CCTV on Trial: Matching Video Images with the Defendant in the Dock.

Josh P Davis and Tim Valentine

Goldsmiths, University of London, London, UK

Running head: CCTV ON TRIAL

Author note

Josh P Davis and Tim Valentine, Department of Psychology, Goldsmiths, University of London, London, SE14 6 NW United Kingdom

This work was funded by a Postgraduate Studentship Grant PTA-030-2002-00990 from the Economic and Social Research Council. We thank the Science Museum, London for providing space and facilities in order to recruit volunteers for Experiment 2. We also gratefully thank Steve Nicola for his valuable assistance in collecting data and recruiting culprits for Experiment 3. Parts of this research have been presented at the Experimental Psychology Society London Conference, University College, London in January 2006 and at the British Psychological Society Annual Conference, University of York in March 2007. We would like to thank Mike Burton and Siegfried Sporer for their comments concerning an earlier version of this article.

Correspondence concerning this article should be addressed to:

Professor Tim Valentine, Department of Psychology, Goldsmiths, University of London, London, SE14 6NW, United Kingdom.

Tel: +44 20 7919 7871

E-mail: t.valentine@gold.ac.uk

Abstract

The experiments reported in this paper investigated simultaneous identity matching of unfamiliar people physically present in person with moving video images typical of that captured by CCTV. This simulates the decision faced by a jury in court when the identity of somebody caught on CCTV is disputed. Namely, "is the defendant in the dock the person depicted in video?" In Experiment 1, the videos depicted medium-range views of a number of actor 'culprits'. Experiment 2 used similar quality images taken a year previously, some of which showed the culprits in disguise. Experiment 3 utilised high quality close-up video images. It was consistently found that in both culprit-present and culprit-absent videos and in optimal conditions, matching the identity of a person in video can be highly susceptible to error.

Keywords: IDENTIFICATION; UNFAMILIAR FACES; RECOGNITION; CCTV

CCTV on Trial: Matching Video Images with the Defendant in the Dock.

Closed Circuit Television (CCTV) surveillance has become increasingly prevalent in most industrialised nations in recent years. There are at least 4,285,000 cameras in the UK (Norris, McCahill & Wood, 2004), 26,000,000 within the USA (*Washington Post*, 8 October 2005) and continued widespread installation is expected worldwide (Norris et al., 2004). CCTV footage of a crime scene provides a permanent record of events and of those involved. The expansion in camera surveillance is having proportional effects on the evidential use of images and it can have a powerful impact in court.

A considerable body of research on human identification in a forensic setting has investigated person recognition by eyewitnesses (for reviews see Cutler & Penrod, 1995; Wells, Small, Penrod, Malpass, Fulero, & Brinacombe, 1998**;** Wells, Memon & Penrod, 2006). Analyses of cases of wrongful conviction show that eyewitness error is a principal cause of miscarriages of justice (e.g., Rattner, 1988; Scheck, Neufeld & Dwyer, 2000). Empirical studies have also confirmed the fallibility of face recognition and that confidence in mistaken identifications can be extremely high (Sporer, Penrod, Read & Cutler, 1995; Wells, Olson & Charman, 2002). Consequently, the psychological science of eyewitness identification has had an impact on legislation and legal procedures (e.g., UK: The codes of practice required under the Police and Criminal Evidence Act (PACE), 1984; USA: Technical Working Group for Eyewitness Evidence, 1999). Similar questions about the reliability of person identification from CCTV evidence have recently been tested in experimental studies (e.g., Davies & Thasen, 2000; Henderson, Newman, & Burton, 2001) and in the English legal system (Attorney General's Reference, 2003). Interest in this issue, particularly in the UK, arises from its relatively long history of intensive public surveillance and there is a requirement for psychological science to inform policy on the forensic use of this type of evidence.

Police procedures have been transformed by the use of CCTV images. For example, identity decisions can be made whilst simultaneously viewing photographs or suspects in custody. Davies and Thasen (2000) note that perceptual matching tasks of this type have close similarities to some in which memory is tested. They suggest that when developing theoretical models of human memory performance, effects that derive from perception should be isolated in order to identify purely mnemonic processes. From an applied perspective there is a need to examine whether factors known to influence memory for people also impact on legal procedures that rely on perception.

Surveillance images can be presented by the prosecution to encourage jurors, who are unfamiliar with the defendant to conclude that the defendant is depicted. In the UK, senior judges have ruled that if footage is 'sufficiently clear', juror's may be invited to decide whether the defendant is the person depicted in CCTV footage and cases may proceed in the absence of further identification evidence (Attorney General's Reference, 2003). Unlike eyewitness testimony this task has no demands on memory, and yet, simultaneous unfamiliar face matching is surprisingly error prone (Bruce, Henderson, Newman, & Burton, 2001; Davies & Thasen, 2000), even if image quality is high (e.g., Bruce, Henderson, Greenwood, Hancock, Burton & Miller, 1999; Henderson et al., 2001). For example, Bruce et al. (1999) reported participant error rates of 30% when simultaneously matching unfamiliar high-quality close-up full-face video stills against arrays of ten full-face photographs. Performance was worse when expression or pose was changed. Face matching with minimal task demands is also error prone (Henderson et al., 2001; Megreya & Burton, 2006; 2007). Henderson et al. (2001; Experiment 4) found that 24% of decisions were incorrect in a forced-choice design requiring an culprit in a television broadcast-quality still image to be identified from two professional portrait photographs. In a follow-up (Experiment 5), approximately 45% of participants mistakenly believed that two images of the same actor were of two different people, while 27.5% incorrectly matched two different actors.

Theoretical explanations have been advanced. Megreya and Burton (2006; 2007) found substantial stable individual differences in accuracy which also correlated with standardised visual perception tests, suggesting it might be possible to predict performance. However, there was only a weak relationship between target present and absent matching performance and no correlation with the identification and processing of familiar or recently learned faces. In contrast to the high error rates associated with unfamiliar face matching, familiar faces tend to be reliably recognised, even in highly degraded images (Bruce et al., 2001; Burton, Wilson, Cowan & Bruce, 1999) and Megreya and Burton propose that familiar and unfamiliar face identification involves qualitatively different operations. Inverting familiar faces disrupts recognition (e.g., Valentine, 1988) and has been explained as interfering with the configural processing of faces so that a less effective feature-by-feature strategy is required. Megreya and Burton found a strong relationship between performance at matching both upright and inverted unfamiliar faces and tasks involving inverted familiar celebrity faces, suggesting that unfamiliar face matching is also constrained by featural processing operations.

These research findings are of consequence, as the majority of incidents caught on CCTV are likely to involve people not known to the police, and definitely not to a jury.

Judges may warn juries that resemblance does not constitute identity. However, other corroborating evidence may be circumstantial and individual jurors could base verdicts on whether they perceive a likeness or not. Nevertheless, the studies reported above used photographs as the test medium, whereas in court, a defendant would be present in person. Only one published study appears to have examined simultaneous face matching using physically present actors. Kemp, Towell and Pike (1997) found that supermarket cashiers were unable to detect the majority (64%) of photo-identity credit cards of another person presented by 'fraudulent' shoppers. However, the photos were 2-cm^2, possibly too small to distinguish detailed features. In addition, a facial photograph cannot provide a range of views, and potential identifying information such as height, weight, gait, idiosyncratic postures or expressions that may be acquired from surveillance videos.

Therefore, in the three experiments reported in this paper, participants had to decide whether a target actor, the 'defendant' present in person was depicted in video footage, simulating the scenario by which jurors may be invited to observe the resemblance between the two. In half of the trials, the defendant was present in the footage (culprit-present). In the remainder, a distracter matched for appearance replaced the defendant (culprit-absent), replicating a situation when an innocent suspect of similar appearance has been wrongly charged. The video sequences used in the first two experiments were full-body medium-range views. The third experiment utilised high-quality, close-up facial images.

Criminal proceedings can take many months from instigation to trial. In this period it is possible that the appearance of the accused will have altered, and jurors, anticipating change, may be more willing to infer identity. Therefore, variations in time between video capture and identification session were also examined, as well as providing knowledge of the age of the footage. Furthermore, interviewed offenders have stated that to avoid detection from CCTV, they would deliberately conceal facial features (Gill & Loveday, 2003). An additional aim was therefore to examine the extent to which the wearing of disguise would reduce identification accuracy.

