Attention to internal face features in unfamiliar face matching

Kingsley I. Fletcher[1]*, Marcus A. Butavicius[2] and Michael D. Lee[3]

[1]School of Psychology, University of Adelaide, Australia
[2]Defence Science and Technology Organisation, Edinburgh, Australia
[3]Department of Cognitive Sciences, University of California, Irvine, California, USA

Accurate matching of unfamiliar faces is vital in security and forensic applications, yet previous research has suggested that humans often perform poorly when matching unfamiliar faces. Hairstyle and facial hair can strongly influence unfamiliar face matching but are potentially unreliable cues. This study investigated whether increased attention to the more stable internal face features of eyes, nose, and mouth was associated with more accurate face-matching performance. Forty-three first-year psychology students decided whether two simultaneously presented faces were of the same person or not. The faces were displayed for either 2 or 6 seconds, and had either similar or dissimilar hairstyles. The level of attention to internal features was measured by the proportion of fixation time spent on the internal face features and the sensitivity of discrimination to changes in external feature similarity. Increased attention to internal features was associated with increased discrimination in the 2-second display-time condition, but no significant relationship was found in the 6-second condition. Individual differences in eye-movements were highly stable across the experimental conditions.

Unfamiliar face matching is unreliable

The ability to match two face images accurately is critical in security and forensic applications. Yet despite our daily interaction with faces, unfamiliar face-matching performance appears to be far from perfect. Previous studies have examined face-matching performance in retail (Kemp, Towell, & Pike, 1997), forensic (Bruce et al., 1999; Henderson, Bruce, & Burton, 2001) and security (Lee, Vast, & Butavicius, 2006) applications. In each of these cases, participants were asked to decide if two simultaneously presented face images were of the same person or different. The studies reported miss rates (where faces of two different people were identified as being of the same person) that ranged from 13% to 49%, and false-alarm rates (where two face images

of the same person were judged to be of different people) that varied from 10% to 44%. It seems that when two unfamiliar face images are presented simultaneously, an incorrect decision is often reached.

Unfamiliar face matching also appears to be subject to potentially large variations in individual performance. Kemp *et al.* (1997) measured the accuracy of six cashiers in deciding whether a photo on a credit card matched the person presenting the card and found that the overall rate of correct decisions for individual cashiers ranged from 59% to 77%. Although this evidence suggests that there are individual differences in performance, each cashier saw a different set of person/photo pairs, so the characteristics of specific images may have contributed to this variation. However, in a within-participants design, Lee *et al.* (2006) found that individual accuracy levels varied from almost perfect to near chance. In a series of experiments on face identification from video images, Bruce *et al.* (1999) noted that 'there are clearly great individual differences in abilities of participants to perform these matching tasks' (p. 356). Understanding whether these individual differences are stable across a range of images, and why these differences exist is a necessary first step towards improving the accuracy of human performance in a face-matching task.

The high error rates seen in unfamiliar face-matching tasks are not seen in familiar face identification which has far higher accuracy (Bruce, Henderson, Newman, & Burton, 2001), even under conditions of poor image quality (Burton, Wilson, Cowan, & Bruce, 1999). Unfamiliar face matching is affected by a range of differences between the images, including differences in pose, hairstyle, and facial hair (Patterson & Baddeley, 1977; Shepherd, Davies, & Ellis, 1981), expression (Bruce *et al.*, 1999), and lighting (Hill & Bruce, 1996), whereas familiar face identification is relatively insensitive to such differences (Bruce & Young, 1986). There appears to be a robust difference in our ability to process familiar and unfamiliar faces. Performance in matching unfamiliar faces is also affected by which parts of the face are visible. Bruce *et al.* asked participants to match a full-face image with a face image that either had only the internal features (eyes, nose, and mouth) or the external features (hair, chin, and ears) visible. Better matching performance was found for images that had the external features visible than those with the internal features visible. Thus when other image differences such as expression, pose, and lighting are held constant, the external features of hairstyle and facial hair appear to dominate unfamiliar face matching. This is in contrast to the internal feature advantage that exists for familiar faces (Campbell *et al.*, 1999; Young, Hay, McWeeny, Flude, & Ellis, 1985).

Possible source of unfamiliar face matching unreliability

Although external features may dominate face-matching decisions for unfamiliar faces at a group level, it may not be so at an individual level. Unfamiliar face matching is thought to involve 'directed visual processing' (Bruce & Young, 1986), where selective attention is used strategically to gather task-relevant information. Hairstyle and facial hair are features that can be easily changed, and therefore are potentially unreliable cues to identity. People who are accurate at matching unfamiliar faces may be able to direct their attention towards the more reliable cues contained in the internal features of the eyes, nose, and mouth and away from the potentially misleading cues contained in the external features of the hairstyle and facial hair. The ability to direct attention towards internal features in preference to external may explain some of the individual differences in unfamiliar face-matching ability.

