Matching Identities of Familiar and Unfamiliar Faces Caught on CCTV Images

Vicki Bruce, Zoë Henderson, and Craig Newman
University of Stirling

A. Mike Burton
University of Glasgow

People can be inaccurate at matching unfamiliar faces shown in high-quality video images, even when viewpoint and facial expressions are closely matched. However, identification of highly familiar faces appears good, even when video quality is poor. Experiment 1 reported a direct comparison between familiar and unfamiliar faces. Participants who were personally familiar with target items appearing on video were highly accurate at a verification task. Unfamiliar participants doing the same task performed very inaccurately. Familiarity affected discriminability, but not bias. Experiments 2 and 3 showed that brief periods of familiarization have little beneficial effect unless “deep” or “social” processing is encouraged. The results show that video evidence can be used effectively as a probe to identity when the faces shown are highly familiar to observers, but caution should be used where images of unfamiliar people are being compared.

The widespread use of closed-circuit television (CCTV) cameras to survey areas where there is potential for crime or public disorder has a number of important consequences for criminal investigation. A video image captured at the scene of a crime can be informative about what happened in a crime (e.g., was there a weapon present or not?) and about who the perpetrators were (e.g., how many, what ages, what gender, what race, etc.). Often, CCTV images are used as evidence for an important final stage in an investigation—that of establishing the identity of the person who committed the crime. It is this latter topic that forms the subject of this article.

Eyewitness identification of faces is notoriously error prone (e.g., see Wells, 1993, for a review), and CCTV images have the potential for removing the need to rely on witness memories for faces at all. In theory, once a CCTV image of a face is recorded, it should be possible to compare this with the appearance of a suspect to verify that this person was indeed linked with this crime. However, recent experiments have pointed out that apparent resemblance between images of unfamiliar faces can be misleading.

Just as witnesses’ memory for faces can be prone to errors, so can people’s perceptions of faces.

Kemp, Towell, and Pike (1997) examined how accurately supermarket cashiers could verify the authenticity of photo-credit cards carried by simulated customers. When the customers carried cards containing a photograph of someone bearing some resemblance to them, they were challenged less than 40% of the time. Bruce et al. (1999) asked participants to compare an image taken from high-quality video footage against a photographic array showing studio portraits of young men all of similar appearance to the target. The target was present in 50% of these arrays (Experiment 1) or in all of them (Experiments 2–4). Errors were made on a substantial proportion of trials even when the image grabbed from video matched the angle and expression of the array photographs. When viewing angle altered, performance dropped further. There were no clear advantages for performance in the matching task when participants viewed animated video clips rather than still images extracted from these.

Henderson, Bruce, and Burton (2001) used footage of a simulated bank robbery to confirm these findings. Typical (i.e., poor) quality CCTV images of two men acting the parts of robbers in a staged armed robbery were shown to participants alongside arrays showing pictures of similar-looking men, within which each robber appeared at one location. Only a minority of participants were correct in their choices. When high-quality (broadcast TV) footage of the same incident was used to provide images of the robbers, performance using the same arrays increased but was only correct on 64% of the trials overall—very comparable with the data obtained by Bruce et al. (1999) using a similar task. Henderson et al. also included studies where the task was to match a video image against just two faces (forced choice—Which one is the robber?) or against a single face (Is this the robber or not?) and showed that even using this kind of task error rates were high. Henderson et al.’s study confirmed that even with good-quality video it can be
very difficult to match accurately with photographs taken on another camera.

In contrast to these difficulties with unfamiliar face matching, Burton, Wilson, Cowan, and Bruce (1999; Experiment 2) showed that identification of familiar faces from CCTV images could be extremely accurate, even when image quality was poor, provided that the video image revealed the face of the target. In Experiment 1, Burton et al. (1999) used a recognition memory test to compare performance with familiar and unfamiliar faces directly. Participants were shown a series of video clips from poor-quality CCTV and later were shown good-quality photographs—their task was to decide which were the people who had been shown on video. Participants familiar with the items used in this study performed almost perfectly at the task, whereas people unfamiliar with them performed only slightly better than chance. However, this comparison of familiar with unfamiliar face images used a task with little forensic validity. Moreover, it might be that the limiting factor determining the good performance of the familiar group was that remembered names could be used to do the recognition memory task, whereas the unfamiliar subjects were forced to rely on remembered faces. It is important to establish whether the benefits of familiarity are found when a more natural perceptual task is used that makes no demands on memory. Moreover, we need to establish whether any benefits of familiarity arise from better discrimination of images of familiar faces or whether familiarity may also bias performance in certain circumstances. Police officers sometimes claim to recognize a person shown on a CCTV image of very imperfect quality. If officers expect a particular villain to appear in a certain context, will this incline them falsely to identify someone who just resembles him?

In this article, we report on three experiments that explored the effects of familiarity on perceptual matching tasks, in which a person shown on video was compared with a single comparison face (Experiments 1 and 2) or with an array of faces (Experiment 3). In Experiment 1, we investigated the effects of high levels of familiarity, using low-quality video images of targets who were the teachers and colleagues of one group of participants in the experiment. Effects of video format (animated or still) and bias were also investigated in Experiment 1. Following the substantial benefits of familiarity observed in Experiment 1, in Experiments 2 and 3 we examined whether brief periods of item familiarization can also yield benefits in perceptual matching tasks. In Experiment 2 we used low-quality video targets and a single-item verification task, and in Experiment 3 we used higher quality video targets in a match-to-array task.

