), 2287-2297.

Loffler, G., Yourganov, G., Wilkinson, F., & Wilson, H.R. (2005b). fMRI evidence for the neural representation of faces. *Nat. Neurosci., 8* (10), 1386-1391.

Lott, L.A., Haegerstrom-Portnoy, G., Schnek, M.E., & Brabyn, J.A. (2005). Face recognition in the elderly. *Optom. Vis. Sci., 82* (10), 874-881.

McCulloch, D.L., Loffler, G., Colquhoun, K., Bruce, N., Dutton, G.N., & Bach, M. (2011). The effects of visual degradation on face discrimination. *Ophthalmic Physiol. Opt., 31* (3), 240-248.

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

Näsänen, R. (1999). Spatial frequency bandwidth used in the recognition of facial images. *Vision Res., 39* (23), 3824-3833.

Nunn, J., Postma, P., & Pearson, R. (2001). Developmental prosopagnosia: Should it be taken at face value? *Neurocase, 7* (1), 15-27.

Or, C.C.-F., & Wilson, H.R. (2013). Implicit face prototype learning from geometric information. *Vision Res., 82*, 1-12.

Pelli, D., & Robson, J. (1988). The design of a new letter chart for measuring contrast sensitivity. *Clinical Vision Sciences,* 2 (pp. 187-199).

Pelli, D.G. (1997). The VideoToolbox software for visual psychophysics: Transforming numbers into movies. *Spat. Vis., 10* (4), 437-442.

Quick, R. (1974). A vector-magnitude model of contrast detection. *Kybernetik, 16* (2), 65-67.

Richler, J.J., & Gauthier, I. (2014). A meta-analysis and review of holistic face processing. *Psychol. Bull., 140* (5), 1281-1302.

Rizzo, S., Venneri, A., & Papagno, C. (2002). Famous face recognition and naming test: a normative study. *Neurol. Sci., 23* (4), 153-159.

Robbins, R., & McKone, E. (2007). No face-like processing for objects-of-expertise in three behavioural tasks. *Cognition, 103* (1), 34-79.

Rossion, B. (2008). Picture-plane inversion leads to qualitative changes of face perception. *Acta Psychol. (Amst.), 128* (2), 274-289.

Roudier, M., Marcie, P., Grancher, A.-S., Tzortzis, C., Starkstein, S., & Boller, F. (1998). Discrimination of facial identity and of emotions in Alzheimer's disease. *J. Neurol. Sci., 154* (2), 151-158.

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

Russell, R., & Sinha, P. (2007). Real-world face recognition: The importance of surface reflectance properties. *Perception, 36* (9)

Sorger, B., Goebel, R., Schiltz, C., & Rossion, B. (2007). Understanding the functional neuroanatomy of acquired prosopagnosia. *Neuroimage, 35* (2), 836-852.

Sprengelmeyer, R., Young, A., Mahn, K., Schroeder, U., Woitalla, D., Büttner, T., Kuhn, W., & Przuntek, H. (2003). Facial expression recognition in people with medicated and unmedicated Parkinson's disease. *Neuropsychologia, 41* (8), 1047-1057.

Tejeria, L., Harper, R., Artes, P., & Dickinson, C. (2002). Face recognition in age related macular degeneration: perceived disability, measured disability, and performance with a bioptic device. *Br. J. Ophthalmol., 86* (9), 1019-1026.

Veres-Injac, B., & Persike, M. (2009). Recognition of briefly presented familiar and unfamiliar faces. *Psihologija, 42* (1), 47-66.

Warrington, E.K. (1984). Recognition Memory Test: Test Booklet 1. (

Watson, A.B., & Pelli, D.G. (1983). QUEST: A Bayesian adaptive psychometric method. *Percept. Psychophys., 33* (2), 113-120.

Weigelt, S., Koldewyn, K., & Kanwisher, N. (2012). Face identity recognition in autism spectrum disorders: a review of behavioral studies. *Neurosci. Biobehav. Rev., 36* (3), 1060-1084.

Wilkinson, F., Wilson, H.R., & Habak, C. (1998). Detection and recognition of radial frequency patterns. *Vision Res., 38* (22), 3555-3568.

Wilson, H.R., Loffler, G., & Wilkinson, F. (2002). Synthetic faces, face cubes, and the geometry of face space. *Vision Res., 42* (27), 2909-2923.

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

Yin, R.K. (1969). Looking at upside-down faces. *J. Exp. Psychol., 81* (1), 141-145.