Eye-movements as a measure of attention to internal face features

When examining a scene or object, the eyes continually perform a series of fixations and saccades which allow various parts of an image to be seen with high resolution as the visual information is processed (Rayner, 1998). The patterns of saccades and fixations are considered to be task dependent and predominantly under cognitive control (Kowler, 1990; Stacey, Walker, & Underwood, 2005; Walker-Smith, Gale, & Findlay, 1977; Yarbus, 1967). It is possible to voluntarily shift attention without moving the eyes (Posner, 1980), and it may also be possible for a saccade to occur without a shift in attention (Kowler, 1990). However, there appears to be a strong relationship between fixations and attention in natural complex information-processing tasks (J. M. Henderson, 2003; Rayner, 1998) and the analysis of eye-movements can have a heuristic value if they are used to test a specific hypothesis (Viviani, 1990). Thus eye-movements can be considered to provide an indication of visual attention. An increase in the proportion of fixations on the internal features of familiar faces when compared with unfamiliar faces has been reported in a matching task (Stacey *et al.*, 2005), suggesting that eye-movements may provide a useful measure of the presence of an internal feature advantage effect. The current study aims to test whether attention to internal face features facilitates unfamiliar face matching by examining whether the proportion of fixation time on internal face features is correlated with accuracy on an unfamiliar face-matching task.

Possible effect of viewing time

Two viewing time conditions were included to explore the effect of time pressure on eye-movements and face-matching performance. Although eye-movements are indicative of visual attention, it must be noted that the location of eye-gaze does not necessarily directly translate into information processing. Eye-movement patterns are likely to reflect a mix of data gathering, data checking, and possibly periods of no data input. As time pressure may limit information processing to the more important and salient aspects of a task, a moderate time-pressure condition was included in an attempt to maximize the proportion of data-gathering fixations compared with checking fixations, thus improving the sensitivity of the eye-movement data. A low time-pressure condition was included to allow more natural strategies to emerge, but with a finite viewing period to avoid excessively long decision times. It appears that all the necessary information to reach a decision can be achieved in less than 6 seconds as no increase in accuracy was found by Lee *et al.* (2006) when viewing time increased from 6 to 15 seconds. Eye-movement pattern differences also seem to appear early, as the effect on eye-movements of prior exposure to faces is measurable in the first 2 seconds of viewing (Althoff & Cohen, 1999). These results and the results of pilot testing led to the choice of a 2-second display time to exert moderate time pressure while still allowing reasonable accuracy to be achieved, and 6 seconds as the low time-pressure condition.

Manipulation of external feature similarity

External feature similarity was manipulated to examine its effect on matching accuracy and eye-movements. The face-matching performance of people who attend predominantly to internal facial features should be less affected by changes to external features than people who attend more to external features. If attention to the internal features enhances face-matching performance, then a small change in matching accuracy

between the similar and dissimilar external feature conditions should be associated with good overall face-matching performance. Conversely, a large change in performance between the similar and dissimilar external feature conditions would suggest a reliance on external features to reach a decision and should be associated with poor face-matching performance. Rather than editing images to achieve the external feature manipulation, this study used the FERET face database corpus (Phillips, Moon, Rizvi, & Rauss, 2000), a dataset which contained images of people with natural differences in external features. Previous studies have commonly edited the images to investigate the effects of external feature similarity, which, while providing a measure of control, may create an unnatural effect that may artificially influence the saliency of external features and may alert participants to the features being manipulated in the stimuli.

Method

Participants

Participants were 48 first-year psychology students (37 females and 11 males) who received course credit for participation. The mean age was 19.6 years ($SD = 3.65$). Data from five participants (four females, one male) were excluded from analysis for the following reasons: Equipment failure caused the loss of data for one participant, two participants recorded the same decision for almost all trials, indicating a lack of engagement with the study, and two participants had consistently poor eye-movement data capture rates and calibration results.

Stimuli

Images were taken from the Color Facial Recognition Technology (FERET) Database (Phillips *et al.*, 2000). The Color FERET Database contains photographic images of 994 participants taken on up to 15 different dates between 1993 and 1996. The images are passport style and have a resolution of 512×768 pixels. The majority of the images in the database are colour, but were converted to greyscale JPEG files to minimize the effect of changes to background colour on the matching decision. Images were screened to include only those of people aged between 16 and 63, with a neutral or slightly smiling expression, who were not wearing glasses and had no distinguishing facial marks.

Same-person image pairs

As the study aimed to measure face-matching discrimination when the effects of day-to-day variation in appearance were present, only people who had images taken on different days were considered. These images were reviewed for differences in hairstyle or facial hair. If there was a change in style, colour, or length the two images were classified as having dissimilar external features. If the hairstyle, colour, or length were judged to be consistent with expected day-to-day variation, the two images were classified as having similar external features. To establish the inter-rater reliability of these subjective classifications, the first author and a person not associated with the project independently classified the images. Agreement between judges was achieved for 86% of images with $\kappa = .65$, which is considered substantial agreement (Landis & Koch, 1977). Classification of images for which agreement was not reached was by discussion with the second author.

