Matching face images taken on the same day or months apart:

The limitations of photo-ID

Ahmed M. Megreya [1&2], Adam Sandford [3], & A. Mike Burton [3]

[1] Department of Psychology, Menoufia University, Egypt

[2] University College in Qunfodah, Umm Al-Qura University, KSA

[3]School of Psychology, Aberdeen University, UK

Please address correspondence to:
A. Mike Burton,
School of Psychology,
Aberdeen University,
Aberdeen AB24 3FX,
UK
Tel: +44 1224 273930
Email: m.burton@abdn.ac.uk

Word count (excluding abstract, references, tables, and figure legends): 3743

Abstract

It is well-established that matching images of unfamiliar faces is rather error prone. However, there is an important mismatch between face matching in laboratory and realistic settings. All of the currently available face-matching databases were designed to establish the baseline level of unfamiliar face perception. Therefore, target and test images for each face identity have been taken on the same day, minimising within-face variations. In realistic settings, on the other hand, faces do vary, even day to day. This study examined the proficiency of matching images of unfamiliar faces, which were taken on the same day or months apart. In two experiments, same-day images were matched substantially more accurately and faster than different-date photographs using the standard 1-in-10 and 1-in-1 face matching tasks. This suggests that experimental studies on face matching underestimate its difficulty in real-world situations. Photographs of *unfamiliar* faces seem to be unreliable proofs of identity, especially if the ID documents do not use very recent images of the holders.

Keywords: face matching; recent/old photos; within-face variations.

Introduction

It has been known for many years that eyewitness identification is highly fallible (Borchard, 1932; Münsterberg, 1908). A very large number of studies have addressed the sources of mistaken identification (for reviews, see Memon, Vrij, & Bull, 2003; Narby, Cutler, & Penrod, 1996; Wells, Wright, & Bradfield, 1999), and many of these have focussed on imperfect *memory* as a source of errors. However, a recent research trend suggests that a significant part of eyewitness misidentifications involves problems of unfamiliar face perception in the first place (Megreya & Burton, 2008). This proposal is based on face matching experiments, which included optimal situations for exploring the baseline level of unfamiliar face identification, without constraints of long term memory (for reviews, see Burton & Jenkins, 2011; Jenkins & Burton, 2008; Megreya, 2012).

In one notable study, Bruce and colleagues (1999) asked participants to match a target face to a 10-face line-up, which might or might not involve a different image of the target himself (see Figure1). There was no time pressure to complete this task, and all images were high quality, taken on the same day, and well matched for viewing angle, lighting and pose. However, the level of performance was very low, with error rates of roughly 30% in both target-present and target-absent line-ups. Poor performance on this 1-in-10 face-matching task has now been replicated many times using a range of stimuli (Bindemann, Sandford, Gillatt, Avetisyan, & Megreya, 2012; Davies & Thasen, 2000; Konar, Bennett, & Sekuler, 2010; Megreya & Bindemann, in press; Megreya & Burton, 2006a&b; 2007; 2008; Megreya, White, & Burton, 2011).

---------- FIGURE1 HERE PLEASE ----------

The difficulty of unfamiliar face matching has been also replicated under different task constraints, for example when the targets were always present in the 10-face line-ups (Bruce et al., 1999; Burton, Miller, Bruce, Hancock, & Henderson, 2001), or when the numbers of faces in the line-ups was reduced from ten to eight (Henderson, Bruce, & Burton, 2001), five (Megreya, Bindemann, Havard, & Burton, 2012), or only two (Henderson et al., 2001). Furthermore, performance on a seemingly simple match/mismatch task (where participants were given two face images and asked whether they showed the same person or not) is also rather low. Many experiments have replicated this difficulty of matching only two images of unfamiliar faces using a variety of databases (Bindemann, Avetisyan & Blackwell, 2010; Bruce, Henderson, Newman, and Burton, 2001; Burton, White & McNeill, 2010; Megreya & Bindemann, 2009; Megreya, Bindemann, & Havard, 2011; Megreya & Burton, 2006b, 2007, 2008; Özbek & Bindemann, 2011; Strathie, McNeill, & White, 2012). These poor levels of performance persist when viewers are asked to match a photo to a live person, as in real ID-check settings (Davis & Valentine, 2009; Kemp, Towell & Pike, 1997; Megreya & Burton, 2008). Evidence seems to indicate that face-matching ability is quite variable among different observers (Bindemann, Brown, Koyas, & Russ, 2012; Burton et al., 2010; Megreya & Bindemann, in press; Megreya & Burton, 2006b), but fairly consistent within participants (Bindemann, Avetisyan, & Rakow, 2012).