In Experiment 1, we also reexamined the effects of animation on the accuracy of matching both familiar and unfamiliar items. Burton et al. (1999) found little or no benefit in a video-arrays matching task if an animated video was used in place of the best still image extracted from it. In further unpublished work conducted since then, Henderson (1999) used a single-item verification task in which a video clip or still image was shown alongside a single full-face comparison photograph. There was no benefit for the moving image over the best (full-face) still condition (overall performance averaged over match and mismatch trials of differing difficulty was as follows: moving, 88% correct; full-face still, 89% correct; angled still, 84% correct). Moreover, research into recognition memory for unfamiliar faces has also found that animated presentation did not lead to better performance than still presentation (Christie & Bruce, 1998; but see Pike, Kemp, Towell, & Phillips, 1997, for evidence that movement can help in such a task). The balance of evidence in past research led us to expect no benefit in the video-matching task from viewing video clips versus the best still image(s) extracted from that clip.

However, for the matching of familiar faces there might be some benefit of movement. Several studies have shown that when famous faces are to be identified from difficult-to-recognize images (e.g., photographic negatives, or one-bit-per-pixel “thresholded” images), people were more successful when shown video clips than when shown single stills (Knight & Johnston, 1997; Lander & Bruce, 2000; Lander, Christie, & Bruce, 1999; Lander, Bruce, & Hill, 2001). Lander et al. (1999) and Lander and Bruce (2000) have provided evidence that the effect may arise in part because viewers can make use of characteristic patterns of facial movement to help to identify celebrities in such circumstances. Because CCTV images of poor quality would appear difficult to recognize, perhaps it would be beneficial to show these moving when the people shown on them are familiar. In the experiment that follows, we compare accuracy of verification performance when the people shown on video appear moving, in three separate stills (multistill), or in the best single still.

The final variable investigated in the experiment concerned the effects of bias. Although familiarity may make it easier to confirm identity, it is also possible that its effects may operate through bias rather than discriminability. Familiarity may make participants more ready to take resemblance to a known person to signal identity. To explore effects of bias in this study, we analyzed data in terms of bias as well as discriminability, but we also manipulated bias by giving half our participants false expectations that matches would be more frequent than they actually were. We particularly wanted to see whether such expectations of matching interacted with the predicted effects of familiarity.

In summary, the experiment manipulated familiarity (participants were either familiar or unfamiliar with a set of video targets), format (video images shown moving, multistill, or single still), and bias (participants were informed that match trials were frequent or infrequent). In each trial, participants compared a video (either of a target—an item chosen to be familiar to some participants—or of an unfamiliar distractor person) with a photograph (either of a target or of a nontarget face, chosen to resemble one of the targets) to decide whether the two matched or not. The design of the study allowed us to examine whether familiarity affected rates of matching or mismatching (false positives).
Participants

One hundred twenty participants were tested. Sixty participants were recruited from the University of Glasgow, including senior undergraduates (forming over 50% of this group of participants), postgraduates, and a small number of staff members (excluding those whose images appeared in the experiment). The remaining 60 participants were undergraduates recruited at the University of Paisley in Paisley, Scotland, who were unfamiliar with any people used in the experiment.¹

Design

We examined the effects of familiarity between participants. Half our participants should have been familiar with the 12 target people but not with other people who appeared in the experiment.² The remaining 60 participants were unfamiliar with any of the faces used in the experiment. We also varied the type of footage between participants. Twenty in each familiarity group saw video images as moving clips, 20 were shown the single best still image, and 20 were shown three different still images selected from the video. A final between-participants factor varied the instructions given to participants. Half the participants were told that video and photographs matched (showed the same person) on “about half” the trials. The remaining participants were told, quite correctly, that matches would be “rare” (they actually occurred in 25% of trials). Thus, there were 10 participants tested in each cell of the design.

Each participant saw 48 trials, 12 of each of four types:

1. Target–target match trials (25%). A video target plus photo of the target (e.g., see Figure 1A and 1B). (Target items were people who were familiar to one group of participants but unfamiliar to the other.)
2. Target–nontarget mismatch trials (25%). A video target plus matched nontarget photo (e.g., see Figure 1A and 1D). (Nontarget items were unfamiliar to both groups of participants.)
3. Distractor–target mismatch trials (25%). Distractor video plus target photo (e.g., Figure 1C and 1B). (Distractor items were unfamiliar to both groups of participants.)
4. Distractor–nontarget mismatch trials (25%). Distractor video plus nontarget on photo (e.g., Figure 1C and 1D). (Both types of item should have been unfamiliar to both groups of participants.)

This last type of mismatching trial, which showed only unfamiliar faces to both groups of participants, provided a means of assessing their baseline performance. All faces appeared in each condition of the experiment, thus images and photos of targets and other faces were repeated during the course of the trial series, though not in the same combination.

Materials and Trial Types

Video images. The experiment made use of an actual CCTV video camera, which records the entry of every person arriving at the Department of Psychology, University of Glasgow. The CCTV system delivers images of relatively low quality, typical of security surveillance systems currently in use. All people shown on video gave their permission for footage to be used in these experiments. Twelve target academic staff who should be familiar to Glasgow participants were filmed entering the Psychology Department. Twelve distractor people who were not members of the department and who should be unfamiliar to all participants were also filmed. Each of these distractors was matched on gender and as closely as possible on other characteristics with one of the targets—but, however, the constraints of the situation (i.e., getting unfamiliar people filmed on the same camera as targets and gaining their permission to use footage) meant that the resemblances were not generally very close.

The monochrome video footage was digitized using Media 100 software on a Macintosh 8100/100 computer from which the three video formats were edited. These consisted of a 3-s moving clip depicting the target’s natural gait, a multistill clip containing three 1-s stills showing disjointed movement, and the best single still, which was extended for 3 s (Figures 1A and 1C give examples of the best stills). The animated and still images showed the full body of the person as well as his or her face. We have already established that it is the information from the face rather than the

Figure 1. Examples of materials used in Experiment 1. Top left, a frame of a familiar target on CCTV; bottom left, a digital photograph of the same person for target match trials; top right, the distractor chosen to be approximately matched to the appearance of this target; bottom right, the nontarget photograph, chosen to match the appearance of this familiar target.

¹ The experimenter was equally unfamiliar to both groups of participants, as she was a research assistant normally based at the University of Stirling, who traveled to recruit and test participants at the University of Glasgow and the University of Paisley for the purposes of this experiment only.