Twenty image pairs with dissimilar and twenty image pairs with similar external features were identified. There were an equal number of male and female image pairs; 68% Caucasian, 20% Asian, 8% African-American, and 5% Hispanic. Examples of a same-person image pair with similar and dissimilar external features are shown in Figure 1.
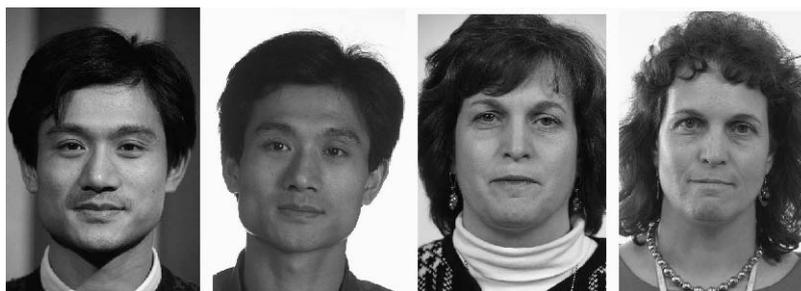
Different-person image pairs
Image pairs of different people who had similar face shape and structure were identified. These were classified as having either similar or dissimilar external features using the criteria described above. The different-people image pairs in each feature category were ranked in the order of similarity by the first author, and the most similar image pairs that matched the sex and ethnicity distribution of the same-person image pairs were selected. An example of image pairs of different people with similar and dissimilar external features is shown in Figure 2.

Cropping and display of images
The images were cropped to achieve a consistent face size and position so that the face and hair filled most of the image. Image pairs were displayed on a computer monitor such that each image had a size of 206 mm height and 138 mm width, with a 17 mm space between. The participant's eyes were positioned 430 mm in front of the monitor and each image subtended a visual angle of approximately 26 degrees height and 18 degrees width. Pilot testing suggested that this was close to the maximum range of eye-movements that could be captured reliably by the eye tracker. This is also similar to the visual angle subtended when viewing a real face at normal conversational distances of around 1 m.

Definition of facial regions
Each face was divided into seven regions of interest: hair, eyes, nose bridge, nose tip, mouth, chin, and ears. Each region was defined for each face individually due to differences in facial structure. Figure 3 demonstrates how the regions were defined. The eyes, nose bridge, nose tip, and mouth were considered to be internal face features. The hair, chin, and ears were considered to be external face features.


Figure 1. Example of same person image pairs with similar external features (left) and dissimilar external features (right).

Figure 2. Example of different-person image pairs with similar external features (left) and dissimilar external features (right).

Apparatus

The eye tracker used was a ViewPoint EyeTracker PC 60 version 2.6, which uses an infrared light source to track corneal reflections from the right eye at a sampling rate of 60 Hz. It was installed on a PC running Windows NT4.0, with a 1.4 GHz processor, 512 Mb RAM, and a Matrox PowerDisk V4.73 graphics driver. A 17″, 75 Hz Hitachi CM772 SVGA monitor with 1,024 × 768 resolution was used. A chin and forehead support bracket provided with the eye tracker was used to stabilize the head position.
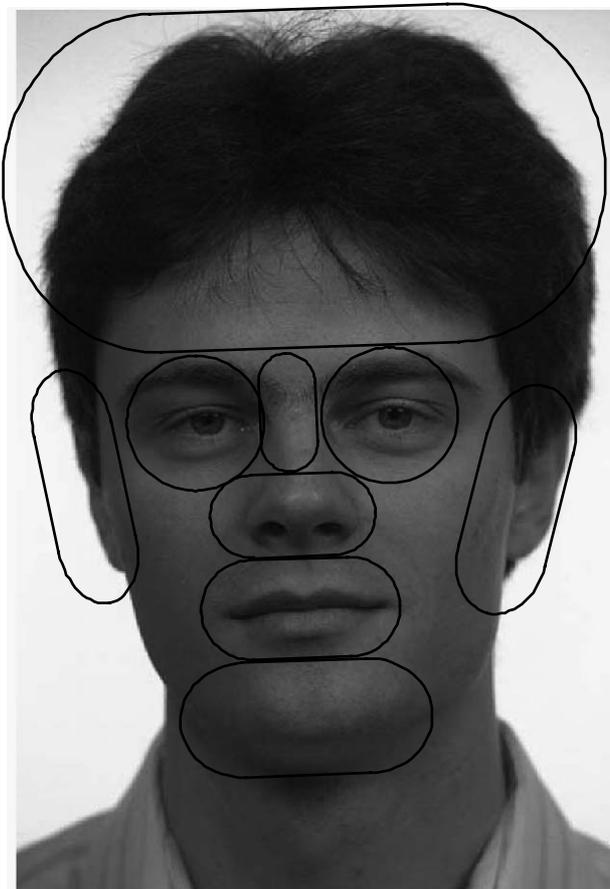
Figure 3. An example of face region classification.