In court, juries may be provided with video stills for close inspection. However, if moving sequences exist, they will be presented. There is evidence that movement, particularly if highly idiosyncratic, can assist recognition of familiar people in degraded images (Cutler, Berman, Penrod & Fisher, 1994; Cutler & Penrod, 1995; Knight & Johnston, 1997; Lander, Bruce & Hill, 2001; Lander & Chuang, 2005). However, with some exceptions (e.g., Schiff, Banka & de Bordes Galai, 1986) movement is not advantageous for unfamiliar face identification (e.g., Bruce et al., 1999; Christie & Bruce, 1998). Nevertheless, to ensure ecological validity, moving images were employed in the experiments reported in this paper, with no viewing time limit.

Finally, within the eyewitness literature, there has been extensive research into the relationship between confidence and identification accuracy (e.g., Brewer, 2006; Sporer et al., 1995). The overall relationship tends to be fairly weak. However, in a meta-analysis, Sporer et al found a stronger relationship between accuracy and confidence in those who make a positive identification decision – ‘choosers’, than in ‘non-choosers’. In addition, the confidence levels of correct choosers are consistently higher than incorrect choosers. There is speculation over what influences confidence, although the main contributor appears to be memory strength (Brewer, 2006). There can be no influence of memory in a matching study. However, analogous to results found in eyewitness studies.

Bruce et al. (1999) found that correct matching decisions were associated with higher confidence, especially when the target was present in arrays. In addition, using a similar single-item verification task to that employed in the current paper, Henderson et al. found that mean confidence in decisions was higher when the two images were of the same person. However, they do not report any statistical analyses and did not separate confidence for correct and incorrect responses. Therefore, to examine if similar differences between choosers and non-choosers would be found in the present research, a combined identity decision and confidence scale was employed and as recommended by Sporer et al. (1995) the accuracy-confidence relationship was examined in those making correct and incorrect decisions and separately for choosers and non-choosers.

Experiment 1

Experiment 1 was designed to examine whether the high error rates in unfamiliar person matching from video images to photographs, would be found when matching a person in medium-range video to a person physically present in the room. In some cases jurors may make identification judgements for more than one defendant and to replicate this, some participants took part in two sessions.

Bruce et al. (1999) reported wide performance variations to individual stimuli in their studies, with 80% of participants failing to match one videoed target with their photograph, whereas accuracy was high in other trials. Therefore, in Experiment 1 multiple culprits were recruited to explore the extent to which error rates may differ between comparisons to individual defendants. In addition, in a forensic context, it is of interest to identify the individual comparisons in which performance is at its most error prone, and therefore the results to each actor when playing the part of the defendant are reported.

Method

Participants. Participants, unfamiliar with the actors were 44 male and 154 female students, aged between 18 and 58 (Mean = 25.3), at Goldsmiths, University of London.

Design. This experiment used a single-item identity-verification design. Participants attended one; or two identification sessions. In each, they viewed a video clip and were required to decide whether a physically present defendant was depicted in the footage. Eight different defendants took part in identification sessions, each attended one culprit-present and one culprit-absent session. For culprit-present conditions a video of the defendant taken approximately three weeks previously was shown. For culprit-absent trials the video was of a distracter of similar appearance. The primary dependent variable was identification accuracy based on converted scores from a combined Identity-Decision and Confidence scale.

Materials and actors. Eight male white Caucasian members of a student rugby club (aged 19 to 21 years) were recruited to act as defendants[1]. All were of slim or medium/muscular build (72 – 95 kg, 1.70 - 1.92 m), clean-shaven with brown or black hair, neither receding, nor below collar length. None had distinguishing facial marks or wore jewellery. From a club membership of about 60, approximately one-third met these criteria, which were chosen in order to select a group of young men who might plausibly be the subject of disputed identification. For example, a number of people may admit presence at or near a crime scene, but dispute identification as an individual recorded by CCTV committing a crime (e.g., an assault outside a nightclub).

To create culprit-absent trials, pilot participants (n = 20) unfamiliar with the culprits paired those 'most likely to be mistaken for one another' using full body colour photographs taken at the time of filming. The videoed culprit for each defendant was the actor paired with that defendant by the most participants.

All of the culprit actors were filmed with a JVC GR-SX22 Compact VHS high-quality domestic camcorder, which was set on a tripod approximately 3.5 metres above ground level and operated using manual zoom and movement controls so that the culprit constantly took up approximately one-half to two-thirds of the screen. This was designed to simulate footage obtained from typical operator-controlled or automatic externally-based CCTV systems Each culprit were instructed to performed choreographed actions for approximately 35-sec over a distance of 50m. This meant that the front of the face, the left and the right facial profile were each depicted for at least 5-sec, while some background environmental details provided cues to height and build. The sequence involved walking at an angle towards the camera for a right three-quarters facial view, continuing behind a 1 metre high wall perpendicular to the camera for a right profile view, standing still and facing the camera, turning, standing and pretending to use a mobile phone held in the right hand for a left profile view; and finally walking towards and underneath the camera. Four stills from the footage taken of one of the culprits (Actor F), associated with the highest proportion of errors in Experiment 1 are shown in Figure 1. Full-face photographs of that culprit and of the defendant (Actor A) present in culprit-absent trials in which this video was used, taken at the time of the identification sessions three weeks later are also provided, as are photographs of the culprit (Actor *I*) and defendant (Actor *G*) associated with the fewest errors.

FIGURE 1 ABOUT HERE

Ex-members of the same rugby club (n = 8), unaware of the identity of those who had volunteered to be filmed were recruited for a pilot study to ensure that the images were of sufficient quality for recognition by acquaintances. In all cases, naming accuracy was 100% within a few seconds of viewing each video.

Using a similar measure employed in face matching research by Bruce et al. (2001), participant responses were collected on an eight-point Identification-Decision and Confidence (IDC) scale ranging from 1 (definitely not the same person) to 8 (definitely the same person). These were categorized as ‘same’ (5 - 8) or ‘different’ (1 - 4) person decisions for accuracy assessment, with those on the extremes (1 & 8) taken to indicate a high confidence in decisions, and coded 4 on a confidence scale of 1 – 4. Those near the centre (4 & 5) were treated as unsure and recoded as 1. Likewise 2 and 7 were recoded 3, and 3 and 6 were recoded 2.

Procedure. Participants were tested individually or in computing lectures with participants seated approximately 1m from individual 18” monitor screens. Participants attended a single or two identification sessions. Videos and culprits in a second session were always different to those in the first. Those attending a second session were asked to mark this on their response sheets. However, as responses were anonymous, collating the two sets of data from each participant was not possible. Participants were informed as to their task and its use in a courtroom scenario and to attempt to be as accurate as possible in deciding whether they believed that the defendant was depicted in the video footage.

Once the video was started, the defendant, blind to whether the footage was of him or not and wearing different clothing from in the video, would enter the room, and stand under good room lighting with arms folded while keeping a neutral expression. Participants viewed the video sequence and recorded their responses on the IDC scale. Those in lectures were invited to move if necessary to ensure a clear view and were required to remain silent and to keep responses private. The video was replayed a minimum of three times and continued until all responses had been collected. Full performance feedback was provided at the end of each session.

Results

Responses were categorised for accuracy and for confidence assessment. Of the 198 participants, 136 attended two sessions. Table 1 shows the converted error rate as a function of whether it was a participant's first/only decision or whether it was their second decision, expressed as a percentage in each of the 16 identification sessions with each defendant denoted by a letter. The proportion of participants responding with scores of 1 or 2 in culprit-present and 7 or 8 in culprit-absent trials are also listed as these are indicative of highly confident incorrect responses. In culprit-present conditions, 22% of participants made false negative identification decisions and 9% of all responses were highly confident incorrect decisions. Performance varied so that with one video no errors were made (Actor *A*). However, with another (Actor *F*), 37% of participants wrongly believed that the defendant was not depicted in the video footage. In culprit-absent conditions, 17% of participants made false positive identification decisions, and 3% of all decisions were highly confident errors. Again, performance varied across defendants with no errors made to one video (Actor _G_), whereas 44% of participants wrongly believed another defendant (Actor _A_) was depicted in the video footage showing a distracter.

To analyse confidence levels, the individual defendant session data were pooled to increase statistical power. Only the data from participant's first or only session was included in this analysis to avoid violating assumptions of independence. Table 2 shows the first or only session mean confidence levels as a function of accuracy and participant decision (choosers vs. non-choosers). A 2 (accuracy) x 2 (participant decision) ANOVA[2] revealed a significant main effect of accuracy, $F(1, 194) = 24.35$, $p < .001$, $r = .33$, correct decisions were associated with higher levels of confidence than incorrect decisions ($M = 2.82$, $SD = 0.94$ vs. $M = 2.00$, $SD = 0.97$). The main effect of decision, $F < 1$, $r = .06$, and the interaction were not significant, $F < 1$, $r = .03$. Simple main effects analyses found higher confidence for correct decisions than incorrect decisions in both choosers, $F(1, 194) = 13.26$, $p < .001$, $r = .25$ and non-choosers, $F(1, 194) = 11.10$, $p < .001$, $r = .23$.