² In fact, a minority of those recruited from the Glasgow staff and postgraduates had encountered some of the other faces (e.g., if a nontarget face showed a member of Stirling faculty with whom a Glasgow faculty member was familiar), but such instances were rare and, in any case, made no material difference to the results. Reanalysis excluding such participants made no difference to the pattern of effects observed (see Experiment 1, Results section).
body that is most useful for identifying familiar people from such images (Burton et al., 1999). People on the video images measured approximately 10 × 15 cm on the screen.

Still images. A photograph of each of the 12 targets was taken. The photographs were color images of the face only, which were printed out to measure approximately 8 × 10 cm. These photographs were used as comparison images on target–target match trials (e.g., see Figure 1A shown with Figure 1B). To create a set of mismatch trials that would potentially lead to confusions, we compared each target photograph with photographs available of staff and postgraduate students at the University of Stirling and obtained for each target a matching nontarget face that was selected to have a resemblance as strong as possible to one of the targets (e.g., similar age, hairstyle; see Figure 1D for the nontarget selected for the target shown in Figures 1A and IB). The target (low-quality CCTV) with nontarget (high-quality photo) mismatching trials provided an opportunity to see whether familiarity might increase, or protect against, false-positive responses.

The (low-quality) distractor videos could be presented with (high-quality) photographs of the targets to examine the effects of familiarity on rates of correct rejection. However, as it was not possible to find and seek quality photographs of the targets to examine the effects of familiarity on performance on each type of trial. Figure 2 plots the overall percentage of correct responses in each condition of the experiment.

Procedure

Participants on each trial were shown either a 3-s video clip, three still images, or a single still image shown in black and white on a Sony 12-in. (0.3048-m) television monitor and video recorder. Viewers sat approximately 1 m away from the screen. While viewing the video footage, which they were free to pause and replay as they wished, participants were presented with a color print of a digital photograph of either the same or a different person. On each trial, participants could take as long as they wished to make their decision. They signaled their decision on a scale ranging from 1 (definitely not the person) to 10 (definitely is the person) and, using this scale, were forced to choose either 5 (not the person) or 6 (is the person) even when they felt unsure of their decision.

Results

We analyzed results in three different ways: overall correct performance, signal-detection analysis of discriminability and bias, and confidence analysis.

Overall Correct Performance

We first mapped numerical confidence scores onto discrete match and mismatch response categories and examined the percentage of correct responses in each condition of the experiment. These means are shown in Table 1, and Figure 2 plots the overall effects of familiarity on performance on each type of trial. Figure 2 indicates that performance was much more accurate for the familiar than for the unfamiliar participants in all conditions except distractor–nontarget mismatching trials, where the advantage for familiar participants was smaller.

### Table 1

<table>
<thead>
<tr>
<th>Format</th>
<th>Target–target (Correct yes)</th>
<th>Target–nontarget (Correct no)</th>
<th>Distractor–target (Correct no)</th>
<th>Distractor–nontarget (Correct no)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Fam</td>
<td>Unf</td>
<td>Fam</td>
<td>Unf</td>
</tr>
<tr>
<td>Mov</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>M</td>
<td>95</td>
<td>83</td>
<td>88</td>
<td>49</td>
</tr>
<tr>
<td>SD</td>
<td>8</td>
<td>9</td>
<td>18</td>
<td>7</td>
</tr>
<tr>
<td>Multi</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>M</td>
<td>97</td>
<td>79</td>
<td>90</td>
<td>52</td>
</tr>
<tr>
<td>SD</td>
<td>6</td>
<td>10</td>
<td>18</td>
<td>11</td>
</tr>
<tr>
<td>Still</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>M</td>
<td>93</td>
<td>73</td>
<td>92</td>
<td>60</td>
</tr>
<tr>
<td>SD</td>
<td>7</td>
<td>16</td>
<td>10</td>
<td>17</td>
</tr>
<tr>
<td>Total</td>
<td>95</td>
<td>78</td>
<td>90</td>
<td>54</td>
</tr>
</tbody>
</table>

“About half” instructions

| Mov    |     |     |     |     |     |     |     |     |
| M      | 88  | 65  | 95  | 59  | 100 | 84  | 91  | 76  |
| SD     | 16  | 17  | 9   | 9   | 0   | 10  | 8   | 10  |
| Multi  |     |     |     |     |     |     |     |     |
| M      | 85  | 71  | 93  | 52  | 98  | 72  | 88  | 73  |
| SD     | 18  | 18  | 13  | 17  | 6   | 15  | 7   | 9   |
| Still  |     |     |     |     |     |     |     |     |
| M      | 92  | 75  | 96  | 62  | 97  | 78  | 90  | 78  |
| SD     | 9   | 24  | 7   | 14  | 6   | 18  | 8   | 13  |
| Total  | 88  | 70  | 95  | 58  | 98  | 78  | 90  | 76  |

Grand mean | 92 | 74 | 92 | 56 | 97 | 74 | 87 | 75 |

Note. Fam = familiar targets; Unf = unfamiliar targets; Mov = video images as moving clips; Multi = three different still images selected from the video; Still = single best still image selected from the video.
These rates of response were analyzed with a 2 (familiarity of the targets) × 2 (instruction type) × 3 (image format) × 4 (trial type) analysis of variance (ANOVA), separately by participant ($F_1$) and by items ($F_2$). This showed a highly significant effect of familiarity with target items—$F_1(1, 108) = 314.95, MSE = 179.1, p < .01$; $F_2(1, 11) = 68.59, MSE = 1,023.9, p < .01$—with mean correct responses higher when participants were familiar with targets (92%, $SD = 11.1$) than when unfamiliar (70%, $SD = 16.4$).