Software was developed which allowed fixations to be identified based on velocity and duration criteria. Minimum fixation durations are rarely less than 100 ms (Rayner, 1998; Salvucci & Goldberg, 2000; Viviani, 1990), and 70 ms was used as the low threshold for fixation duration based on inspection of the data. Velocity within fixations is significantly different from saccades, and a static criterion based on exploratory data analysis can provide a robust measure (Salvucci & Goldberg, 2000). The angular velocity threshold was chosen to be 55 degrees per second, based on inspection of the obtained eye-movement data.

Procedure

Participants sat at the eye tracker which was adjusted for comfort. The task was explained to them but the hypothesis was not revealed. Each trial was initiated by the participant, and began with a sequential display of circular points at the four outside corners of the frame where the image pairs were to be displayed followed by a point at the centre of the screen. Each point was displayed for 1 second, which allowed the direction of gaze to be aligned with the image corners for each trial, and the centre point provided a consistent starting-point for each participant's gaze.

An image pair was then displayed 300 ms after presentation of the centre calibration point and remained on the screen until either the participant reached a same/different decision or the display time elapsed, whichever occurred first. At this point, the image pair was removed from the screen and eye-movement recording was stopped. Participants indicated a same decision by single-clicking the mouse button. Double-clicking the mouse button indicated a different decision. Accuracy and the time taken to reach the decision were recorded.

The screen then displayed the decision and required the participant to use the mouse pointer to indicate the confidence in their decision. A vertically presented 5-point Likert scale was used with options ordered from bottom to top of 'very low', 'low', 'moderate', 'high', and 'very high'. A 'wrong' button was provided at the bottom of the screen, which allowed participants to indicate if they had made a mistake with the mouse buttons. After recording confidence, the participant clicked on a screen button to proceed to the next trial.

Trials were grouped into two blocks, each with a maximum image pair display time of either 2 or 6 seconds. Each image pair was displayed for the same duration to all participants. A tone was sounded 1 second before the end of the display period. The order of presentation of blocks was counterbalanced between participants. Each block consisted of 84 trials, with trials 1–4 presenting one practice image pair from each category (same-person similar external features, same-person dissimilar external features, different-people similar external features, and different-people dissimilar external features) and trials 5–84 presenting 20 image pairs from each category in random order. In addition to being calibrated at the start of each experimental session, the eye tracker was recalibrated before trials 1, 30, and 60 of each block. Participants were given the opportunity to rest prior to recalibration and an experimental session lasted approximately 45 minutes.

Results

Accuracy, response time, and eye-movement data are reported here. All results reported are at the level of the individual participant, with each participant's score calculated as

the mean of the 20 trials in each condition. It would be equally possible to perform the analysis at the level of individual trials averaged across all participants; however, as the primary aim of the current study was to explore possible reasons for individual differences in face-matching accuracy, the performance of each participant across trials in each condition is reported here. A significance level of $p = .05$ was used for all statistical tests.

Review of data for errors and outliers

The data were inspected to check for errors and outliers in both the eye tracking and response time results. Trials were excluded if the eye tracker failed to capture any fixations (8% of trials), if the error of any of the alignment points was extreme (4% of trials), or if a response was not made within 2.5 seconds after the images were removed (3% of trials). Eye-tracker data capture failure was identified by visual inspection, and exclusion of alignment and response time outliers was based on individual trial values exceeding two standard deviations from the mean. The combination of these three criteria resulted in the exclusion of 684 trials (10%).

When a participant indicated a wrong decision in the confidence rating scale, the recorded same/different decision was changed prior to calculating decision accuracy. Participants selected 'wrong' in only 2% of all trials.

Relationship between attention to internal face features and discrimination

As measured by proportion of fixation time on internal face features
Overall, fixations occurred in one of the defined regions for 87% ($SD = 3.87\%$) of the total time spent fixating. The internal face features were fixated most and accounted for 80% ($SD = 9.14\%$) of total fixation time. In contrast, the external features received relatively few fixations, accounting for 7% ($SD = 6.31\%$) of fixation time. The location of the 13% ($SD = 3.86\%$) of fixations spent outside the defined regions did not appear to have a clear pattern, and was a combination of locations close to one of the defined regions, on the cheeks, or at unusual locations such as between the two images, or not on a face.