In addition, two point-biserial correlations revealed a highly significant accuracy-confidence relationship in both choosers, $r(94) = .39$, $p < .001$ and in non-choosers, $r(104) = .29$, $p < .005$, although the correlation was higher for those making a positive identification decision[3].

Discussion

Experiment 1 demonstrated that when an unfamiliar person is physically present, identity verification from video footage can be highly susceptible to error. This finding is consistent with previous research using photographs (e.g., Bruce _et al_., 1999; Henderson _et al_., 2001). Across the eight culprit-absent sessions, 17% of participants, some highly confident, wrongly identified a distracter as being in the video. No errors were made in one trial, whereas in contrast, in three sessions, over 30% of participants made identification errors. All of the actors were recruited from the same sports club and yet even with a limited pool of 60 members it was perhaps surprisingly simple to find some who were mistaken for one another. Research has found that unfamiliar face identification judgments are predominantly based on external features (e.g., hairstyle, hairline, face shape; Bruce et al., 1999; Ellis, Shepherd & Davies, 1979; Young, Hay, McWeeny, Flude & Ellis, 1985), with hairstyle the most salient cue (O'Donnell & Bruce, 2001). The high levels of misidentifications in this experiment may be due to the homogeneous appearance of these actors, enhanced by a tendency for similar hairstyles.

In culprit-present trials, 22% of matching judgements were incorrect, 40% of which were highly confident decisions. Again performance varied across sessions and was at ceiling in one comparison only. A quarter or more of participants in trials involving four other defendants incorrectly responded that these actors were not in the videos. Filming had taken place three weeks previously and appearance may have altered. However, most judicial proceedings will take longer to process and so this experiment possessed ecological validity.

Consistent with the findings of Bruce et al. (1999) and most eyewitness research (e.g., Sporer et al., 1995), correct decisions were associated with higher confidence than incorrect decisions and the relationship between accuracy and confidence was higher in choosers than in non-choosers. If replicated in a court room this reassuring finding suggests that, in the absence of other information, jurors might be more cautious in inferring guilt than innocence on the basis of video footage alone.

Participants who were highly familiar with the culprits were 100% accurate in naming them from their video image in a pilot task. Performance will have been aided by knowledge that all were members of their rugby club. However, the same stimuli were used in separate experiments conducted at the Science Museum in London in which museum visitors participated (Davis, 2007). In these experiments, eight volunteers who, by chance knew the actors, spontaneously named the 'culprits' when viewing the images. In one case, 'culprits' in four out of six videos were identified. It is not possible to know if others failed to recognise acquaintances. However, these findings support previous research finding reliable familiar face recognition in CCTV (Bruce et al., 2001; Burton et al, 1999).

One limitation of Experiment 1 is that there were low participant numbers in some trials. Furthermore, the experiment took place less than a month after filming. In a forensic setting, even if a suspect is quickly apprehended, criminal investigations can take months, or even years and there may be substantial natural or deliberate changes to their appearance. Therefore, in Experiment 2 considerably more participants were recruited from a more diverse demographic background, with identification sessions taking place a year after filming. A further aim of Experiment 2 was to investigate the effects of criminals' attempts to conceal their appearance from CCTV surveillance. Therefore, in some of the videos the culprits were depicted in a hat or dark glasses

Experiment 2

To avoid detection, criminals often conceal facial features. They are probably correct to believe that this could reduce the likelihood of prosecution. There is extensive evidence that the removal or addition of disguises impairs facial recognition (e.g., wigs, hats, glasses: Diamond & Carey, 1977; dark glasses: Hockley, Hemsworth & Consoli, 1999; Metzger, 2001; clear glasses and beards: Terry, 1994; for a meta-analysis see Shapiro & Penrod, 1986). In addition, Henderson et al. (2001; Experiment 3) found that unfamiliar simultaneous face matching accuracy was worse when actors were disguised with hats. Hats predominantly conceal external facial features and these results are unsurprising, as in contrast to familiar face recognition that is primarily based on the processing of the internal features of a face, unfamiliar face identification is based on external feature processing (Bruce et al., 1999; Ellis et al., 1979; Young et al., 1985). Indeed, Bruce et al. (1999; Experiment 4) found that while unfamiliar face matching performance was best when whole faces were presented, judgements were more accurate when internal facial features rather than external features were concealed. However, the images were taken on the same day, and as hairstyle can be easily manipulated, the same strategy with images taken some time apart might be flawed.

Experiment 2 therefore examined the effect on matching accuracy of dark glasses which partially obscure internal features by comparing performance to matching actors in no disguise or depicted in a hat which partially obscured external features. A further aim was to examine whether judgements would be influenced by knowing that the footage was not contemporary and therefore half of the participants were correctly warned that the videos were one-year old.

In an eyewitness identification study, Charman and Wells (2007) found that providing instructions that a suspect's appearance may have changed, increased distracter selections from both culprit-present and culprit-absent lineups. The authors provide two explanations. One is that this instruction induced a general reduction in decision criterion.

The second is that it somehow increased the ecphoric similarity of the memory trace of the target to the selected distracter. However, neither account can explain why this did not also increase correct culprit selections over distracters. Experiment 2 was not designed to decide between these two accounts. However, the warning that the videos were a year old can be seen as an implicit version of the appearance change instruction. If a more liberal criterion was applied when making a 'same person' or choosing decision, more errors would be predicted when the videoed culprit was not the defendant, and fewer errors predicted when the culprit was the defendant. These effects would be expected for disguised and undisguised videoed culprits. By combining the hit or correct identification rate in culprit present trials with the false alarm or incorrect identification rate in culprit-absent trials it was also possible to assess response bias in each condition to measure any such criterion change (B'', Donaldson, 1992).

Wearing a hat was predicted to have the most negative influence on accuracy and confidence of identification, as external features are of primary importance in unfamiliar face processing. Wearing dark glasses could be expected to impair performance compared to an actor wearing no disguise because the eyes are occluded. However, the effect of glasses was not predicted to be as detrimental as wearing a hat. Nevertheless, an alternative prediction in relation to wearing glasses is possible. Wearing glasses might compel participants to use a more efficient strategy of processing external features (Bruce et al., 1999). If hairstyle had not substantially changed in the one-year intervening period, performance would potentially be better when the actor was wearing glasses than when in no disguise. However, if hairstyle had changed, identification of an actor wearing glasses would be worse than wearing no disguise.

Method

Participants. Participant, unfamiliar with the actors, were 303 male and 288 female visitors aged between 18 and 70 (Mean age = 34.2) to the Live Science exhibit located within the Who am I? Gallery at the Wellcome Wing of the Science Museum, London.

Culprits and materials. Two defendants, paired in Experiment 1 were recruited for this study (Actors *C* and *E*). In culprit-absent trials the video for Actor *C* was played while Actor *E* was present in the room as the defendant and vice versa. The no disguise videos were those used in Experiment 1. Similar sequences had been recorded at the same time with the culprits wearing dark glasses and a hat. Photographs of Actor *E* in each disguise condition are shown in Figure 2 and photographs of both defendants taken at the time of filming, and at the identification sessions a year later are shown in Figure 3.

FIGURES 2 AND 3 ABOUT HERE

Design. This experiment utilised a 2 (culprit presence) x 2 (video age information) x 3 (disguise) x 2 (defendant) independent measures design, using the same identity-verification task and measures as Experiment 1. The first factor, Culprit Presence was whether the defendant was present or absent in the video. The second factor, Video Age Information was whether participants were warned in advance that the video had been filmed one year previously or whether no warning was given. The third factor was the Disguise of the videoed culprit. There were three levels: no disguise; dark glasses or hat. The two Defendants were treated as levels of a fourth factor.

Procedure. The procedure was similar to Experiment 1 except that the footage sometimes showed the culprit in disguise and approximately half the participants were verbally warned in advance that the films were a year old. The remainder were provided with no additional information. In all conditions the defendant wearing no disguise stood still with arms folded next to the 18" television/video screen. Participants were seated silently in a row in groups of up to eight, approximately 2m in front of the screen.

Results

As described in Experiment 1, scale scores were converted for accuracy and confidence and Table 3 shows the percentage error rate as a function of disguise, warning and culprit presence. The proportion of participants responding with highly confident incorrect responses is also listed. In addition, as the two actors performed as matched culprit and defendant in each condition, the non-parametric measure of response bias - B'' (Donaldson, 1992) is reported. Values can range from + 1.0 to – 1.0. Negative values indicate a liberal bias, positive values a conservative bias.