In addition to the practical importance of face matching tasks, they provide considerable theoretical contributions to the face recognition literature. For example, they have been used to replicate several face processing phenomena, which were previously thought to occur during memory, such as the inversion effect (Megreya & Burton, 2006b), the female same-sex bias (Megreya, Bindemann, & Havard, 2011),

and the other-race effect (Megreya, White, & Burton, 2011). The well-established leftward bias for face processing was also replicated (Megreya & Havard, 2011). Furthermore, face matching has been used to highlight the dissociation between familiar and unfamiliar face processing (Clutterbuck & Johnston, 2002; 2004; 2005). For example, Megreya and Burton (2006b) found no correlation between participants' performance on familiar and unfamiliar face matching tasks. Importantly however, when these tasks were presented upside down, viewers' ability to match familiar and unfamiliar inverted faces correlated strongly. Accordingly, we suggested that the processes involved in upright unfamiliar face processing appear to be qualitatively similar to those underlying the recognition of inverted familiar and unfamiliar faces, but different from those responsible for upright familiar face processing. Consistent with this suggestion, Megreya and Burton (2007) found that a short familiarization procedure was successful in producing the *expected* negative correlation between hits and false positives in matching upright unfamiliar faces. However, this so-called "mirror effect" disappeared when faces were turned upside down. Together, faces in the unfamiliar face-matching task seem to be treated as "images" or "simple visual patterns", and matched on this basis without domain-specific expertise (Megreya & Burton, 2006b, see also Hancock, Bruce, & Burton, 2000). This suggestion was further supported by a subsequent study (Konar et al., 2010), where holistic processing was not correlated with unfamiliar face matching.

However, there is an important mismatch between face matching in the laboratory and realistic settings. All of the currently available face-matching databases (e.g. Bruce et al., 1999; Burton et al., 2010; Megreya & Burton, 2008) were designed to reveal the best results that could be achieved by observers. So, photos for matching have always been taken on the same day, eliminating transient differences

due to changes in (for example) hair, weight, or health. In fact, in realistic settings, the general appearance of faces does change, even day to day. This suggests that laboratory face-matching experiments might underestimate the problems of face matching in real-world situations, where same-day photos are rarely used. For example, people can use their passports as long as they are valid, often for many years.

Using a rather limited set of face identities, Davis and Valentine (2009) asked participants to match recent or week-old video images of unfamiliar faces to static photographs or physically-present people. Face matching was more error-prone in the "old" condition, regardless of whether the targets were presented in static images or live. In addition, Davis and Valentine (2009) examined how well participants could match one year-old images of unfamiliar faces to live people when videos were presented without disguise or while wearing a hat or dark glasses. Participants were or were not given a warning that the videos had been taken one year previously. Interestingly, this warning did not generally affect the accuracy of face matching. However, wearing glasses produced fewer matching errors than the no disguise condition (but see Hockley, Hemsworth & Consoli, 1999; Terry, 1994), and these two conditions did not differ from the hat condition (but see Henderson et al., 2001; Megreya et al., 2011).

In this paper, we report two experiments, aiming to reduce the gap between laboratory and realistic settings. In Experiment 1, we used a 1-in-10 face-matching task (e.g., see Bruce et al., 1999), where the target *and/or* (IS THIS REALLY INCORRECT?) the line-up image were taken on the same day or several months apart. In Experiment 2, the 1-in-10 arrays were reduced to face-matching pairs, similar to Davis and Valentine (2009), but with a much larger database of faces.

Experiment 1

Method

Participants

Eighty undergraduate students (47 females, 33 males) from Menoufia University (Egypt) participated in this experiment. Participants' age ranged from 18-21 years, with an average of 19.7 years (SD = 0.6). All participants reported normal or corrected-to-normal vision.

Stimuli & procedure

This experiment uses 1-in-10 face matching line-ups (see figure 2). Participants' task was to judge whether the target face was present among the 10-face array, and if so, to choose the match. Consistent with previous research, target and array photos were taken with two different cameras (to eliminate image matching), though pose was very similar, and all images were of high quality.

---------- FIGURE 2 HERE PLEASE ----------

Stimuli were constructed from 4 photos of each of 56 young men. Half of these stimuli (Time 1 images) were photos taken from the Egyptian face database, constructed by Megreya and Burton (2008). Each person was photographed twice, on the same day, using two different cameras (photos A and B). Photos A were taken with a high-quality digital camera, wheras photos B were stills taken from a high-

quality camcorder. The images were matched for lighting, pose, and viewing conditions; see Megreya and Burton (2008) for full description of this database. The other half of the stimuli (Time 2 images) were photos of the same 56 young men taken several months later (mean 17.2 months, sd 7.3). Once again, each person was photographed twice, using the same cameras (A and B) as used in the original session. Once again, images were matched for lighting, pose and viewing conditions. All photos showed young, clean-shaven Egyptian men, in full-face view and with a neutral facial expression. They measured approximately 5 x 7 cm and were shown in grayscale.

1-in-10 matching arrays were constructed from these photos (see Figure 2). The 10-face arrays were always constructed from Camera A photos, while target faces were always Camera B photos. Two sets of same-time 1-in-10 line-ups were constructed (both Target and Array photos taken either at Time 1 or Time 2). Two sets of different-time line-ups constructed by crossing Target and Array photos from the first two sets, i.e. in these line-ups, Target and Array photos had been taken at different times. For each of the 56 target faces, four target present arrays were constructed (i.e. 224 arrays: target photo at Time 1 or 2; array photos at Time 1 or 2). A further 224 target-absent arrays were constructed by replacing a single photo from each of the target present arrays.