It was predicted that the ability of participants to discriminate same from different face pairs would be positively correlated with the proportion of fixation time on internal face features. A non-parametric form of discrimination from signal detection theory, A', was used (Stanislaw & Todorov, 1999) and the correlations between discrimination and proportion of fixation time on the internal features are shown in Table 1. For the

Table 1. Correlation between proportion of fixation time on internal features and discrimination (A') (95th percentile confidence intervals are in parentheses)

| External feature condition | Display time condition | | |
	2 seconds	6 seconds	All
Similar external features	.260 (−.044, .520)	−.050 (−.345, .254)	.073 (−.232, .365)
Dissimilar external features	.339* (.043, .580)	.013 (−.288, .312)	.205 (−.102, .476)
All external features	.341* (.045, .582)	−.042 (−.338, .262)	.144 (−.163, .426)

*$p < .05$.

2-second condition a significant correlation was evident between mean percentage of fixation time on the internal features and mean discrimination. However, for the 6-second condition this correlation was close to zero and not significant. Overall, the correlation was positive but non-significant for the dissimilar external feature conditions and near-zero for the similar external feature condition.

As measured by sensitivity to changes in external feature similarity

In addition to measuring eye-movements, the effect of manipulating external feature similarity was also used as an indicator of attention to internal features. It was hypothesized that increased attention to internal features would reduce the effect of changes to external feature similarity on discrimination. If the level of attention to internal features improves face-matching accuracy, then the change in discrimination between similar and dissimilar external feature face pairs should be negatively correlated with discrimination ability. In the 2-second display-time condition, a significant negative correlation was found between discrimination and the absolute change in discrimination between similar and dissimilar external feature conditions, $r(41) = -.360$, $p = .018$, ($CI_{95\%} = -.596, -.067$). However, the relationship was not significant in the 6-second display-time condition, $r(41) = -.219$, $p = .158$, ($CI_{95\%} = -.487, .087$).

Individual differences in fixation on internal features

Mean proportion of fixation time spent on internal features over all conditions varied from a minimum of 50% to a maximum of 90%. This suggests that participants used a variety of fixation strategies in the face-matching task, and this was confirmed by visual observation of raw fixation data. Despite the use of different fixation strategies between participants, each participant appeared to exhibit a relatively stable fixation strategy across conditions. A significant correlation was obtained between the relative proportion of fixation time that each participant made on the internal features in the 2- and 6-second display-time conditions, $r(41) = .689$, $p < .001$, ($CI_{95\%} = .490, .820$), between the similar and dissimilar external feature display-time condition $r(41) = .937$, $p < .001$, ($CI_{95\%} = .886, .966$), and between same-person and different-person image pairs, $r(41) = .934$, $p < .001$ ($CI_{95\%} = .881, .964$). Participants exhibited highly stable fixation patterns across all conditions.

Effect of display time and external feature similarity

Accuracy, discrimination, and bias

The mean percentage accuracy for each condition is shown in Table 2. These were converted into non-parametric forms of discrimination and bias, A' and B'' (Stanislaw & Todorov, 1999), which are shown in Table 3.

A 2×2 (display time × external feature similarity) repeated measures analysis of variance was performed on discrimination and bias. Discrimination was significantly higher in the 6-second display-time condition when compared with the 2-second display-time condition, $F(1, 42) = 4.57$, $p = .038$, $\eta_p^2 = .098$. There was no significant difference in discrimination between the similar and dissimilar external feature conditions, $F(1, 42) = 0.34$, $p = .564$, $\eta_p^2 = .008$ and the interaction was not significant, $F(1, 42) = 0.033$, $p = .857$, $\eta_p^2 = .001$.

Table 2. Mean percentage accuracy, response time, and fixation length for same- and different-person image pairs in each experimental condition (standard deviations are in parentheses)

	2-second display time		6-second display time		Total	
	Same person	Different person	Same person	Different person	Same person	Different person
Mean percentage accuracy						
Similar external features	85.10 (14.14)	83.08 (13.69)	87.17 (13.31)	87.19 (12.09)	86.32 (11.21)	84.91 (10.83)
Dissimilar external features	81.91 (14.66)	85.41 (10.62)	84.03 (12.52)	89.39 (13.03)	82.68 (12.38)	87.29 (10.47)
Total	83.57 (13.11)	84.29 (10.63)	85.57 (12.18)	88.30 (11.81)	84.47 (11.29)	86.15 (9.89)
Mean response time (s)						
Similar external features	2.06 (0.68)	2.10 (0.68)	3.34 (1.35)	3.33 (1.23)	2.72 (0.93)	2.73 (0.88)
Dissimilar external features	2.16 (0.67)	2.01 (0.65)	3.67 (1.26)	3.40 (1.26)	2.95 (0.88)	2.72 (0.87)
Total	2.11 (0.67)	2.05 (0.66)	3.50 (1.29)	3.36 (1.23)	2.83 (0.90)	2.72 (0.87)
Mean fixation length (ms)						
Similar external features	229.49 (35.42)	237.42 (37.06)	242.79 (31.69)	253.49 (35.79)	236.35 (30.30)	245.18 (34.10)
Dissimilar external features	227.58 (31.49)	239.34 (34.77)	247.80 (32.65)	255.30 (32.69)	237.54 (30.51)	247.44 (31.79)
Total	228.53 (31.99)	238.42 (35.23)	245.32 (31.57)	254.32 (33.59)	236.96 (29.92)	246.31 (32.58)