Overall, 44% of participants in culprit-present trials incorrectly believed that the defendant was not present in the video. A further 33 % of participants who took part in culprit-absent trials wrongly believed that the defendant and the culprit in video were the same person. Of those making errors, 51% were highly confident in that they responded on the extreme intervals of the IDC scale.

TABLE 3 ABOUT HERE

A 3 (disguise) x 2 (culprit presence) x 2 (warning) x 2 (defendant) independent measures ANOVA conducted on these data revealed a significant main effect of disguise, $F(2, 567) = 5.24$, $p = .006$, $\eta^2 = .018$. Bonferroni post-hoc t-tests found that significantly fewer errors were made in the glasses condition (30.2%) than in the no disguise condition (44.0%, $p < .05$, $r = .14$) and a trend towards being fewer than in the hat condition although this was not significant (41.2%, $p < .1$, $r = .11$). The no disguise and the hat conditions did not significantly differ ($p > .2$, $r = .03$). The main effect of culprit presence was also significant, $F(1, 567) = 7.69$, $p = .006$, $r = .12$. More errors were made when targets were present than when they were absent. The main effects of warning and of defendant were not significant, $F < 1$.

However, the interaction between warning and defendant was significant, $F(1, 567) = 7.03$, $p = .008$, $r = .11$. Simple main effects analyses of the interaction revealed that when Actor E was the defendant, error rates were significantly lower when a warning was given (33%) than when no warning was given (46%; $F(1, 567) = 5.38$, $p = .021$, $r = .10$). The trend was reversed with Actor C (warning: 42% vs. no warning: 34%), although the difference was not significant, $F(1,567) = 2.19$, $p > .1$, $r = .06$. All the remaining two-way and higher order interactions were not significant ($p > .1$).

Although the interaction between disguise and warning was not significant, examination of the response bias measure, B'' in Table 3 demonstrates that a conservative criterion was applied in all conditions except when the defendants were wearing glasses and a warning was given. This measure is derived by combining correct culprit-present (hit) rates with incorrect culprit-absent (false alarm) rates and this statistic indicates that except in this specific condition, participants tended to be cautious in making positive identification decisions.

To examine confidence levels, data from the separate actors were pooled to increase statistical power. A 2 (accuracy) x 2 (decision type) x 2 (warning) x 3 (disguise) ANOVA on the converted confidence data found a significant main effect of accuracy, $F(1, 567) = 8.81$, $p = .003$, $r = .12$. Correct responses were associated with higher confidence than incorrect responses (*M* = 2.90, *SD* = 0.97 vs. *M* = 2.68, *SD* = 1.00). The main effect of disguise was also significant, $F(2, 567) = 3.96$, $p = .020$, $\eta^2 = .014$. Bonferroni-corrected t-tests found that confidence was higher in the no disguise condition than in the dark glasses condition (*M* = 2.96, *SD* = 0.94 vs. *M* = 2.71, *SD* = 0.99; $p < .05$, $r = .12$). Confidence levels in the hat condition (*M* = 2.78, *SD* = 1.02) did not significantly differ from the other two disguises ($p > .2$). The main effect of decision was significant, $F(1, 567) = 18.44$, $p < .001$, $r = .18$. Confidence for non-choosers was higher than for choosers (*M* = 2.97, *SD* = 0.97 vs. *M* = 2.63, *SD* = 0.97). The main effect of warning, $F < 1$, and all the interactions were non-significant ($p > .1$). Simple main effects analyses examined the non-significant interaction between decision and accuracy. There was found to be no significant difference in confidence levels for correct or incorrect choosers, $F(1, 567) = 2.43$, $p > .1$, $r = .06$. However, correct decisions were associated with higher confidence than incorrect decisions in non-choosers, $F(1, 567) = 6.73$, $p = .01$, $r = .11$.

In addition, two point-biserial correlations revealed that there was no significant relationship between confidence and accuracy in choosers, $r(328) = .09$, $p > .1$. However, the accuracy-confidence relationship was significant for non-choosers, $r(263) = .14$, $p < .05$ indicative of a higher correlation for those making a negative identification decision.

Discussion

Experiment 2 confirmed the difficulties inherent in simultaneous identity matching from medium-range video to people present in person. In culprit-present trials, 44% of decisions were in error, of which 60% were highly confident responses, partly explained as the consequence of the one-year time interval between video capture and identification session, reflecting alterations to the appearance of the actors. Indeed, in Experiment 2, 48% of responses were incorrect when no warning was given and the culprits in video were in no disguise, almost twice as many as with the same actors and conditions in Experiment 1 (25%). However, examination of Figure 3 indicates that changes to appearance were not substantial. Fewer errors were made in culprit-absent trials. Nonetheless, one-in-three participants still incorrectly responded that the defendant was depicted in the video, 45% of which were highly confident decisions. From Figure 3 it is clear that both actors have similar face shape and hair. However, they do not closely resemble one another.

Performance in this task was found to depend on the specific defendant present in person, whether the culprit on video was disguised or not and whether participants were warned that that the videos were a year old. As in Experiment 1, confidence was highest for correct non-choosers. However, confidence was also highest when the videoed culprit was depicted wearing no disguise and lowest in the dark glasses condition. And yet, the fewest identification errors were made in the dark glasses condition and the most in the no disguise condition in both culprit-present and culprit-absent trials.

These results are not consistent with those found in previous studies which have found that disguise interferes with the matching and recognition of faces in photographs (e.g., Diamond & Carey, 1977; Henderson et al., 2001, Hockley et al., 1999; Metzger, 2001; Terry, 1994). In Experiment 2, the full bodies of the actors in the video could be compared with the bodies of the live actors. The contradictory results suggest that disguise has less influence on identity decisions with people present in person, possibly as more cues are available than in photographs.

A second explanation is that when the actors were shown in glasses, participants were compelled to use a strategy relying more on the external features of the faces. Indeed, unfamiliar face matching has been found to be more accurate when external features are displayed than when internal features are available (e.g., Bruce et al., 1999). Such a strategy will only be effective if hairstyle is consistent. This appears to have occurred with the actors in this study whose hairstyle was relatively unchanged (Figure 3). Nevertheless, accuracy was highest in the whole face condition in the Bruce et al. study in which participants were required to match faces in arrays of ten. In the present research, there was no requirement to locate a target within an array which will have reduced task demands. These contradictory results suggest that although responses were driven by a more efficient external feature strategy, participants were also using some of the available internal features when comparing the appearance of the defendant and the video image. Indeed, performance in the hat condition did not substantially differ from when the videoed culprits were in no disguise and yet, based on previous unfamiliar face matching research (e.g., Henderson et al., 2001) a disguise that primarily covers external features would be expected to increase errors.

No generalised reduction in decision criterion from giving an age of video warning across all three disguise conditions was found, as would have been predicted if results were consistent with those found by Charman and Wells (2007) in an eyewitness study. Instead, when the culprit was depicted in dark glasses, 41% of participants made errors when no warning was given in culprit-present trials. The error rate was reduced by half to 20% with a warning, although these differences were not significant. Examination of the measure of response bias (B'', Donaldson, 1992) demonstrates that in comparison to all other conditions, participants applied a liberal criterion only when the defendant was wearing dark glasses and when they were provided with the warning. Indeed, no changes to the error rate from providing the warning were found in the other disguise conditions. This does suggest that if participants were utilising a more successful external feature strategy when the culprit was wearing glasses, the warning may have additionally reduced caution in making a positive identification decision.

However, no effects on confidence levels were found from providing the age of video warning. Nevertheless, analysis of response accuracy revealed that regardless of culprit presence, significantly more correct decisions were made when the warning was given in trials when Actor *D* was the defendant. In contrast, the warning reduced the proportion of accurate responses involving Actor *C*, although this difference was not significant. Reasons for these effects are unclear. They can’t be the result of participants making decisions based on anticipating hairstyle or other feature changes as these effects were also consistent when hairstyle was obscured by the culprit wearing a hat. Charman and Wells (2007) found similar actor-specific effects in their eyewitness lineup experiments in which participants were warned that the target’s appearance might have changed between study and test. These results do suggest caution in generalising effects observed from single actor trials. Effects can be specific to individuals.

Experiments 1 and 2 illustrated that when medium-range video footage, similar to that obtained from open-street CCTV systems is available; unfamiliar face matching is prone to error even if the target is physically present. The issue of poor-quality CCTV footage is regularly noted in media crime reports, often resulting in requests for improved systems that will provide high-quality images. Indeed, the Home Office recently reported that 80% of images examined by the police are inadequate for identification purposes, and suggesting that national quality standards are developed (Home Office, 2007). This raises the question, examined in Experiment 3, of whether face matching would still be error prone with close-up images.