The effect of target familiarity interacted with that of trial type: $F_1(3, 324) = 24.67, MSE = 141.7, p < .01$; $F_2(3, 33) = 6.09, MSE = 722.1, p < .05$ as shown in Figure 2. When target faces were familiar, performance was worse in the distractor–nontarget trials ($M = 86%, SD = 9.8$) than in the distractor–target trials ($M = 97%, SD = 7$), which could benefit from the target familiarity. When all faces were unfamiliar, however, there was no difference between performance in these two conditions (distractor–nontarget, $M = 75%, SD = 10.5$; distractor–target, $M = 74%, SD = 15.2$), suggesting that these kinds of trials were equivalent in their degree of visual resemblance. However, for unfamiliar participants, performance was significantly worse in the target–nontarget mismatching trials ($M = 56%, SD = 13.3$), where the pairs of faces bore greater similarity to each other. The higher performance on the distractor–nontarget trials that was found for the familiar (86%) compared with the unfamiliar (75%) participants was unexpected, and we return to discuss this later.

There was also a significant interaction between instructions and trial type: $F_1(3, 324) = 7.18, MSE = 141.7, p < .01$; $F_2(3, 33) = 8.88, MSE = 140.1, p < .01$. Performance on target–target (match) trials was significantly higher when “about half” instructions were given ($M = 87%$) than when “rare” instructions were given ($M = 80%$). However, the reverse trend was shown in target–nontarget mismatch trials, where performance was significantly better with the rare instructions ($M = 88%$) than with the half instructions ($M = 83%$). There was no significant effect of instructions on the remaining two types of mismatching trial. The effect of format (moving, multistill, or single still) was significant on the items analysis only, $F_1(2, 108) = 2.305, MSE = 179.1, p > .01$; $F_2(2, 22) = 4.56, MSE = 110.5, p < .05$, with the single still condition (82%) marginally better than the moving (81%) and multistill (79%) conditions.

**Signal Detection Analysis**

Next, we combined “hits” (correct responses to target–target trials) with “false alarms” (incorrect responses to each of the other kinds of trial) to yield discriminability ($A'$) and bias ($B'$) for the task of distinguishing match from nonmatch trials of different types, using nonparametric measures of these (Hodos, 1970). Familiarity with targets significantly affected discriminability, $F(1, 108) = 158.49, MSE = 0.02, p < .001$; mean $A'$ was 0.93 for familiar participants and 0.75 for unfamiliar participants. Target familiarity had no effect on bias; $F < 1$, mean $B'$ was $-0.002$ for familiar participants and $-0.043$ for unfamiliar participants. In contrast, our manipulation of instructions significantly affected bias in the predicted direction (more conservative responses when match trials were described as rare: $F(1, 108) = 8.28, MSE = 0.20, p < .01$; mean $B'$ half instructions $= 0.090$, rare instructions $= 0.045$. However, they had no effect on discrimination ($F < 1$, mean $A' = 0.84$ for both half and rare instructions). Other effects described earlier were replicated in the $A'$ analysis, including the effect of format, $F(2, 108) = 3.47, MSE = 0.02, p < .05$, where discrimination of targets from distractors in the single still condition (mean $A' = 0.86$) was better than in the moving condition (mean $A' = 0.84$) and multistill conditions (mean $A' = 0.81$).

In the bias ($B'$) analyses, there were additional highly significant effects of trial type and its interaction with target familiarity. This effect arose because the familiar group was relatively cautious when the trials showed a distractor on video with a nontarget (i.e., both items were unfamiliar, $B' = -0.21$) but relatively cautious on trials when a video showed a distractor but the photograph showed a familiar target ($B' = 0.14$). The unfamiliar group showed neutral bias ($B'$ of approximately 0) across all three types of mismatch trial. Thus, familiarity with the photograph (clear image) did not appear to bias participants to make false matches with the video (poor image).

**Confidence**

The final analyses conducted were on the participants' confidence scores, which were analyzed separately for correct target match and correct mismatched trials. An ANOVA of the target match trials revealed a significant main effect of target familiarity, $F(1, 108) = 119.0, MSE = 1.23, p < .01$. Participants familiar with the targets were significantly more confident ($M = 8.67, SD = 1.17$) than those who were unfamiliar ($M = 6.46, SD = 1.06$). There was also a significant main effect of instructions, $F(1, 108) = 5.16, MSE = 1.23, p < .05$, with the half instructions ($M = 7.79, SD = 1.37$) yielding significantly higher confidence compared with rare instructions ($M = 7.33, SD = 1.73$).

For correct mismatched trials, lower scores (below 5) signal greater confidence. The ANOVA again revealed a significant main effect of target familiarity, $F(1, 108) = 291.9, MSE = 1.13, p < .01$; for familiar participants, $M = 1.98, SD = 0.89$; for unfamiliar participants, $M = 3.90, SD = 0.92$. There was also a significant

---

**Figure 2.** Overall correct response rates in each condition of Experiment 1. Hatched bars represent familiar targets; gray bars represent unfamiliar targets. Targ = target; Nontarg = nontarget; Dist = distractor.
main effect of trial type, $F(2, 216) = 69.89$, $MSE = 0.26$, $p < .01$, with confidence lowest for the target–nontarget pairs (in line with accuracy analyses, distractor–target $M = 2.54$, $SD = 1.23$; distractor–nontarget $M = 2.95$, $SD = 0.90$; and target–nontarget $M = 3.32$, $SD = 1.61$). There was a significant main effect of instruction type, $F(1, 108) = 11.00$, $MSE = 1.13$, $p < .05$, with greater confidence being given with the “rare” instructions ($M = 2.75$, $SD = 1.30$) compared with the “half” instructions ($M = 3.12$, $SD = 1.31$).