The 448 line-ups were divided into 8 counter-balanced sets, such that each participant saw only 56 trials. So, across the experiment, each target face was seen equally often in target-present and target-absent arrays, and each target was seen equally often with same-time or different-time arrays. Stimulus sets were constructed such that half the participants saw target images taken at time 1 and half saw target images taken at time 2. Time of the 10-face arrays was manipulated within subjects.

The experiment was run individually on an iBook G4 laptop computer using Superlab Pro software. Each participant was presented with 56 trials (14 arrays in each experimental conditions) with a different random order of trials. Each line-up was displayed on the computer screen until a participant made a response. Participants' task was to judge whether the target face was present among the 10-face array, and if so, to choose the match. There was no time pressure and participants were encouraged to perform as accurately as possible.

Results

Consistent with previous research, data from target-present and target-absent trials were calculated separately (e.g., Megreya & Burton, 2007). In target-present trials, three measures were calculated. *Hits* refer to the correct identification of the target face. *Miss* refers to the incorrect response that the target is not present. *Misidentification* (Misid) refers to the incorrect identification of one of the line-up foils. In the target-absent trials, we report *Correct Rejection* (CR), which is the correct negative response that the target is not present among the 10-face line-up. In addition, response times (RTs) of all correct responses (hits and CR) are reported.

---------- TABLE 1 HERE PLEASE ----------

Table 1 shows descriptive statistics for performance across conditions. These data were subjected separately to 2 (time of target photo) x 2 (time of array images) mixed-design analyses of variances (ANOVAs). For *accuracy*, there were no main effects for any of the measures (all Fs < 1), however an interaction was observed for each of the three targe-present measures (*hits*: $F(1,78)=147.2$, $p<0.01$; *misses*:

$F(1,78)=111.7$, $p<0.01$; *misids*: $F(1,78)=9.43$, $p<0.01$). No interaction was observed for *CRs* ($F(1,78)=3.52$, $p> 0.05$). Simple main effect analysis (SMEs) revealed, in each case, that the interactions were due to higher performance on same-date target/array line-ups, than on different date line-ups (e.g., for time 1 targets, SMEs showed more accurate performance on time 1 arrays: *hits* $F(1,78)=70.5$, $p<0.01$, $\eta_p^2=0.47$; *misses*: $F(1,78)=48.2$, $p<0.01$, $\eta_p^2=0.38$; *misids*: $F(1,78)=6.5$, $p<0.05$, $\eta_p^2=0.08$).

For *reaction times*, there was a main effect for time1 vs time 2 targets in hits [$F(1,78)= 51.25$, $p<0.01$; $\eta_p^2 = 0.40$]. Participants who were presented with Time2 targets were faster than those who were shown Time 1 targets. There was also an interaction between time of target and time of array ($F(1,78)= 97.6$, $p<0.01$). Consistent with the accuracy data, this was due to faster responses in same-time line-ups (e.g., for time 1 targets, SMEs showed faster responses on time 1 arrays: $F(1,78)=54.5$, $p<0.01$, $\eta_p^2=0.41$; whereas for time 2 targets, the opposite pattern was observed: $F(1,78)=43.3$, $p<0.01$, $\eta_p^2=0.36$). There were no significant main effects or interactions for reaction times when making correct rejections (all Fs < 1).

Discussion

The aim of this experiment was to examine the ability of viewers to match a target image of an unfamiliar face to an array of possible matches. Consistent with previous studies, this turned out to be difficult for subjects, even given no requirement to remember images between presentations, and good quality images.

When images were taken on the same day, participants identified the correct face (hits) on 79% of occasions, whereas they made miss and misidentification responses on 11.3% and 9.7% of trials (respectively). This low level of performance replicates previous research (e.g., Bruce et al., 1999; Megreya & Burton, 2006a&b; 2008). What we have shown in the present experiment, is that this (already poor) level of accuracy falls very severely when images for matching were not taken at the same time. When images were taken several months apart, the hit rates dropped to 58.6%, whereas miss and misidentification rates jumped to 27.7% and 13.7% (respectively; see Table 1). Consistent with this pattern, participants took much longer to match the images of faces taken on different days compared to those taken on the same day, with roughly a 4-second difference (see Table 1). These results suggest that the poor performance observed in face matching experiments which use same-day photos and optimal viewing conditions, greatly underestimates the real world problem. Experiment 2 aimed to replicate these findings using a paired face-matching task, which is more common in everyday identification than the 1-in-10 face-matching paradigm.

Experiment 2

Method

Participants

Eighty undergraduate students (35 females and 45 males) from Menoufia University volunteered to participate in this experiment. Average age was 19.6 years (SD = 0.8). All participants had normal or corrected to normal vision, and none had participated in Experiment 1.

Stimuli and procedure

The same set of face images were used here as in Experiment 1. However, in this experiment, participants were shown pairs of face images, and simply asked to decide whether the images were the same or different people. Half the pairs show images taken at the same time (1 or 2) and the other half showed images taken at two different times (1 and 2). Figure 3 shows examples. As in Experiment 1, eight sets of stimuli (n= 56) were created for counter-balancing purposes. Therefore, across participants, each target face was seen equally often in same and different trials, and combinations of time 1 and 2 images occurred equally often.