Experiment 3

Experiment 3 was designed to examine identification matching using close-up high-quality footage of faces, slowly turning from left-to-right profile and back. Bruce et al. (1999) acquired similar-quality close-up videos for their studies on the same day as the photographs, describing them as the ‘gold standard’ for measuring performance in this context. Likewise, for Experiment 3, footage was taken a few minutes prior to the identification sessions. Performance in this culprit-present ‘immediate’ condition was compared with footage taken a week previously for a ‘time-lapse’ condition. Between filming sessions, the culprits shaved and slightly re-styled their hair (without cutting it), simulating minor everyday differences.

Four novel culprits were recruited for this experiment, two acted as defendants, two as distracters for culprit-absent trials. Each pair had been identified by acquaintances of the experimenter for their similarity of appearance.

During the course of a police investigation, decisions concerning offender identity in surveillance footage are likely to be made by reference to photographs. Jurors may also be asked to compare contemporaneous defendant photographs with CCTV images, particularly if the defendant’s appearance is alleged to have changed. Studies comparing eyewitness memory in which lineups are made up of actors present in person or are shown in photographs have generally found no differences or small effects that were not consistent across studies. (see Cutler et al., 1994 for a review). Nevertheless, in contrast to those studies in which memory was tested, it is not required in a perceptual matching study and there may be differences. Therefore, a final factor was included in which matching performance to the defendants present in person was compared with the same defendants shown in close-up full-face photographs taken on the day of the identification sessions.

Previous research has either found that simultaneous matching with one colour and one monochrome image is more accurate than when both images are in colour (Bruce et al., 1999; Experiment 1) or has found no differences at all (e.g., Bruce et al., 1999; Experiment 2; Davies & Thasen, 2000). Bruce et al. suggest that the contradictory findings in their experiments may have been due to a mismatch of hue across image formats. To avoid this confounding effect in Experiment 3, photographs were monochrome. The photograph identification trials replicate conditions reported by Henderson et al. (2001; Experiment 5), except that in the current experiment, video footage was moving. In their experiment, Henderson et al. found that 45% of participants, often highly confident, mistakenly believed that images of the same actor were of two different people.

For the current experiment, more errors were predicted for the culprit-present time-lapse condition than the immediate condition, as the appearance of the culprits had been slightly altered. Finally, based on the results of previous research using similar high-quality video images, substantial errors were also expected in culprit-absent trials.

Method

Participants. Participants, unfamiliar with the actors were 99 male and 277 female students, staff and visitors, aged 18 - 64 (Mean = 26.8 years) to Goldsmiths, University of London.

Design. This experiment employed a 3 (video condition) x 2 (presentation mode) x 2 (defendant) independent measures single-item identity-verification design. Participants viewed video footage under one of three Video Conditions. The footage depicted the defendant and was taken a few minutes prior to the identification session (culprit-present – immediate), it depicted the defendant, but was taken a week previously (culprit-present – time-lapse), or was of a distracter matched for similarity of appearance (culprit-absent). Presentation Mode had two levels. In the 'live' presentation mode participants had to decide whether the defendant, present in person was depicted in the video. In the 'photograph' mode, the same decision was made to a full-face monochrome photograph. Two Defendants, treated as levels of a third factor, were recruited and paired on the basis of facial similarity with two separate distracters for culprit-absent trials. Responses were collected using the IDC scale.

Actors and materials. Four white male actors aged 20 – 21 years, height 1.83m – 1.87m, weight; 60kg – 73kg were recruited for this experiment. Two actors, the defendants, were videoed twice, once on the day of the identification session for the immediate condition and once approximately a week previously for the time-lapse condition. The remaining actors were recruited to be distracters for culprit-absent trials and were videoed once only. All were filmed using a JVC GR-SX22 Compact VHS camcorder from a distance of 2m so that facial images took up approximately three-quarters of the screen. Keeping a neutral expression for about 5 seconds in each view, the actors firstly faced the camera, turned slowly for a right profile view, back to a left profile view and then faced the camera again. Images, edited so that they could be viewed in a continuous loop were transferred using Adobe Premiere Pro software into a digital format for playback using Microsoft PowerPoint.

The full-face photographs were taken from a distance of 3m with defendants adopting a neutral expression using a Samsung Fino 1O5OXL (zoom 38 - 105mm) camera on full zoom loaded with black-and-white Ilford FP4-125 35mm film. They were scanned using a Hewlett-Packard HP ScanJet 3570c and were trimmed to remove extraneous cues with Adobe Photoshop software. Images were printed onto 720 dpi photo-quality paper sized 29.7 x 21.0 cm (*A4* sheet) and presented in a plastic folder. Photographs of all four actors are shown in Figure 4.

Procedure. The procedure was similar to that described for Experiment 1. The video footage was played for an unlimited time, while in the live mode participants sat in front of the defendant while viewing the video. In the photograph mode, participants were handed a photograph to compare with the video.

Results

As described in Experiment 1, scale scores were converted for accuracy and Table 4 shows the percentage error rate as a function of presentation mode and video condition. The proportion of participants responding with highly confident incorrect responses is also listed.

TABLE 4 ABOUT HERE

Overall 28% of the converted responses were incorrect, 17% in the culprit-present immediate condition and approximately 33% in the other two conditions. A 3 (video condition) x 2 (presentation mode) x 2 (defendant) ANOVA on these data found a significant effect of video condition, $F(2, 364) = 6.31$, $p = .002$, $\eta^2 = .034$. Bonferroni-corrected t-tests found significantly fewer errors in the culprit-present immediate condition than in the time-lapse ($p < .01$, $r = .19$) and culprit-absent conditions (p < .01, $r = .20$). There were no significant differences between the culprit-present time-lapse and culprit-absent conditions ($p > .2$). The main effects of presentation mode, $F < 1$, and defendant, $F(1, 364) = 1.28$, $p = .26$, $r = .06$ were not significant.

The interaction between presentation mode and video condition was significant, $F(2, 364) = 3.49$, $p = .032$, $\eta^2 = .019$. Simple effects analyses revealed no differences between live and video presentation modes in the immediate condition, $F < 1$. However, in the time-lapse condition, marginally more errors were made in the live presentation mode, $F(1, 364) = 3.54$, $p = .061$, $r = .10$, whereas in the culprit-absent condition, marginally more errors were made in the photograph mode, $F(1, 364) = 3.41$, $p = .066$, $r = .10$)

There was a significant interaction between video condition and defendant, $F(2, 364) = 4.33$, $p = .014$, $\eta^2 = .023$. Simple effects analyses revealed no significant differences in error rates across defendants in the culprit-present immediate, $F < 1$ and culprit-absent conditions, $F(1, 364) = 2.44$, $p > .1$, $r = .08$. However, significantly more errors were made to Defendant *W* than to Defendant *Y* in the time-lapse condition (43% vs. 23% respectively), $F(1, 364) = 6.55$, $p = .011$, $r = .13$, suggesting a more substantial appearance change in the intervening period for Defendant *W*. The remaining 2-way and the 3-way interactions were not significant ($p > .1$).

To examine confidence levels, data from the separate actors were pooled to increase statistical power. A 2 (accuracy) x 2 (decision type) x 2 (presentation mode) ANOVA conducted on these data revealed a highly significant main effect of accuracy, $F(1, 368) = 24.25$, $p < .001$, $r = .25$. Correct decisions were associated with higher confidence than incorrect decisions ($M = 3.02$, $SD = 1.06$ vs. 2.32, $SD = 0.94$ respectively). The effect of decision type was also significant, $F(1, 368) = 5.63$, $p = .018$, $r = .012$. Choosers had higher confidence than non-choosers ($M = 3.00$, $SD = 1.03$ vs. 2.54, $SD = 1.09$). The effect of presentation mode and none of the interactions were significant ($p > .2$). Simple main effects analyses on the non-significant interaction between decision and accuracy found higher confidence for correct decisions than incorrect decisions in both choosers, $F(1, 368) = 13.06$, $p < .001$, $r = .18$ and non-choosers, $F(1, 368) = 11.21$, $p < .001$, $r = .17$.

In addition, two point-biserial correlations revealed highly significant relationships between confidence and accuracy in choosers, $r(231) = .25$, $p < .001$ and in non-choosers, $r(145) =.25$, $p = .002$ with no differences in correlation coefficients.

Discussion

Experiment 3 demonstrated that even with close-up images, unfamiliar face matching from moving video to either photographs or to actors present in person is prone to substantial error. When the footage was a week old, 26% of participants wrongly believed that an actor shown in high-quality video was not standing in front of them, 25% of which were highly confident incorrect decisions. Even more were incorrect when presented with a photograph (40%), with more than half (60%) being highly confident. This high error rate may be due to alterations to hairstyle and to the actors shaving their facial hair. Indeed, the error rate differed between defendants in this condition, and although both had short hair and alterations were not substantial, this suggests that one actor had changed appearance more than the other.