The analysis of mismatched trials also revealed a significant interaction between target familiarity and image format, $F(2, 108) = 3.77$, $MSE = 1.13$, $p < .05$. Further analysis revealed that this was because format had a significant effect for unfamiliar participants, who showed greater confidence ($p < .01$, Newman–Keuls) in the still condition ($M = 3.64$, $SD = 0.94$), compared with the multistatic condition ($M = 4.24$, $SD = 0.79$), with the moving condition nonsignificantly different from either still condition ($M = 3.80$, $SD = 0.94$). There was also a significant three-way interaction in the confidence analysis of mismatched trials between target familiarity, instructions, and the type of mismatched trial, $F(2, 216) = 4.04$, $MSE = 0.26$, $p < .05$. A breakdown analysis revealed that this was due to the participants who were familiar with the targets being more confident ($p < .05$) when given rare instructions and when attempting distractor–nontarget trials.

In designing this study, we had expected our two groups of participants to perform equally poorly on the distractor–nontarget trials, where all faces should have been unfamiliar to both groups. Although this condition gave rise to significantly poorer matching by the familiar participants than the other conditions did, they still outperformed participants in the unfamiliar group by 12%. It became clear during the study that a small number of faculty participants in the familiar group might have been familiar with some of the nontargets (staff at another university) and also that three of the distractor people used could have been previously encountered by some participants in the familiar group. We therefore repeated all our statistical analyses using data from just half of the participants (30 familiar, 30 unfamiliar—half of the initial sample), excluding all those from the familiar group who could have known any faces other than those of the 12 targets and excluding data also from three distractor items where there was a possibility that undergraduate students might have encountered them in Glasgow. This made no difference to the statistical pattern of data at all and made very little difference to absolute levels of performance (no cell means deviated by more than 2.5% from the figures given in Table 1). In particular, there remained the same significant advantage for the familiar over unfamiliar participants in the distractor–nontarget condition, where their performance should have been closely matched. Thus this difference does not appear to be due to any accidental familiarity effects creeping in to affect performance in this cell.

There are several possible reasons for the difference, however. First, the high accuracy of the familiar participants in the other conditions may have led to them being better able to calibrate their performance and guess “sensibly” in this condition. The bias effects support this argument, as the familiar group was relatively incautious in this condition. Also, familiar participants may have been used to respond that two faces did not match (only familiar items were used in match trials)—yielding higher accuracy but also yielding the slight shift in bias in this condition. It is also possible that exposure to the unfamiliar items in trials where their differences from familiar faces were easy to signal made participants more attentive to distinguishing features of the distractor and nontarget faces, which became useful when these were paired together in later trials.

The important aspect of the results, however, is that for unfamiliar participants, the distractor–nontarget trials were of equal difficulty to the distractor–target trials. In contrast, familiar participants were significantly better at the distractor–target trials, compared with the distractor–nontarget ones.

### Discussion

Familiarity with target faces has a substantial effect on this task. When participants knew either the video target or the comparison photograph, they were able to match or reject matches with over 90% accuracy (mean $A' = 0.93$, where maximum score is 1) despite the poor quality of the video images used here. When participants were unfamiliar with these images, overall performance was very much poorer (mean $A' = 0.75$).

Familiarity with the targets aided discriminability, and its only effects on bias were in the cautious direction when one of the images was familiar. In general, the effects of deliberate biasing of instructions were as predicted for both types of item. When participants expected matches, then they made more correct match responses, but at the cost of more false ones. Although this behavior is sensible, it is also worrying, as there is considerable room for expectation to affect matching in police and court room situations. However, this effect of biasing did not interact with familiarity. Indeed, if anything, it appears that the familiar participants were appropriately cautious in the situation that we were most concerned about (false positives when the comparison image was familiar).

Again, we found no difference between moving and still images for the matching of unfamiliar faces in this study, replicating previous findings (Bruce et al., 1999; Christie & Bruce, 1998; but see Pike et al., 1997). However, the absence of any advantage of viewing the moving images for familiar people is more surprising. This contrasts with previous data (Knight & Johnston, 1997; Lander et al., 1999) showing that the identification of famous faces in difficult-to-recognize situations is helped if the images are shown moving rather than still.

One reason why we did not find this effect here may be that performance with the familiar faces on CCTV is too good—performance is effectively at ceiling, so seeing the people move cannot improve things further. A more interesting explanation is that the CCTV images show the wrong kind of movement—it may be expressive and other nonrigid movements of faces that aid in identification, rather than the gait-related movements of the head and body which characterize people walking toward a CCTV camera. Burton et al. (1999) found that recognition of individuals from their moving bodies alone (faces concealed) was very poor indeed, whereas recognition of individuals from the CCTV images of their faces alone was extremely accurate, suggesting that the patterns of movement that arise as a person walks toward a camera are not very helpful for identification in this task.
The remaining experiments in this article follow up the initial demonstration of the advantage of target familiarity for the video-matching task. In Experiments 2 and 3, we investigated whether brief periods of familiarization with targets were sufficient to yield benefits in such tasks.

Experiment 2

Experiment 1 confirmed the suggestion by Burton et al. (1999) that high levels of familiarity allow difficult images of faces to be identified accurately. However, we do not know how familiar a face must be before such benefits are observed. Although our undergraduate students in Experiment 1 were probably not lifelong friends of our targets, they had nonetheless been exposed to them for extensive periods of time on repeated occasions. What if the people had been encountered much more casually? This is an important question, as a police officer might claim to recognize someone on a CCTV image who was known from a different criminal conviction, but the officer's level of familiarity with the face might nonetheless be based on relatively brief prior exposure. Can we achieve the effects of familiarization by showing an animated image of the person to be encountered on poor CCTV later? In this experiment, we investigated this question. Participants were given a face verification task, rather like that used in Experiment 1, in which they had to decide whether a low-quality video still, shown in a 30° angle with some image degradation, matched a comparison full-face high-quality photograph. Half the people shown in the video stills had been studied earlier in animated, high-quality videos under instructions emphasizing that these faces would require identification later. The remaining items were completely unfamiliar in the test phase. The earlier exposure to the animated video could have allowed the extraction of some kind of 3-D or structural representation of that face, which might have helped compensate for the change in viewpoint for the familiarized targets.