---------- FIGURE 3 HERE PLEASE ----------

Each participant was presented with 56 trials (14 matching pairs in each condition, see Figure 3 for examples), and asked to make same/different responses. Participants were tested individually and stimuli were presented on a Macintosh computer. The task was self-paced, and participants were encouraged to respond as accurately as possible.

---------- TABLE 2 HERE PLEASE ----------

Results

Table 2 shows summary statistics for performance in Experiment 2. These data were subjected to a series of 2 (time of face 1) x 2 (time of face 2) mixed-design ANOVAs. For the *accuracy* data, there were no main effects of hits (time of target: $F(78)= 2.80$, $p> 0.05$; time of array, $F < 1$). However, there was a significant interaction ($F(1,78)= 109.6$, $p< 0.01$). Simple main effects revealed that this was due to higher performance in same-time pairs (e.g. for time 1 targets, SMEs showed higher performance for time 1 arrays ($F(1,78)=54.3$, $p < 0.01$, $\eta_p^2=0.41$; whereas for time 2 targets, the pattern was reversed ($F(1,78)=55.3$, $p < 0.01$, $\eta_p^2=0.42$). For correct rejections, both main effects and interaction were non-significant.

For *reaction times*, the only significant effect was an interaction in the *hits* data ($F(78)= 14.32$, $p< 0.001$). Analysis of SMEs revealed that this was due to faster responses when matching time 1 targets to time 1 arrays ($F(1,78)=15.6$, $p < 0.01$, $\eta_p^2=0.17$). For time 2 targets, the difference between time1 and time 2 arrays was non-significant ($F(1,78)=1.97$, $p > 0.05$).

Discussion

In this experiment, we presented participants with a a face matching task, which consisted of photos taken on the same day or several months apart. In the same-day conditions, the level of all participants' performance was generally low, with roughly 10% and 15% errors for the same and different face-matching pairs (respectively). These low rates converge with previous reports (e.g., Burton et al., 2010; Megreya & Burton, 2006b; 2008). Importantly however, when participants had to match images taken at different times, hit rates dropped and RTs jumped (see Tables 2), consistent with the results of Experiment 1.

General Discussion

The aim of this research was to examine the ability of viewers to match images of unfamiliar faces that were taken on the same day or several months apart using 1-in-10 arrays (Experiment 1) and 1-in-1 pairs (Experiment 2). This study, therefore, tries to simulate face matching in realistic settings. When face images were taken on the same day, hit rates of 79% and 90% were recorded using the 1-in-10 and pairwise face matching tasks, respectively. This low level of performance converges with the results of previous studies, which have only used same-day photos in order to establish the baseline level of unfamiliar face perception (for reviews, see Burton & Jenkins, 2011; Megreya, 2012).

When face images were taken several months apart, hit rates dropped to 58% and 70% on the 1-in-10 and 1-in-1 face matching tasks, respectively. Interestingly, the magnitude of determent between matching faces that were taken on the same or different dates was the same (20%), using both face matching tasks. This determent in accuracy was also accompanied by much longer response times during matching different-date images. Consistently, Davis and Valentine (2009) reported similar detrimental effects even when the face images were only one-week old. These results, therefore, suggest that previous experiments on face matching underestimated the problems of this ability in realistic settings, where same-day photos are rarely used.

The analyses of misses and misidentifications in Experiment 1 tell us more details about the bases of these hit determents. When face arrays included same-date or different-date images of the same faces (respectively), miss rates of 11.3% and 27.7% were recorded along with 9.7% and 13.7% misidentifications. Miss responses, therefore, seem to be much more susceptible to complications when different-date

images were used. On the other hand, using same- or different-date photos had no effects on false positives using both the 1-in-10 and 1-in-1 face matching tasks. In turn, these results suggest that the difficulty of matching different-date images of faces relates to within-person variations, rather than between-person similarities.

A recent study by Jenkins, White, Van Monfort, and Burton, (2011) revealed some interesting findings that could support this suggestion. Using a card sorting task, observers were presented with 40 full-face photographs of 2 unfamiliar people, and asked to group them according to identity. Jenkins et al (2011) did not tell their participants how many identities to expect. Rather, they were free to group the images however they wish. Surprisingly, participants tended to subdivide these two identities into several perceived identities (ranging from 3 to 16), suggesting that images of the same face identity were often considered too dissimilar to go together. In contrast, Jenkins et al (2011) reported a very low level of misidentifications between both identities. Accordingly, Jenkins et al (2011) suggested that the difficulty of this face sorting was to find commonalities among images of the same identity, rather than to find differences between them. Consistently, miss rates in our present study increased more than misidentifications when face arrays involved images taken several months apart.

However, recent evidence seems to suggest that matching several images for the same face identity is also difficult when there is no probability for miss responses. Bindemann and Sandford (2011) asked participants to match three different IDs of the same person (one ID at a time) to a set of 30 photographs, among which the target face was always present. Only 38% of observers matched the correct target to all three IDs, and each ID was mistaken for a variety of different foils. In addition, there were

different recognition rates for the three IDs (67%, 46%, and 58%), suggesting that observers failed to realise that all IDs depicted the same person.