Seventeen per cent of participants also made errors in the culprit-present immediate condition with footage less than one-hour-old, for both defendants and presentation modes. One-third of these incorrect participants reported high levels of decision confidence and these false negative results illustrate that even in ‘gold standard’ conditions, with minimal task demands and with no alterations to appearance, face matching is surprisingly difficult. These results are consistent with the photograph-based matching research of using the same standard of video image (Bruce et al., 1999) and matching design (Henderson et al., 2001, Megreya & Burton, 2006; 2007).

One explanation for the high error rates in the immediate condition is that the participants were mainly psychology undergraduates and were cautious in attributing identity when carrying out a task that may be perceived as straightforward. Indeed, this group routinely have their perceptions tested and occasionally ‘tricked’. However, if participants were ‘suspicious’ of experimenter motives, the results of the culprit-absent trials are of even greater concern, as over 40% wrongly believed that an actor present in person was depicted in video and more than 40% of those making an incorrect decision reported high levels of confidence. Accuracy in culprit-absent trials was marginally better if defendants were depicted in photographs rather than being physically present. When taken alongside the superior performance found to actors present in person than to photographs in the culprit-present time-lapse condition, this suggests that participants were more cautious in attributing identity when presented with a photograph. However, more than 25% of decisions were still incorrect.

Consistent with Experiments 1 and 2, correct decisions were associated with a higher level of confidence. However, in contrast to the findings employing the medium-range footage, choosers were more confident than non-choosers. These results suggest that with higher resolution images, participants were more assured that their belief of a match would be correct, and yet the high error rate in the culprit-absent condition indicates that many were mistaken.

General Discussion

The main aim of the experiments reported in this paper was to determine whether the high rate of identification matching errors from video footage to photographs found in previous studies would be replicated when video was matched to somebody physically present. The latter comparison is frequently made in criminal cases involving disputed identification from CCTV. In addition, factors influential on the recognition of faces were evaluated to examine if they have similar effects in a task with no mnemonic requirements. In Experiment 1, footage was taken 3 weeks prior to the identification sessions and using a single-actor identification-verification design, 22% of participants in culprit-present conditions and 18% in culprit-absent conditions made incorrect decisions.

Experiment 2 examined the effects of disguise, and of providing or withholding the factually correct advice that the videos had been taken one year previously. Approximately 44% of participants made errors in culprit-present trials and 33% made errors in culprit-absent trials. Accuracy was highest when the videos depicted the actors in dark glasses, possibly due to participants being compelled to use an external facial feature strategy, which is known to be more effective in the recognition of unfamiliar faces (Bruce et al., 1999; Ellis et al., 1979; Young et al., 1985). In some conditions, regardless of whether the actor was present in the video or not, the age of video warning made the criterion for a positive identification less conservative, particularly when the culprit in video was wearing dark glasses. Similar effects on eyewitness memory have been found in studies examining lineup identifications (Charman & Wells, 2007). It appears that a lack of resemblance between a suspect and an image may be disregarded by knowing that the appearance of an individual may have changed. It would be of concern if such information can influence decisions in real cases.

When identity from CCTV is in dispute and footage is ‘sufficiently’ clear, juries can be invited to compare images with the defendant in the dock, or to a contemporary photograph taken at the time of the crime. In some countries, such as the UK (Attorney General’s Reference, 2003) and Australia (*Smith v The Queen*, 2001) no further evidence of identity is required. The results of previous research in this area (e.g., Bruce et al. 1999; Henderson et al., 2001; Megreya & Burton, 2006; 2007) suggest that even when images are of the highest quality, substantial errors in identification still occur. In Experiment 3, close-up facial video footage of the actors was obtained minutes prior to identification sessions and 17% of participants incorrectly responded that the same actor, either present in person or depicted in a photograph, was not present. Furthermore, there were substantial increases in error rates when footage was a week old. The results of the culprit-absent trials are of greater concern, as over 50% of participants wrongly believed that one particular actor present in person was simultaneously depicted in video when in fact a different person was shown. Of all the incorrect identification decisions in Experiment 3, 38% were highly confident.

In contrast to Experiments 1 and 2, chooser's confidence was higher than non-choosers. These results suggest that with high resolution images, participants may be less cautious in deciding that a physically present person is depicted in video, regardless of whether this is correct or not. Further research is required to examine at what point this confidence 'switch' occurs.

Nevertheless, from the results of the experiments reported in this paper, it is not possible to conclude that a jury made up of 12 different people would, following deliberation, come to a combined incorrect verdict based on presentation of CCTV. Indeed, a jury in most cases would actually view the defendant for longer than in these experiments. A meta-analysis examining simulated jury decision making found that if prior to deliberation more than 10 to 2 are in agreement of a guilty verdict, or 5 to 7 for an acquittal, the final result will normally be in that direction (Devine, Clayton, Dunford, Seying, & Pryce, 2001). This current research could be extended to examine the results within the context of decision making amongst a jury of 12 people. Preliminary work has suggested that some groups of 12 do provide an incorrect guilty verdict when informed that a case rests on identification in CCTV (Davis, 2007). In one culprit-absent trial, this involved a photograph of one actor and a video of a distracter of the quality used in Experiments 1 and 2. With the increasing use of video footage in society and in court as evidence, further research is required to investigate these processes.

There are three other circumstances in which CCTV images are admissible in a UK court. Firstly, if the identification is made by someone familiar with the defendant, they may give evidence as if an actual witness to the incident. Although a sizeable body of research has found that highly familiar faces are reliably recognised in extremely poor quality images (Bruce et al., 2001; Burton et al., 1999), performance is rarely 100% and further research is needed to determine how much contact is required before a witness could claim to be sufficiently familiar. A related form of evidence follows substantial viewing of images by a witness, not previously known to the defendant (often a police officer) presenting opinion of identity in court (Attorney General's Reference, 2003). This implicitly assumes that familiarisation can occur from inspection of images alone. Bruce et al. (1999) found evidence that learning faces in a social context may be crucial. Additional research has identified perceptual processes involved in face familiarisation (e.g., Bonner, Burton & Bruce, 2003; Clutterbuck & Johnston, 2005), and that brief exposure can substantially improve matching performance (Megreya & Burton, 2006). However, it is unclear whether from viewing images alone accuracy would be sufficient to be forensically reliable.

Finally, in the UK, experts in facial comparison may provide opinion evidence in court. However, different witnesses using similar techniques often come to different opinions as to identity (Attorney General's Reference, 2003). Some methodology has been published (e.g., Mardia, Coombs, Kirkbride, Linney, & Bowie, 1996), although this is not typical of the analyses presented in court and often poor image quality precludes application. Alternative techniques used by practitioners in court have been published (Davis & Valentine, in preparation: Halberstein, 2001; Porter & Doran, 2000). However, no comprehensive investigation of the distribution of facial measurements in the population has been published. Without such data, assessments of the probability that two different images are of the same person can only be estimates. Trial judges have noted the shortcomings of current practice in facial comparison and recommended further research to ensure evidence is reliable (e.g., *R v Gray*, 2003). Research is currently being undertaken to develop computer software that can perform automatic face recognition. A series of US Government sponsored tests on commercial systems have been undertaken against large databases of two and three-dimensional facial images (Phillips, Scruggs, O'Toole, Flynn, Bowyer, Schott, & Sharpe, 2007). In ideal conditions performance has been better than human observers. However, performance is reduced considerably when faces are obscured, viewed from different angles or when environmental lighting effects are changed. Therefore for the foreseeable future, CCTV evidence in court is likely to be reliant on identifications by humans.

Notwithstanding the findings reported within this paper, CCTV images provide a permanent record of events, potentially act as a crime deterrent and to some extent overcome the errors in memory that are often found with eyewitness evidence. However, in common with other researchers expressing concern in this field (e.g., Bruce et al., 1999; Bruce et al., 2001; Davies & Thasen, 2000), conclusions from the experiments reported here suggest that inviting juries to compare a defendant with CCTV images as the only form of identification evidence would appear to be a risky policy, and on the basis of current research should be avoided, even if the footage is of the highest quality.

Footnotes

1. An extra culprit (Actor *I* in Figure 1) was filmed to enhance the likelihood that each defendant would be matched with a physically similar distracter. However, although the video of this culprit was used in a culprit-absent trial, this actor did not attend an identification session as a defendant.