Method

Participants

These were 24 undergraduates at the University of Stirling who volunteered to participate.

Materials

The materials for this experiment were drawn from the pool of items described by Bruce et al. (1999). For the familiarization phase, 24 video clips of police trainees were used in high-quality original format. Each showed a trainee seated in a rotating chair against a clear blue background. The person was filmed full face and then rotated 360° in 10° intervals. He then looked up, looked down, faced the camera, and smiled. For each actor, the video was edited to exclude the portion from 90° to 270° where only the back of the head was visible, to leave a clip of approximately 30 s in length for each. These 24 were divided into two sets so that each face served as familiarized or novel between subsets of participants.

For each of the 24 targets, an image of the person at 30° was captured from the video, and this was reduced in quality by increasing its color saturation and background in Adobe Photoshop by 60% (see Figure 3). Each image was printed out (approximately 8 × 10 cm) on a sheet of white A4 paper.

Figure 3. Examples from the test phase of Experiment 2. A target image appears at the top. At the bottom are shown a photograph of the target (left) and the distractor (right) used for match and mis-match trials with this target.

For the test phase, high-quality full-face studio shots of these 24 men were available, and for each of the 24, there was a paired distractor photograph chosen from faces with which this target has been confused in earlier matching studies (see Bruce et al., 1999, for a description of these studies).

Test booklets were constructed in which each pair of pages showed one 30° video image alongside a comparison photograph of the same or a different person. All 24 target video faces were tested in a booklet, half appearing with a distractor and half appearing with a photograph of themselves. Different booklets were used between participants so that targets were tested equally often in match and nonmatch trials.

Design

This was a mixed design with a within-participant factor of familiarization (half the target faces at test had been familiarized on video, the remainder were novel) and a between-participants factor of length of familiarization videos. Half of the participants saw each face for 30 s on video prior to test; the rest saw each video twice, for a total of 1-min exposure to the faces. All items served as both familiarized and novel faces by counterbalancing which items were shown on the videos between subsets of 6 participants.

In the test phase, low-quality 30° video images of each of the familiarized and novel items were paired with a photograph of either the same person or a distractor.

Procedure

Participants were tested individually. Prior to the familiarization phase, they were told that they should study all the faces appearing on video very
carefully because at a later stage they would be asked to recognize the faces appearing. They were then shown the videos of 12 men, with each clip shown either once (30 s) or twice (1 min), depending on participant group. A 2-s blank interval separated each of the video clips on the tape. In the test phase, participants were given 24 pairs of video-photographs and told that the two matched on 50% of trials. They were asked to decide whether they thought each pair showed the same person or two different people. Having made this decision, they were then asked to give their confidence on a 10-point scale on which 1 = a guess and 10 = absolute certainty.

Results

The mean performance in each condition of interest is summarized in Table 2. Performance is comfortably above chance (50%) but below ceiling, leaving scope for improvement as a result of familiarization. An ANOVA performed on the overall correct responses (maximum of six in each cell) in each condition revealed no significant effects (all ps > .15). The effect of prior familiarization was not significant, $F(1, 22) = 2.07, MSE = 1.29, p = .164$, though clearly there is some trend in the overall means in the expected direction.

Each participant's performance was also expressed as discriminability ($A'$) and bias ($B$) by combining the number of correct match responses (hits) with the number of incorrect mismatch responses (false alarms) for familiarized and unfamiliar targets. ANOVAs of the resulting $A'$ and $B$ scores showed no significant effects (all ps > .2) except that the two-way interaction between exposure (1 min/30 s) and familiarity in the bias ($B$) analysis approached significance, $F(1, 22) = 2.99, MSE = 0.007, p = .098$. The pattern of this interaction is that the 1-min group was more cautious with the familiar than with the unfamiliar target faces, whereas the 30-s group was more liberal with the familiar faces. Although there is some suggestion that the familiarization may have had slight, although inconsistent, effects on response criteria, there is no hint of the effect of familiarity on discriminability, which is where it had its effects in Experiment 1. Analysis of confidence of response showed no significant effects (all ps > .2).

Discussion

In Experiment 2, we examined whether a brief period of familiarization through exposure to videos of the faces to be matched later in the experiment was sufficient to create the highly accurate performance seen with familiar faces in Experiment 1. There was no clear advantage for the familiarized items, and those trends that appeared seemed less evident for the longer exposure group. The matching task used in this experiment required that participants compensate for a change in viewpoint from the 30° degraded video targets to the full-face clear photographs, and so study of these men rotating in front of the camera to reveal their faces at different angles might have been predicted to be especially helpful.

Nonetheless, the lack of familiarization effect is consistent with the failure to find benefits of using moving footage in earlier experiments (Bruce et al., 1999; Henderson, 1999). If study of a moving video image of a target could readily create the kind of visual representation enjoyed by familiar faces, then viewing targets in moving rather than static video for comparison should reveal benefits that our studies have failed to do. Experiment 2 shows additionally that there is little to be gained from the memory of a moving sequence showing the face to be compared. Thus, whatever the basis of the familiarity effects in such video tasks, it must rely on more than mere brief exposure to the items.