It has been previously suggested that the difference in image format (video versus digital) may be the cause for making face matching difficult, and that using a constant format may make it easier (Bruce et al., 1999; Megreya & Burton, 2006b). Importantly, however, the results of the present study suggest that matching constant images of faces could also be difficult if they were taken on different dates (see also, Bindemann & Sandford, 2011; Jenkins et al., 2011). On the other hand, the date of images could not have any effects on familiar face recognition. Indeed, Jenkins et al (2011) found that the familiarity with faces involved in the card sorting task made it trivial. This further supports the dissociation between familiar and unfamiliar face processing (e.g., see Megreya & Burton, 2006b; 2007).

In most countries, ID documents such as passports and driving licenses are valid through a 10-year period. During this time, facial appearance is almost certainly changed to a much greater extent than the faces that were included in this study (which used a 17-month difference, on average). Therefore, the present results provide strong evidence that age-related changes in facial appearance could increase the mistaken identification of unfamiliar faces, replicating the results of Davis and Valentine (2009). Future research would also be useful to examine how deliberate changes in faces affect matching accuracy as compared to these normal age-related changes. Davis and Valentine (2009) reported mixed results on how wearing a hat or dark eye-glasses influence matching one-year old images to "live" faces. We anticipate that the better understanding of face matching in the context of realistic within-person variability could lead to significant advances in face verification by

both humans and automatic systems. Our experiments provide only a starting point here.

ACKNOWLEDGEMENT

This research was supported by an ESRC Professorial Fellowship to Burton (ES/J022950/1)

References

Bindemann, M., Avetisyan, M., & Blackwell, K. (2010). Finding needles in haystacks: Identity mismatch frequency and facial identity verification. *Journal of Experimental Psychology: Applied, 16*, 378-386.

Bindemann, M., Avetisyan, M., & Rakow, T. (2012). Who can recognize unfamiliar faces? Individual differences and observer consistency in person identification. *Journal of Experimental Psychology: Applied, 18*, 277-291.

Bindemann, M., Brown, C., Koyas, T., & Russ, A. (2012). Individual differences in face identification postdict eyewitness accuracy. *Journal of Applied Research in Memory and Cognition, 1*, 96-103.

Bindemann, M. & Sandford, A. (2011). Me, myself, and I: Different recognition rates for three photo-IDs of the same person. *Perception, 40*, 625-627.

Bindemann, M., Sandford, A., Gillatt, K., Avetisyan, M., & Megreya, A. M. (2012). Recognizing faces seen alone or with others: Why are two heads worse than one? *Perception, 41*, 415-535.

Borchard, E. M. (1932). Convicting the innocent: Sixty-five actual errors of criminal justice. New York: Yale University Press.

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

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

Burton, A. M., Miller, P., Bruce, V., Hancock, P. J. B., & Henderson, Z. (2001). Human and automatic face recognition: A comparison across image format. *Vision Research, 41,* 3185–3195.

Burton, A. M., White, D., & McNeill, A. (2010). The Glasgow Face Matching Test. *Behavior Research Methods, 42*, 286-291.

Burton, A.M. & Jenkins, R. (2011). Unfamiliar face perception. In A.J. Calder, G. Rhodes, M.H. Johnson & J. Haxby (Eds.). *The Oxford Handbook of Face Perception.* (pp 287-306). Oxford: Oxford University Press.

Clutterbuck, R. & Johnston, R. A. (2002). Exploring levels of face familiarity by using an indirect face-matching measure. *Perception, 312*, 985-994.

Clutterbuck, R. & Johnston, R. A. (2004). Demonstrating how unfamiliar faces become familiar using a face matching task. *European Journal of Cognitive Psychology, 17,* 97-116

Clutterbuck, R. & Johnston, R. A. (2004). Matching as an index of face familiarity. *Visual Cognition,11*, 857-869.

Davies, G., & Thasen, S. (2000). Closed-circuit television: How effective an identification aid? *British Journal of Psychology, 91*, 411-426.

Davis, J. P., & Valentine, T. (2009). CCTV on trial: Matching video images with the defendant in the dock. *Applied Cognitive Psychology, 23*, 482–505.

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

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

Hockley, W.E., Hemsworth, D.H., & Consoli, A. (1999). Shades of the mirror effect: Recognition of faces with and without sunglasses. *Memory & Cognition, 27*, 128–138.

Jenkins, R., & Burton, A. M. (2008). Limitations in facial identification: The Evidence. *Justice of the Peace, 172*, 4-6.

Jenkins, R., White, D., Van Monfort, X., Burton, A. M. (2011). Variability in photos of the same person. *Cognition, 121*, 313-323.

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

Konar, Y., Bennett, P. J., & Sekuler, A. B. (2010). Holistic processing is not correlated with face-identification accuracy. *Psychological Science, 21*, 38-43.

Megreya, A. M. & Burton, A. M. (2006a). Recognising faces seen alone or with others: When two heads are worse than one. *Applied Cognitive Psychology, 20,* 957-972.

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

Megreya, A. M. & Havard, C. (2011). Left face matching bias: Right hemisphere dominance or scanning habits? *Laterality: Asymmetries of Body, Brain and Cognition, 16,* 75-92.