2. A second similar ANOVA was conducted on the data from the 136 participants who attended a second session. A highly significant main effect of accuracy was again obtained, $F(1, 132) = 24.10$, $p < .001$, $r = .39$ with higher confidence levels associated with correct responses ($M = 3.07$, $SD = 0.97$) than with incorrect responses ($M = 2.04$, $SD = 1.04$). In addition, in contrast to the non-significant effect found with the first or only session data , the main effect of decision was also highly significant, $F(1, 132) = 14.42$, $p < .001$, $r = .30$. In this case, non-choosers ($n = 76$, $M = 3.21$, $SD = 0.98$) were more confident than choosers ($n = 60$, $M = 2.48$, $SD = 1.02$). The interaction was not significant, $F < 1$). Simple main effects analyses found higher confidence for correct decisions than incorrect decisions in both choosers, $F(1, 132) = 13.84$, $p < .001$, $r = .31$ and for non-choosers, $F(1, 132) = 10.28$, $p = .002$, $r = .27$. The discrepancy between the two analyses in terms of the main effect of decision making being significant only in the second session can be explained as a result of a large minority (34%) of the second session participants taking part in the culprit-absent trial involving Defendant *G* and Culprit *I*. In this trial, no participant made a false positive identification error and most made highly confident decisions to what was the most easily distinguishable pairing (see Figure 1).

3. Similar highly significant confidence-accuracy correlations were found in choosers, $r(60) = .44$, $p < .001$ and non-choosers, $r(76) = .35$, $p < .005$ with the second session data.

References

Attorney General's Reference [No.2 of 2002], 2003. 1 Cr. App. R. 321, England.

Bonner, L., Burton, A.M., & Bruce, V. (2003). Getting to know you: How we learn new faces. *Visual Cognition, 10*, 527–536.

Bruce, V., Henderson, Z., Greenwood, K., Hancock, P.J.B., Burton, AM., & 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., Wilson, S., Cowan, M., & Bruce, V. (1999). Face recognition in poor-quality video: Evidence from security surveillance. *Psychological Science, 10*, 243-248.

Charman, S.D., & Wells, G.L. (2007). Eyewitness lineups: Is the appearance-change instruction a good idea? *Law and Human Behavior, 7*, 3-22.

Christie, F., & Bruce, V. (1998). The role of dynamic information in the recognition of unfamiliar faces. *Memory and Cognition, 26,* 780-790.

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

Cohen, M.E., & Carr, W.J. (1975). Facial recognition and the von Restorff effect. *Bulletin of the Psychonomic Society, 6*, 383–384.

Cutler, B.L., Berman, G.L., Penrod, S.D., & Fisher, R.P. (1994). Conceptual, practical and empirical issues associated with eyewitness identification test media. In D.F. Ross, J.D. Read and M.P. Toglia (Eds.), *Adult Eyewitness Testimony: Current Trends and Developments* (pp. 163–181). Cambridge: Cambridge University Press.

Cutler, B.L., & Penrod, S.D. (1995). *Mistaken identification: The eyewitness, psychology and the law*. Cambridge: Cambridge University Press.

Davies, G.M., Shepherd, J.W., & Ellis, H.D. (1979). Similarity effects in face recognition. *American Journal of Psychology, 92*, 507-523.

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

Davis, J.P. (2007). *The forensic identification of CCTV images of unfamiliar faces*. Unpublished PhD Thesis. Goldsmiths, University of London.

Davis, J.P., & Valentine, T. (submitted). CCTV on trial II: Facial mapping of photographic images to compare with the defendant.

Devine, D.J., Clayton, L.D., Dunford, B.B., Seying, R. & Pryce. J. (2001). Jury Decision Making: 45 Years of Empirical Research on Deliberating Groups. *Psychology, Public Policy, and Law 7*, 622-727.

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

Donaldson, W. (1992). Measuring recognition memory. *Journal of Experimental Psychology: General, 121*, 275-277.

Ellis, H.D., Shepherd, J.W., & Davies, G.M. (1979). Identification of familiar and unfamiliar faces from internal and external features: Some implications for theories of face recognition. *Perception, 8*, 431–439.

Gill, M., & Loveday, K. (2003). What do offenders think about CCTV? In M. Gill (Ed.), *CCTV* (pp. 81-91). Leicester: Perpetuity Press.

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 and Cognition, 27*, 128–138.

Home Office (2007). *National CCTV Strategy.* London: Home Office.

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

Knight, B., & Johnston, A. (1997). The role of movement in face recognition. *Visual Cognition, 4*, 265–273.

Lander, K., Bruce, V., & Hill, H. (2001). Evaluating the effectiveness of pixelation and blurring on masking the identity of familiar faces. *Applied Cognitive Psychology, 15,* 101-116.

Lander, K., & Chuang, L. (2005). Why are moving faces easier to recognize? *Visual Cognition, 12*, 429–442.

Light, L.L., Kayra-Stuart, F., & Hollander, S. (1979). Recognition memory for typical and unusual faces. *Journal of Experimental Psychology: Human Learning and Memory, 5*, 212–228.

Liu, C.H., Seetzen, H., Burton, A.M., & Chaudhuri, A. (2003). Face recognition is robust with incongruent image resolution: Relationship to security video images. *Journal of Experimental Psychology: Applied, 9*, 33-41.

Mardia, K.V., Coombs, A., Kirkbride, J., Linney, A., & Bowie, J.L. (1996). On statistical problems with face identification from photographs. *Journal of Applied Statistics, 23*, 655–676.

Megreya, A.M., & Burton, A.M. (2006). Unfamiliar faces are no faces: Evidence from a matching task. *Memory & Cognition, 34*, 865-876.

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

Metzger, M.M. (2001). Which transformations of stimuli are the most disruptive to facial recognition? *Perceptual and Motor Skills, 92*, 517–526.

Newell, F.N., Chiroro, P., & Valentine, T. (1999). Recognising unfamiliar faces. The effects of distinctiveness and view. *Quarterly Journal of Experimental Psychology, 52A*, 509–534.

Norris, C., McCahill, M., & Wood, D. (2004). Editorial. The growth of CCTV: A global perspective on the international diffusion of video surveillance in publicly accessible space. *Surveillance & Society, 2*, 110-135

O'Donnell, C., & Bruce, V. (2001). Familiarisation with faces selectively enhances sensitivity to changes made to the eyes. *Perception, 30*, 755-764.

Police and Criminal Evidence Act (PACE) (1984). London: HMSO.

Phillips, P. J., Scruggs, W.T., O'Toole, A.J., Flynn, P.J., Bowyer, K.W., Schott, C.L., &

Rosenthal, R. (1991). *Meta-analytic procedures for social research.* Newbury Park, CA: Sage.

Sharpe, M. (2007). *FRVT 2006 and ICE 2006 large-scale results*. Retrieved April 15 2007, from http://iris.nist.gov/ice/FRVT2006andICE2006LargeScaleReport.pdf

R v Gray (2003). EWCA Crim. 1001

Rattner, A. (1988). Convicted but innocent: Wrongful conviction and the criminal justice system. *Law and Human Behavior, 12*, 283–293.

Rhodes, G., Brennan, S., & Carey, S. (1987). Identification and ratings of caricatures: Implications for mental representations. *Cognitive Psychology, 19*, 473-497.

Scheck, B., Neufeld, P., & Dwyer, J. (2000). *Actual innocence: Five days to execution and other dispatches from the wrongly convicted.* New York: Doubleday.

Schiff, W., Banka, L., & de Bordes Galdi, G.D. (1986) Recognising people seen in events via dynamic "mug shots". *American Journal of Psychology, 99,* 219–231.

Shapiro, P.N., & Penrod, S. (1986). Meta-analysis of facial identification studies. *Psychological Bulletin, 100*, 139-156.

Smith v The Queen (2001). HCA 50, Australia.

Sporer, S.L., Penrod, S., Read, D., & Cutler, B. (1995). Choosing, confidence and accuracy: A meta-analysis of the confidence-accuracy relation in eyewitness identification studies. *Psychological Bulletin, 118,* 315–327.

Technical Working Group for Eyewitness Evidence (1999). *Eyewitness evidence: A guide for law enforcement*. Washington, DC: United States Department of Justice, Office of Justice Programs.

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

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

Valentine, T. (2001). Face-space models of face recognition. In M.J. Wenger & J.T. Townsend (Eds.), *Computational, geometric, and present perspectives on facial cognition: Contexts and challenges*. Hillsdale, New Jersey: Lawrence Erlbaum Associates Inc. pp. 83-113.

Valentine, T., & Bruce, V. (1986). The effects of distinctiveness in recognising and classifying faces. *Perception, 15*, 525–535.

Valentine, T., & Endo, M. (1992). Towards an exemplar model of face processing: the effects of race and distinctiveness. *Quarterly Journal of Experimental Psychology, 44A*, 671-703.