Table 3. Mean discrimination (A′) and bias (B″) for each experimental condition (standard deviations are in parentheses)

	2-Second display time		6-Second display time		Total	
	A′	B″	A′	B″	A′	B″
Similar external features	0.90 (0.07)	0.07 (0.47)	0.92 (0.07)	0.01 (0.42)	0.91 (0.06)	0.06 (0.44)
Dissimilar external features	0.89 (0.08)	−0.03 (0.35)	0.92 (0.07)	−0.18 (0.45)	0.91 (0.07)	−0.12 (0.41)
Total	0.90 (0.07)	0.04 (0.42)	0.92 (0.07)	−0.12 (0.49)	0.91 (0.06)	−0.02 (0.40)

Bias exhibited the opposite pattern of results. There was no significant difference between the 2- and 6-second display-time conditions, $F(1, 42) = 0.804$, $p = .375$, $\eta_p^2 = .019$. However, participants showed a significant difference in bias between the similar and dissimilar external feature conditions. Similar external feature image pairs produced a bias towards the same person decision compared to dissimilar external feature image pairs which yielded a bias towards a different person decision, $F(1, 42) = 9.38$, $p = .004$, $\eta_p^2 = .183$. Again the interaction was not significant, $F(1, 42) = 0.117$, $p = .735$, $\eta_p^2 = .003$.

Response time

The mean response times for correct responses in each condition are shown in Table 2. A $2 \times 2 \times 2$ (display time × same/different person × external feature similarity) repeated measures ANOVA revealed that response time was significantly longer in the 6-second condition when compared with the 2-second condition, $F(1, 42) = 85.57$, $p < .001$, $\eta_p^2 = .671$. There was no significant difference in response time between same- and different-person image pairs, $F(1, 42) = 3.11$, $p = .085$, $\eta_p^2 = .069$, but image pairs with dissimilar external features had longer reaction times than image pairs with similar external features, $F(1, 42) = 27.88$, $p < .001$, $\eta_p^2 = .399$. There was a significant display time by feature interaction, $F(1, 42) = 21.76$, $p < .001$, $\eta_p^2 = .341$, and a significant person by feature interaction, $F(1, 42) = 17.15$, $p < .001$, $\eta_p^2 = .290$. Response time for dissimilar feature image pairs increased in the 6-second display time but not the 2-second display time condition. Response time also increased in the dissimilar feature condition only for same-person image pairs, but not different. The display time by person interaction and the display time by person by feature interaction were not significant, $F(1, 42) = 1.20$, $p = .280$, $\eta_p^2 = .028$ and $F(1, 42) = .71$, $p = .404$, $\eta_p^2 = .017$, respectively.

In order to test for a possible speed–accuracy trade off, a correlation was performed between the mean response time and mean discrimination of each participant for each viewing time condition. In the 2-second condition, response time and discrimination were positively correlated, $r(41) = .367$, $p = .016$ ($CI_{95\%} = .075, .601$), with longer response times associated with higher discrimination. The correlation was not significant in the 6-second condition, $r(41) = .164$, $p = .293$, ($CI_{95\%} = -.143, .443$).

Mean fixation length

The mean length of each fixation in each condition is shown in Table 2. A $2 \times 2 \times 2$ (display time × same/different person × external feature similarity) repeated measures ANOVA found that mean fixation length was significantly longer in the 6-second display-time condition when compared with the 2-second display-time condition, $F(1, 42) = 26.55$, $p < .001$, $\eta_p^2 = .387$, and also longer for different-person image pairs than same-person image pairs, $F(1, 42) = 59.40$, $p < .001$, $\eta_p^2 = .586$. There was no significant difference in mean fixation length between the similar and dissimilar external feature conditions, $F(1, 42) = 2.12$, $p = .153$, $\eta_p^2 = .048$. None of the interactions were significant.

Discussion

Attention to internal face features is associated with improved discrimination with short viewing times

This study tested the hypothesis that increased attention to the internal face features (eyes, nose, and mouth) would be associated with an increased ability to discriminate

same and different face pairs, as these features are more stable than the external features (hair, chin, and ears) which can be more readily altered. It examined people's ability to decide whether pairs of faces that exhibited natural day-to-day variation in appearance were of the same person or different. The face pairs had either similar or dissimilar hairstyle, or facial hair, and were displayed for either 2 or 6 seconds. The extent of attention to internal face features was measured by the proportion of fixation time on the internal features and whether discrimination ability differed between face pairs with similar and dissimilar external features.