Megreya, A. M. (2012). Accuracy of face recognition. In A. Quaglia & C. M. Epifano (Eds). *Face Recognition: Methods, Applications and Technology* (pp. 1-28). New York: Nova Publishers.

Megreya, A. M., & Bindemann, M. (2009). Revisiting the processing of internal and external features of unfamiliar faces: The headscarf effect. *Perception, 38*, 1831-1848.

Megreya, A. M., & Bindemann, M. (in press). Individual differences in personality and face identification. *Journal of Cognitive Psychology*.

Megreya, A. M., & Burton, A. M. (2007). Hits and false positives in face matching: A familiarity-based dissociation. *Perception & Psychophysics, 69*, 1175-1184.

Megreya, A. M., & Burton, A. M. (2008). Matching faces to photographs: Poor performance in eyewitness memory (without the memory). *Journal of Experimental Psychology: Applied, 14*, 364-372.

Megreya, A. M., Bindemann, M. & Havard, C. (2011). Sex differences in unfamiliar face identification: Evidence from matching tasks. *Acta Psychologica, 137*, 83-89.

Megreya, A. M., Bindemann, M., Havard, C. & Burton, A. M. (2012). Identity-lineup location influences target selection: Evidence from eye movements. *Journal of Police & Criminal Psychology, 27,* 167-178.

Megreya, A. M., White, D., & Burton, A. M. (2011). The other race effect does not rely on memory: evidence from a matching task. *Quarterly Journal of Experimental Psychology, 64,* 1473-1483.

Memon, A., Vrij, A., & Bull, R. (2003). *Psychology and law: Truthfulness, accuracy and credibility* (2nd ed.). Chichester, England: Wiley.

Münsterberg, H. (1908). On the witness stand: Essays on psychology and crime. New York: Doubleday, Page & Company.

Narby, D. J., Cutler, B. L., & Penrod, S. D. (1996). The effects of witness, target, and situational factors on eyewitness identifications. In S. L. Sporer, R. S. Malpass & G. Koehnken (Eds.), *Psychological issues in eyewitness identification* (pp. 23–52). Mahwah, NJ: Erlbaum.

Özbek, M. & Bindemann, M. (2011).Exploring the time course of face matching: Temporal constraints impair unfamiliar face identification under temporally unconstrained viewing. *Vision Research, 51*, 2145–2155.

Strathie, A., McNeill, A. & White, D. (2012). In the dock: Chimeric image composites reduce identification accuracy. *Applied Cognitive Psychology, 26*, 140–148.

Terry, R.L. (1994). Effects of facial transformations on accuracy of recognition. *Journal of Social Psychology, 134*, 483–492.

Wells, G. L., Wright, E. F., & Bradfield, A. L. (1999). Witnesses to crime: Social and cognitive factors governing the validity of people's reports. In R. Roesch, S. D. Hart, & J. R. P. Ogloff (Eds.), *Psychology and law: The state of the discipline* (pp. 53–87). New York: Kluwer Academic/Plenum.

Figure 1. Examples of the 1-in-10 face matching arrays that were used in Bruce et al's (1999) study. The person shown at the top may or may not be one of the ten below. Subjects' task is to decide if he is present, and if so, which he is.