Experiment 3

Although Experiment 2 suggested that periods of exposure as long as 1 min did not serve to create the benefits for familiar items seen in other studies, it is possible that the familiarization period used was too brief or that participants were not encoding the items in a way that helped them in the later task. When we meet people in everyday life, we do not usually just observe them; more often, we interact with them in meaningful ways. In our final experiment, we investigated the effects of familiarization, with variation in encoding activity, by using the "arrays" task methodology that we have previously reported with unfamiliar faces (Bruce et al., 1999). Essentially, the task was to decide whether a video target appeared in a lineup ("array") of faces, with the target appearing in the array on 50% of the trials. We assessed three groups of participants on a baseline arrays task to determine that they were comparable, using materials not needed for the experimental phase. The three groups were then all shown a set of videos in a familiarization phase prior to a further arrays task that contained items that had been familiarized for some of the participants. The experimental social-exposure group viewed videos in pairs and were encouraged to chat about the people whose faces appeared. This was in an attempt to simulate something more like a social encounter with the faces to be familiarized. These faces then appeared as targets in the second arrays test. Performance of this group was compared with that of two further groups. One single-exposure group saw the same faces on video under similar instructions, but each participant viewed the videos in isolation—thus, this condition is analogous to the brief familiarization condition in Experiment 2. A control group viewed videos in pairs, but the videos showed faces of people who did not appear in the later arrays task, and thus any familiarity acquired with the video faces would be irrelevant to the later matching task, which used faces that were entirely novel for this group of participants.

Method

Participants

Sixty fellow students and/or acquaintances of the experimenter volunteered to participate in the experiment without payment.
Materials and Design

The design involved a between-groups comparison of matching performance. The three groups were given the same baseline test before proceeding to three different experimental conditions, depending on the nature of the familiarization phase.

Two sets of materials were used. For the baseline performance test, 20 target-face arrays were selected from those previously used by Bruce et al. (1999; Experiment 1). In brief, each of these arrays show a full-face target image grabbed from high-quality video alongside a set of 10 digital images of men of similar appearance to the target. In half of the arrays used the target was present in the array, and in half the target was absent. These arrays were used to check that the three groups of participants were matched in terms of their face-matching abilities at the start of the experiment.

For the familiarization phase and the test phase, a new set of materials was collected, as similar as possible to the items used in previous arrays tasks. The materials comprised videos and still photographs of police trainees at a local college. A total of 80 men were selected from this database. Twenty formed the targets for the second arrays test, and 40 were nontargets used to create the arrays for this test. The final set of 20 were faces used in video for the control group only, who were familiarized with irrelevant faces.

The 40 video clips (20 of the targets, 20 for the control group familiarization) were edited so that each showed a repeated 7-s clip of the face as it made expressive and speech movements in angles close to full face, looped four times to create a clip of approximately 28 s in length. The 60 still images (20 of the targets, 40 nontargets for the array test) showing full-face neutral expression (including stills of the 20 targets) were rated for similarity using the sorting method described by Bruce et al. (1999). Participants were asked to sort 60 images into piles, and the number of times each face was grouped with each other face was counted. This enabled us to select a set of the eight most similar nontarget faces for each of the 20 target faces, and these faces were used to construct arrays for each target face. Because of the relatively small pool of items (60 in total), arrays for the test phase were constructed using just 8 items (either 7 plus a target or 8 nontargets) to avoid individual items being repeated too often. Half the arrays were constructed with the target present and half with the target absent. The 8 items in each array were arranged in two rows of 4 items per row and were shown alongside an image of a target face (9 × 12 cm) that had been grabbed from the videotape to match the full-face unsmiling viewpoint as closely as possible.

In the arrays used in the test phase, faces measured approximately 6 × 8 cm and were printed in color on 720-dpi paper, using an Epson Stylus printer. Two arrays were created for each person videotaped (one present, one absent), with the position of the target face varied across arrays. Two test books were created, each containing 20 target faces plus 20 arrays (10 target faces present, 10 absent). Half the participants saw one set of targets in present arrays and the rest in absent arrays, and the others saw the remaining targets with present arrays.

For both the baseline and test phases, response booklets were created so that participants could work at their own pace through the set of arrays. This enabled participants to be tested in pairs with different test books. Each page of the booklet showed an array of numbered faces in silhouette, and participants were instructed to either mark the face they thought matched the target in the array on the “present” option and then look for the target in the “not present” box as appropriate for each numbered array that they examined. Where participants had been recruited in pairs, they were seated so that they were not able to see each other’s work in the baseline and test phases. In the familiarization phase, participants were asked to watch the video carefully, as it would be relevant to their task in the final phase of the experiment. Those recruited in pairs were asked to “chat about the faces between yourselves as you watch the video.” They were assured that their words were not being recorded. Then all participants were given the 20 arrays of the test phase and were asked to work through these in a similar way to the baseline task. The whole procedure took approximately 30 min.

Results

Table 3 summarizes the percent correct performance on target-present and target-absent trials for each group at baseline and in the test phase.

Baseline Performance

The three groups were fairly closely matched in their baseline accuracies as indicated in Table 3. A 3 (participant group) × 2 (target-present vs. target-absent arrays) ANOVA on correct scores (maximum of 10) showed only a main effect of array type, F(1, 57) = 39.32, MSE = 3.10, p < .001; performance on target...
Postfamiliarization Performance

Table 3 indicates that the three groups differed considerably, and an ANOVA showed a main effect of participant group, $F(2, 57) = 14.43$, $MSE = 3.71$, $p < .001$; and of array type, $F(1, 57) = 114.783$, $MSE = 3.48$, $p < .001$; but no interaction between these, $F(2, 57) = 1.055$, $MSE = 3.48$, $p = .355$. Performance on present arrays (89%) was again more accurate than on absent arrays (52.5%), but this was shown across all three groups. Scheffé tests showed that the experimental social-exposure group performed significantly better than both the other two groups ($p < .01$). Using a Scheffé test, we found that the single-exposure group was not significantly better than the control group, although a less conservative post hoc $t$ test showed $t(57) = 2.08$, significant at $p < .05$.

The clear advantage of the social-exposure group is even more evident when the individual participant data are examined. Seventeen of 20 participants in the social-exposure condition scored 100% correct on target present arrays, compared with only 4 of the single-exposure participants and just 1 in the control group.