Both measures supported the predicted relationship in the 2-second viewing condition, but not in the 6-second. Thus under limited viewing time, participants were better able to discriminate same from different face pairs when increased attention was paid to internal features. The size of the effect was only moderate, accounting for approximately 12% of individual variance in discrimination, but it was reliably produced in both the measures used to identify the level attention paid to internal features.

Previous studies have established that internal face features play a greater role in familiar face matching than in unfamiliar face matching (Young *et al.*, 1985). It has also been proposed that the process of familiarization may involve increasing the ability to encode the internal features of a face, which leads to better matching performance for familiar faces when compared with unfamiliar faces (Clutterbuck & Johnston, 2005). The results of the current study support the role that internal features play in facilitating face matching, and also suggest that internal features facilitate identification and matching of unfamiliar as well as familiar faces.

This is not to say that external features are not salient when matching unfamiliar faces. The manipulation of external feature similarity had a significant effect on matching accuracy at a group level, with dissimilar external features creating a bias towards identifying the face pairs as being different people, but having no effect on discrimination. Again this result is consistent with previous studies which found that external features significantly affected the ability to match unfamiliar faces (e.g. Patterson & Baddeley, 1977), but suggests that external feature differences appear to have more of an affect on the matching decision criterion rather than change what aspects of unfamiliar faces are encoded.

However, at an individual level it appears that the effect of external features can be reduced by directing attention towards the internal features. For image pairs with dissimilar external features viewed under limited viewing time, discrimination was significantly correlated with the proportion of fixation time spent on internal features. In addition, the relationship between attention to internal features and discrimination was stronger for image pairs with dissimilar external features than for image pairs with similar external features, for which the correlation did not achieve statistical significance.

No association found between attention to internal face features and discrimination with longer viewing time

The predicted relationship between attention to internal face features and discrimination was only evident under the 2-second display-time condition and not in the 6-second display-time condition. This result may have been caused by the different amounts of information that could be processed in each display-time condition. Discrimination was generally higher when more viewing time was available. Discrimination was higher in the 6-second display-time condition than it was in the 2-second display-time condition, and in the 2-second condition faster responses were associated with lower discrimination,

indicating a speed–accuracy trade off. Although ability can confound a between-participants speed–accuracy correlation, this correlation is in the opposite direction to that expected if it was due to participants with higher ability taking less time to reach a decision. Thus it supports the conclusion that longer viewing times allowed higher discrimination, although not to a large extent as accuracy only increased by 3% between the 2- and 6-second display-time conditions. However, the necessary visual information may be able to be collected in under 6 seconds, as the mean response time in the 6-second display-time condition was approximately 3.5 seconds.

These results suggest that 2 seconds of viewing time may not have allowed for complete face processing, and that given a limited viewing time prioritizing attention to internal face features over external face features was associated with better discrimination. However, when more time was available a more complete processing of other facial aspects could be performed. This view that unfamiliar face matching is not a rapid, automatic process appears to be consistent with previous findings that people appear to have a limited capacity to process unfamiliar faces (Bindemann, Burton, & Jenkins, 2005; Boutet & Chaudhuri, 2001; Palermo & Rhodes, 2002). The additional processing led to improved discrimination, but also meant that internal face feature processing may have made a relatively lower contribution to discrimination performance. This would lead to the result that attention to internal face features was not a significant predictor of discrimination if 6 seconds of viewing time was allowed. Thus attention to internal features appears to contribute to effective face matching, but face matching appears to be supplemented and improved by processing of additional face aspects if time permits.

Conclusion

Understanding how people are able to recognize and discriminate faces is an important topic for cognitive science. Face processing is a core human activity, and an understanding of the process involved, and its limitations, has great potential for application to a variety of real-world problems. Previous studies have demonstrated wide variation in individual abilities to accurately match unfamiliar faces. The current study provides further evidence that substantial variation in individual performance exists, and also demonstrates that eye-movement patterns vary substantially between people, but that individuals exhibit remarkably stable eye-movement patterns across a range of diverse face images. Thus the manner in which people process unfamiliar faces appears to vary widely, but individuals seem to be highly consistent in the face aspects that they attend to. The current study also provides results that suggest unfamiliar face matching is a decision process requiring several seconds to achieve maximum performance. Greater attention to the internal features appears to improve discrimination of unfamiliar faces, but other aspects of face processing are also clearly involved and attention to internal features accounts for only a modest amount of individual variation. However, individual differences in unfamiliar face processing and matching ability appear to be highly stable, which offers encouragement that it may be possible to identify the more important sources of variation.

Acknowledgements

References

Althoff, R. R., & Cohen, N. (1999). Eye-movement-based memory effect: A reprocessing effect in face perception. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *25*(4), 997–1010.