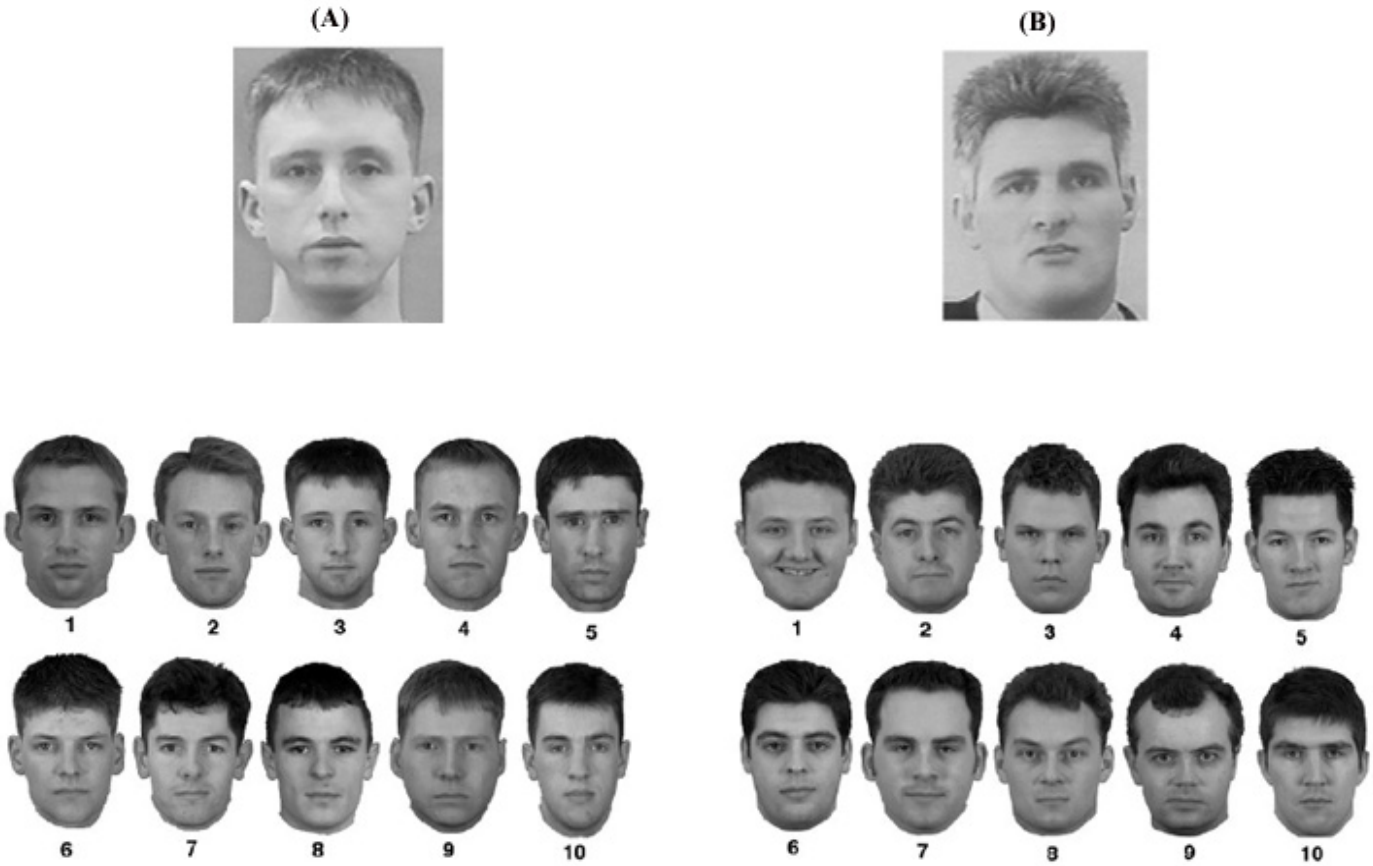

Figure 2. Examples of the 1-in-10 target-present arrays in Experiment 1. A+B= Time1 target/ Time1 array; C+D= Time2 target/ Time2 array; A+D= Time1 target/ Time2 array; C+B= Time2 target/ Time1 array.