Washington Post. (8 October 2005). Security camera new witness. Retrieved 10 October 2005 from, http://www.washingtonpost.com/wp-yn/content/article2005/10/07/AR2005100701895

Wells, G.L., Memon, A., & Penrod, S.D. (2006). Eyewitness evidence: Improving its probative value. *Psychological Science in the Public Interest, 7*, 45-75.

Wells, G.L., Olson, E., & Charman, S. (2002). Eyewitness identification confidence. *Current Directions in Psychological Science, 11*, 151–154.

Wells, G. L., Small, M., Penrod, S., Malpass, R. S. Fulero, S. M. & Brinacombe, C. A. E. (1998). Eyewitness identification procedures: recommendations for lineups and photospreads. *Law and Human Behavior 22*, 603-647.

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

Table 1

Percentage error rate and percentage of highly confident errors to each defendant in participant's first/only and second trials in Experiment 1.

Session	Culprit-present			Culprit-absent		
	n	Incorrect Trials	Confident Errors	n	Incorrect Trials	Confident Errors
Defendant *A*						
1st/only	13	0	0	3	67	0
2nd	0	-	-	13	38	15
Total	13	0	0	16	44	12
Defendant *B*						
1st/only	30	13	0	20	35	5
2nd	0	-	-	0	-	-
Total	30	13	0	20	35	5
Defendant *C*						
1st/only	0	-	-	2	0	0
2nd	19	21	10	15	40	0
Total	19	21	10.	17	35	0
Defendant *D*						
1st/only	0	-	-	21	23	9
2nd	20	25	20	0	-	-
Total	20	25	20	21	24	9
Defendant *E*						
1st/only	17	29	12	10	30	0
2nd	0	-	-	0	-	-
Total	17	29	12	10	30	0
Defendant *F*						
1st/only	24	37	17	21	5	5
2nd	0	-	-	0	-	-
Total	24	37	17	21	5	5
Defendant *G*						
1st/only	13	31	8	3	0	0
2nd	12	0	0	46	0	0
Total	25	16	4	49	0	0
Defendant *H*						
1st/only	0	-	-	21	5	0
2nd	11	36	9	0	-	-
Total	11	36	9	21	5	0
Grand Mean						
1st/only	97	23	7	101	19	4
2nd	62	21	11	74	15	3
Total	159	22	9	175	17	3

Note: Confident errors refer to responses on the two extreme intervals of the IAC scale.

Table 2

Mean correct and incorrect confidence level data in Experiment 1 as a function of first or only session participant decisions (Max = 4; SD in parentheses)

	Choosers		*Non-choosers*	
	Correct *n = 75*	*Incorrect* *n = 19*	*Correct* *n = 82*	*Incorrect* *n = 22*
Confidence	*2.79* *(0.89)*	*1.89* *(0.74)*	*2.85* *(1.00)*	*2.09* *(1.15)*

Table 3

Percentage error rate and highly confident errors and response bias (B'') in Experiment 2 as a function of disguise and whether an age of video warning was given or not.

	Culprit-present		Culprit-absent		B''		Total		Grand
Warning	No	Yes	No	Yes	No	Yes	No	Yes	Mean
No disguise									
n	46	50	61	43			107	93	200
Incorrect Trials	48	50	39	21	.17	.57	43	45	44
Conf Errors	33	34	39	19			26	27	26
Glasses									
n	54	45	43	50			97	95	192
Incorrect Trials	41	20	23	34	.39	-.35	33	27	30
Conf Errors	22	11	5	6			44	8	11
Hat									
n	53	45	51	50			104	95	199
Incorrect Trials	55	49	33	28	.41	.42	44	38	41
Conf Errors	28	31	12	16			20	23	22
Total									
n	153	140	155	143			308	283	
Incorrect Trials	48	40	33	34	.30	.14	40	37	
Conf Errors	27	26	13	13			20	19	
Grand Mean									
n	293		298						
Incorrect Trials	44		33		.23				
Conf Errors	27		13						

Note: Yes = age of video warning given; No = age of video warning not given. Conf = confident errors refer to responses on the two extreme intervals of the IAC scale.

.

Table 4
Percentage error rate and highly confident errors to each defendant in Experiment 3 as a function of video condition and presentation mode

Presentation Mode	Culprit-present Immediate		Culprit-present Time Lapse		Culprit-absent Distracter	
Defendant	*W*	*Y*	*W*	*Y*	*W*	*Y*
Live Mode						
n	32	31	32	30	32	31
Incorrect Trials	19	13	34	17	31	52
Conf Errors	12	3	9	3	12	23
Photograph Mode						
n	32	31	31	31	32	31
Incorrect Trials	22	13	52	29	25	29
Conf Errors	3	3	35	13	6	29
Grand Mean						
n	126		124		126	
Incorrect Trials	17		33		34	
Conf Errors	6		15		17	

Note: Conf = confident errors refer to responses on the two extreme intervals of the IAC scale.

Figure 1
Stills from the colour video footage taken of Actor F (top) as well as full-face photographs (below) taken at the time of the identification three weeks later of both Actor F (bottom left) and of Actor A (bottom 2nd left) who was the matched defendant in culprit-absent trials when this footage was presented. In both culprit-present (37%) and in culprit-absent trials (44%), this footage was associated with the highest proportion of errors. Actor G (bottom 2nd right) and Actor I (bottom right) – matched in the only culprit-absent trial in which no errors were made are also depicted.

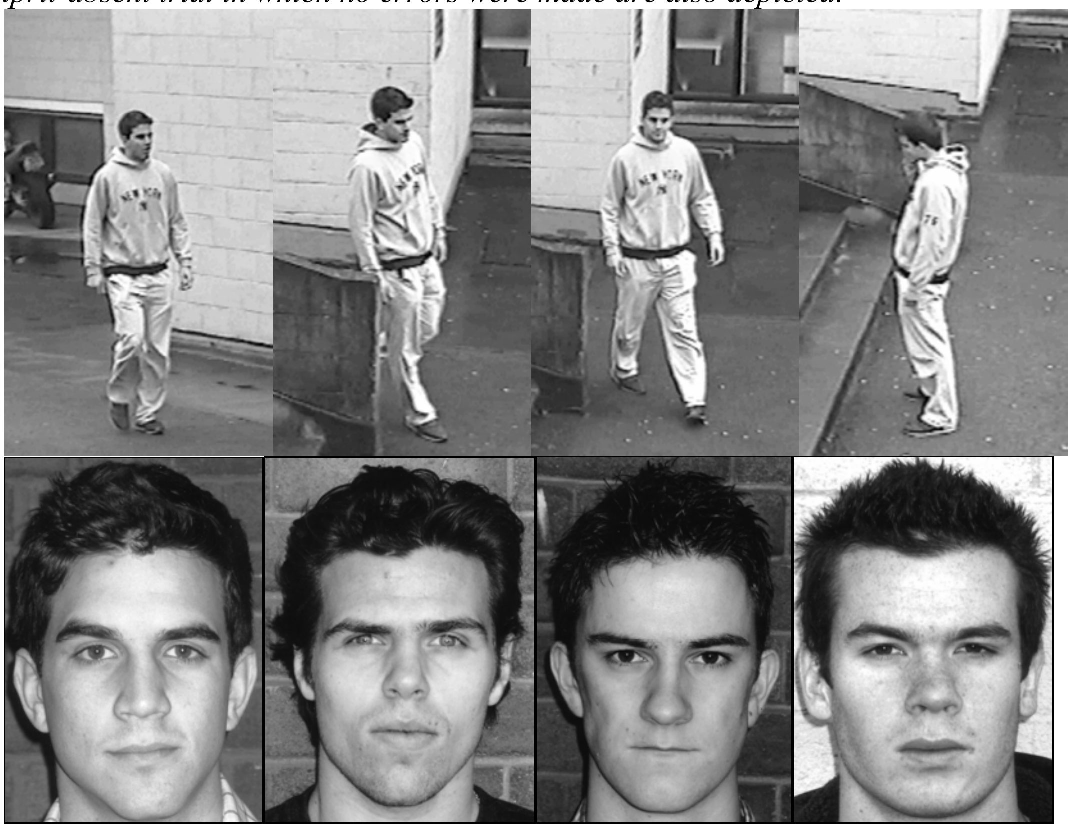

Figure 2

Disguise conditions in Experiment 2

Figure 3
Photographs of Actors C (left) and E taken at the time of the original video filming sessions; and together at the time of the Experiment 2 identification sessions, 11-months later.

Figure 4

Photographs of the four actors recruited for Experiment 3 (from left to right: Defendant W with matched Culprit X; and Defendant Y with matched Culprit Z).