Bindemann, M., Burton, A. M., & Jenkins, R. (2005). Capacity limits for face processing. *Cognition*, *98*, 177–197.

Boutet, I., & Chaudhuri, A. (2001). Multistability of overlapped face stimuli is dependent upon orientation. *Perception*, *30*, 743–753.

Bruce, V., Henderson, Z., Greenwood, K., Hancock, P. J. B., Burton, M. A., & Miller, P. (1999). Verification of face identities from images captured on video. *Journal of Experimental Psychology: Applied*, *5*(4), 339–360.

Bruce, V., Henderson, Z., Newman, C., & Burton, M. A. (2001). Matching identities of familiar and unfamiliar faces caught on CCTV images. *Journal of Experimental Psychology: Applied*, *7*(3), 207–218.

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

Burton, A. M., Wilson, S., Cowan, M., & Bruce, V. (1999). Face recognition in poor quality video: Evidence from security surveillance. *Psychological Science*, *10*, 243–248.

Campbell, R., Walker, J., Benson, P. J., Wallace, S., Coleman, M., Michelotti, J., *et al.* (1999). When does the inner-face advantage in familiar face recognition arise – and why? *Visual Cognition*, *6*, 196–216.

Clutterbuck, R., & Johnston, R. A. (2005). Demonstrating how unfamiliar faces become familiar using a face matching task. *European Journal of Cognitive Psychology*, *17*(1), 97–116.

Henderson, J. M. (2003). Human gaze control during real-world scene perception. *Trends in Cognitive Science*, *7*(11), 498–504.

Henderson, Z., Bruce, V., & Burton, M. A. (2001). Matching the faces of robbers captured on video. *Applied Cognitive Psychology*, *15*, 445–464.

Hill, H., & Bruce, V. (1996). Effects of lighting on the perception of facial surfaces. *Journal of Experimental Psychology: Human Perception and Performance*, *22*, 986–1004.

Kemp, R., Towell, N., & Pike, G. (1997). When seeing should not be believing: Photographs, credit cards and fraud. *Applied Cognitive Psychology*, *11*, 211–222.

Kowler, E. (1990). The role of visual and cognitive processes in the control of eye movement. In E. Kowler (Ed.), *Eye movements and their role in visual and cognitive processes* (pp. 1–70). Amsterdam: Elsevier.

Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, *33*, 159–174.

Lee, M. D., Vast, R. L., & Butavicius, M. A. (2006). *Face matching under time pressure and task demands*. In R. Sun (Ed.), *Proceedings of the 28th Annual Conference of the Cognitive Science Society* (pp. 1675–1680). Mahwah, NJ: Erlbaum.

Palermo, R., & Rhodes, G. (2002). The influence of divided attention on holistic face perception. *Cognition*, *82*, 225–257.

Patterson, K. E., & Baddeley, A. D. (1977). When face recognition fails. *Journal of Experimental Psychology: Human Learning and Memory*, *4*(4), 406–417.

Phillips, P. J., Moon, H., Rizvi, S. A., & Rauss, P. J. (2000). The FERET evaluation methodology for face recognition algorithms. *IEEE Transactions on Pattern Analysis and Machine Intelligence.*, *22*, 1090–1104.

Posner, M. I. (1980). Orientating of attention. *Quarterly Journal of Experimental Psychology, 32*, 3–25.

Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin*, *124*(3), 372–422.

Salvucci, D., & Goldberg, J. (2000). *Identifying fixations and saccades in eye-tracking protocols*. In *Proceedings of the Eye Tracking Research and Applications Symposium* (pp. 71-78). New York: ACM Press.

Shepherd, J. W., Davies, G. M., & Ellis, H. D. (1981). Studies of cue saliency. In G. M. Davies, H. D. Ellis, & J. W. Shepherd (Eds.), *Percieving and remembering faces* (pp. 105–132). London: Academic Press.

Stacey, P. C., Walker, S., & Underwood, J. D. M. (2005). Face processing and familiarity: Evidence from eye-movement data. *British Journal of Psychology, 96*, 407–422.

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

Viviani, P. (1990). Eye movements in visual search: Cognitive, perceptual and motor control aspects. In E. Kowler (Ed.), *Eye movements and their role in visual and cognitive processes* (pp. 353–393). Amsterdam: Elsevier.

Walker-Smith, G. J., Gale, A. G., & Findlay, J. M. (1977). Eye movement strategies involved in face perception. *Perception, 6*, 313–326.

Yarbus, A. L. (1967). *Eye Movements and Vision (B. Haigh, Trans.)*. New York: Olenum Press.

Young, A. W., Hay, D. C., McWeeny, K. H., Flude, B. M., & Ellis, H. D. (1985). Matching familiar and unfamiliar faces on internal an external features. *Perception, 14*, 737–746.