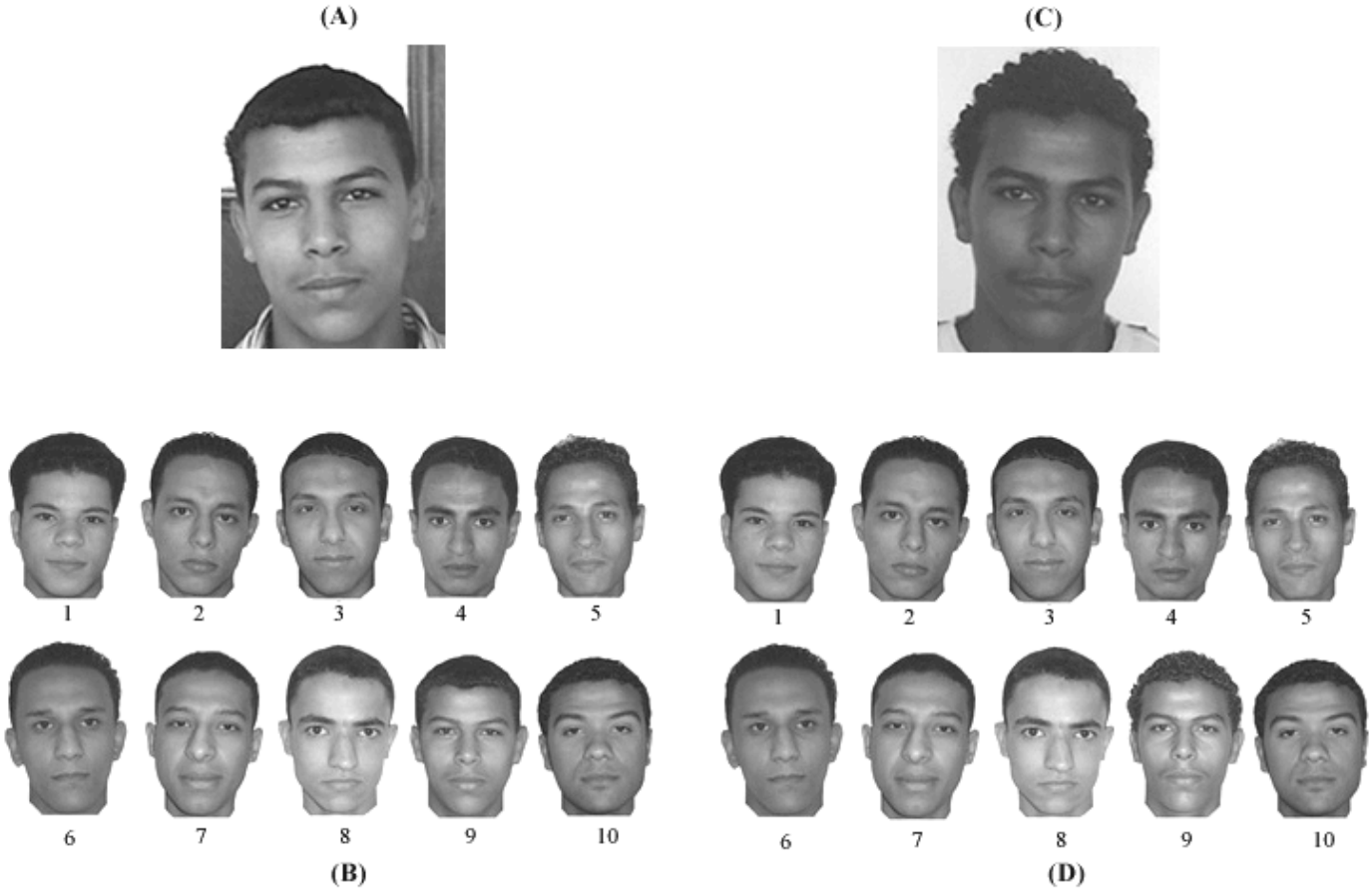

Figure 3. Examples of face matching pairs used in Experiment 2 (Camera A/Camera B).

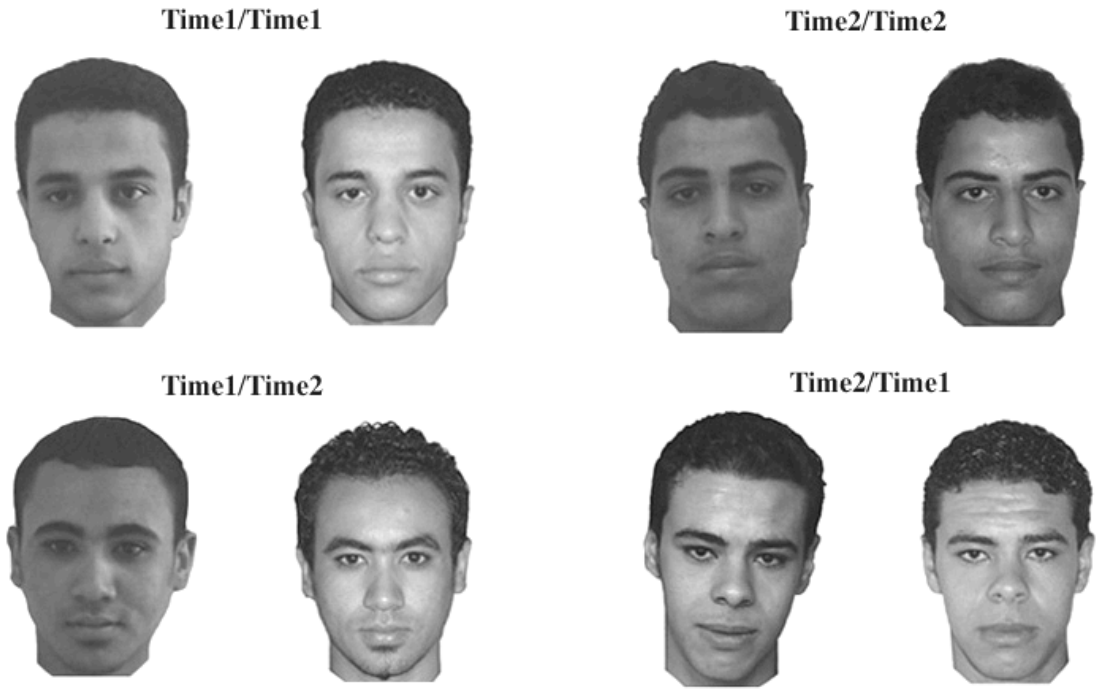

Table 1. Mean performance across Target/Array Conditions (*sd* in parenthesis) in Experiment 1.

Measures	Time1/Time1	Time1/Time2	Time2/Time2	Time2/Time1
Accuracy (%)				
Hits	77.5 (*15.4*)	57.5 (*18/4*)	80.5 (*15.2)*	59.6 (*20.2*)
Miss	13.7 (*13.1*)	28.9 (*17.3*)	8.9 (*11.1*)	26.5 (*19.4*)
Misid	8.8 (*11.9*)	13.6 (*14.2*)	10.6 (*12*)	13.9 (*12.1*)
CR	66.8 (*25.4*)	65.7 (*27*)	68.2 (*23.1*)	62.3 (*22.8*)
RTs (sec)				
Hit	10.7 (*5.6*)	14.9 (*7.1*)	8.0 (*2.6*)	11.8 (*4.1*)
CR	21.2 (*11.2*)	21.3 (*13.2*)	19.9 (*10*)	20.7 (*11.7*)

Table 2. Mean performance across Conditions (*sd* in parenthesis) in Experiment 2.

Measures	Time1/Time1	Time1/Time2	Time2/Time2	Time2/Time1
Accuracy (%)				
Hits	87.5 (*13.8*)	67.5 (*19.5*)	92.3 (*9.8*)	72.1 (*16.4*)
CR	84.3 (*14.2*)	84.1 (*13.1*)	87.3 (*11.9*)	88.6 (*10.8*)
RTs (sec)				
Hits	2.7 (*1.8*)	3.4 (*2.1*)	2.9 (*1.3*)	3.1 (*1.7*)
CR	2.9 (*1.2*)	2.9 (*1.4*)	3.0 (*1.5*)	3.1 (*1.8*)