Discussion

In Experiment 3, we used high-quality video images and an array task to extend the investigation of effects of brief familiarization reported in Experiment 2. The single-exposure condition of Experiment 3 is analogous to the familiarization used in Experiment 2 and gave similar results—there was a marginal improvement for familiarized compared with unfamiliarized faces (single-exposure vs. control condition gave a marginally significant 9% advantage in a between-groups comparison at test in Experiment 3 compared with the nonsignificant trend of 6% overall in a within-subjects contrast of familiar and unfamiliar items in Experiment 2). However, Experiment 3 added the quite striking observation that the same duration of physical exposure to faces can have a dramatic benefit in some circumstances. The social-exposure condition of Experiment 3 yielded a substantial gain in performance, particularly on target-present trials in which performance became almost perfect. We do not know what the critical aspects of the social-exposure condition were that yielded the benefits for these participants' later matching performance. It may have been that this condition provoked deeper processing of the faces in terms of semantic and personality characteristics. We know that focusing on personality dimensions promotes better memory for faces than focusing on their physical characteristics, and it may be only the social condition of this experiment that encouraged such deep processing (cf. Bower & Karlin, 1974; Patterson & Baddeley, 1977; Winograd, 1981). It may have been that talking about the faces forced participants to create labels or identities for individual faces, or it may have been something more genuinely related to the social context within which the faces were viewed. Further research is needed in order to understand the basis of the learning effects observed in Experiment 3. For now, however, Experiment 3 suggests that it is likely to be the nature, rather than the length, of exposure to faces that will lead to the benefits of familiarization.

General Discussion

Experiment 1 showed that personal familiarity with the people shown in images even of low quality led to highly accurate matching. This was not at the expense of increased false matches in cases where the video face was not familiar but the comparison photograph was. Familiar items were both matched more accurately and rejected more correctly in Experiment 1—the effects of familiarity operated wholly through discriminability and not bias. Presumably, performance is good with familiar faces in this task because, for most images, there was sufficient low-spatial frequency information in the CCTV image either to match with the stored visual appearance of the person in memory (for the target images, cf. Harmon & Julesz, 1973) or to decide that the image was unfamiliar (for the distractor images). If the person on the CCTV image can be matched to the visual memory of "Mike
Burton," then this obviates any need to compare the image directly with a comparison photograph. The match or mismatch can be decided by individual identification of each separate image and comparisons mediated by the nonvisual identities accessed. When Burton et al. (1999) showed highly accurate performance in recognition memory for familiar faces shown on CCTV, it seemed likely that it was these nonvisual mediating codes that allowed for accurate memory decisions (cf. Bruce, 1982). Ironically, then, it is where performance with CCTV images can interact with individual stored memories for (familiar) faces that it is likely to be most reliable, even though the apparent advantage of CCTV cameras is that they remove the need to rely on witness memory.

Bias effects were demonstrated in Experiment 1 through manipulating the expectations that participants had of the frequency of matches. This shows that matching in these tasks is susceptible to influences of bias, but such effects appear to operate independently of the influence of familiarity.

However, we are aware that the distractors shown on video in Experiment 1 were not closely matched in visual appearance with the familiar targets. We still need to establish the effects of familiarity and bias on decisions made in response to poor-quality video images chosen to resemble closely people familiar to participants. Such a study is difficult to set up, but it is important nonetheless. Similarly, our faces and participants were chosen to represent relatively high levels of familiarity with faces in some conditions. Nonetheless, the results of Experiment 1 strongly suggest that where a person is recognized on a CCTV image by someone familiar with them, these identifications should be taken very seriously, even if the CCTV image is of low quality. Interestingly, it is in such situations that CCTV has had some of its most important successes. For example, the London nail bomber was identified and subsequently apprehended and convicted after a close personal acquaintance identified the man from CCTV images that were shown in the newspaper.5

This led us to investigate the effects of familiarization in Experiments 2 and 3. Could the benefits of familiarity be created through a period of prior exposure to the faces to be matched later? What happens when a video shows an image of a person known only slightly—as may be true of police familiarity with some local people with previous convictions? Experiments 2 and 3 suggest that brief periods of simple exposure will not necessarily yield the benefits of familiarity but that when encoding conditions more closely simulate a social situation, then quite dramatic effects of familiarization can be obtained. The precise mechanism by which the social-exposure condition of Experiment 3 generated its benefits awaits clarification, but for now, the practical implication is that one must take seriously reported identifications of people seen on video, even by others who know them slightly. However, the benefits found in Experiment 3 occurred in a task in which high-quality images were available for inspection at test. As yet, there is no evidence that brief periods of familiarization can assist the recognition or matching of unfamiliar faces in poor-quality images.

The findings reported here add an important gloss to cautions that should be applied when using CCTV images to establish identities in police work and in court. CCTV images are extremely useful in establishing what happened at a crime and what the person or people who committed the crime looked like. Our research (Burton et al., 1999; Experiment 1 of this article) also has shown that even very poor images can be readily identified by people who are familiar with the faces shown. This is an important demonstration, because cases often rely on such identifications. (In a recent, high-profile case in the United Kingdom, well-known football players were accused of an assault, and CCTV images showing people in the vicinity of the criminal incident were referred to in press coverage of the court case).6

Used appropriately to trigger recognition among the public or police, CCTV can be an extremely powerful aid to identification and therefore is appropriate as an aid to an investigation. Even where CCTV images are of low quality, the use of broadcast or print images to provoke identification in public and police contexts should be encouraged. Indeed, our experiments suggest that there is little advantage for presenting moving rather than still images to provoke successful recognition, suggesting that circulation of still images from CCTV by means of newspaper or other media may be just as effective as showing video footage on television programs.

Our concerns are that CCTV images are often not used in this way but are used as evidence of identity in court, where people unfamiliar with the alleged offenders are invited to compare the appearance of a CCTV image (often low in quality) with that of the defendant. On the basis of research presented here and earlier (Bruce et al., 1999; see also Davies & Thasen, 2000), we argue that this latter practice should be avoided, as resemblances between images of otherwise unfamiliar faces can be misleading and such judgments are highly prone to error.

References


Received May 3, 1999
Revision received May 10, 2001
Accepted May 10, 2001