

Running head: Live showups and video lineups

Live showups and their influence on a subsequent video lineup

Tim Valentine (Goldsmiths, University of London)

Josh P Davis (University of Greenwich)

Amina Memon (Royal Holloway, University of London)

&

Andrew Roberts (University of Warwick)

Corresponding author: Professor Tim Valentine
Department of Psychology,
Goldsmiths, University of London,
New Cross,
London
SE14 6NW
Email: t.valentine@gold.ac.uk

Acknowledgments

We wish to thank the Nuffield Foundation for funding this research project (Grant no. AJU33483). We also thank the members of the project advisory group, John Armstrong, Paul Bogan, Graham Davies, Mark Lindley and Ade Adetosoye for their valuable contribution to the project.

Abstract

A live showup (known as a street identification in the UK) allows the perpetrator to be identified shortly after a street crime. If the suspect disputes the identification, a video lineup often ensues. Four experiments examined the reliability of live showups and their influence on a subsequent video lineup using realistic procedures and conditions. Similar proportions of culprits and innocent suspects were identified from live showups and video lineups. Both culprits and innocent suspects previously identified were likely to be identified again in a subsequent lineup, with delays from a few minutes to a month. Only a weak effect of clothing bias was observed. There was strong evidence of commitment to a previous identification, but no reliable evidence of source monitoring errors. The results suggest that a live showup is not less fair than a lineup, but use of repeated identification procedures introduces an unfair bias against innocent suspects.

Introduction

Eyewitness testimony can play a crucial role in securing a conviction in court. However, it has become increasingly apparent that eyewitness memory can be prone to serious error. A large body of empirical research has also shown that an unreliable identification can be made with high levels of confidence (for reviews see Dysart & Lindsay, 2007a; Valentine, 2006; Wells, Memon & Penrod, 2006; Wells, Small, Penrod, Malpass, Fulero, & Brinacombe, 1998). In England and Wales, cases of mistaken identification resulting in miscarriages of justice led to the publication of the Devlin Report (1976) and with it a number of recommendations for improving police practice. DNA exoneration cases in the USA have also demonstrated that eyewitness error has resulted in a substantial number of wrongful convictions (e.g., Scheck, Neufeld & Dwyer, 2000; Innocence Project, 2011). Rates of wrongful conviction are likely to be far higher in cases that have a less severe judicial penalty (Naughton, 2003).

In England and Wales, the Police and Criminal Evidence Act (PACE, 1984; Codes of Practice, 2008) established guidelines for the conduct of identification procedures. The Code permits use of a live showup (known as a street identification), normally held soon after a crime and if practical, after obtaining a description of the culprit. (See Davis, Valentine, Memon & Roberts (submitted) and Bogan (2004) for an account of current practice in England and Wales.) The procedure often involves a victim being taken on a drive around the area in order to obtain identification evidence to justify arrest of a suspect. If there is sufficient evidence to justify arrest, a live showup is prohibited in England and Wales. Usually, after arrest a video lineup should be conducted (*R v Forbes*, 2001). A showup *must* be conducted with the suspect physically present (i.e. it must be a live showup). Showing a single photograph of a suspect to a witness is not permitted in England and Wales, although it is a common procedure in other jurisdictions including the USA.

A live showup conducted shortly after a crime can be the only practical means to investigate a crime, and may appear to have a number of advantages. Witnesses are likely to be most motivated to assist immediately after an incident; their memory will be fresh and innocent suspects can be quickly eliminated from further investigation. For these reasons, showups are frequently used. In the USA, showups have been estimated to account for between 30% and 77% of all identification procedures (Gonzalez, Ellsworth & Pembroke, 1993; Stebley Dysart, Fulero, & Lindsay, 2003; Dysart & Lindsay, 2007a). In England and Wales, over 20,000 suspects per annum are estimated to be charged or cautioned after being identified from a showup (Davis, *et al.*, submitted). However, it has been argued that a live showup is an inherently suggestive procedure, and that a lineup should be used to provide a safeguard against the risk of a mistaken identification (Dysart & Lindsay, 2007a; Steblay *et al.*, 2003; Wolchover & Heaton-Armstrong, 2004).

A meta-analysis of eight US empirical studies identified 12 comparisons of showups and lineups of which 11 involved photograph showups and photograph lineups (Steblay *et al.*, 2003). In culprit-present trials, an identification (of the culprit or a foil) was made to 71% of six-person lineups. However, because 24% of witnesses identified a foil, only 47% of culprits were correctly identified from lineups. This figure closely compares with 46% of culprits identified from showups. In target-absent trials, 57% of lineups and 85% of showups were correctly rejected. However, taking account of mistaken identifications of foils from lineups, the proportion of innocent suspects mistakenly identified was 10% for lineups¹. This figure compares with 15% of innocent suspects mistakenly identified from showups.

In both the UK and the USA, witnesses may be asked to attempt to identify the same suspect a second time in a lineup (Behrman & Davey, 2001, Davis, *et al.*, submitted). Following a disputed identification from a live showup, there is an obligation in England and

¹ The figure of 10% innocent suspects mistakenly identified is from a subset of 5 tests that allowed the separation of foils and innocent suspects.

Wales that a lineup must be carried out (*R v Forbes*, 2001). A question arises as to the probative value of a second identification procedure. It could be argued that any identification from a subsequent lineup could be influenced by the first identification. The same issue arises if a witness identifies a photograph from a mugshot album and subsequently views the suspect in a lineup.

Carry-over effects can influence an identification procedure in a number of ways. A witness who has made a prior identification may show a *commitment* to identify the same person again, even when the identification is mistaken. There is a strong evidence for a commitment to a mistaken identification, from a mugshot search (Blunt & McAllister, 2009; Brown, Deffenbacher & Sturgill, 1977; Dysart, Lindsay, Hammond, & Dupuis, 2001; Gorenstein & Ellsworth, 1980; Goodsell, Neuschatz & Gronlund, 2009; Haw, Dickinson & Meissner, 2007; Hinz & Pezdek, 2001; Memon, Hope, Bartlett & Bull, 2002), from a showup (Godfrey & Clark, 2010; Haw *et al.*, 2007), and from a prior lineup (Hinz & Pezdek, 2001; Pezdek & Blandon-Gitlin, 2005). Another aspect of commitment has been observed amongst participants who reject culprit-absent mugshots and were subsequently more likely to reject a culprit-absent lineup (Brigham & Cairns, 1988). In this case, the commitment is to *not* identify foils even if they had been seen previously. (For a review and meta-analysis see Deffenbacher, Bornstein and Penrod, 2006).

The commitment effect cannot explain a mistaken identification of an innocent suspect or foil in a subsequent lineup if they were seen but *not selected* in the previous identification task. Errors of this nature are commonly interpreted within the Source Monitoring Framework (e.g., Johnson, Hashtroudi & Lindsay, 1993) in which an eyewitness may correctly recognise a person as someone they have seen earlier, but incorrectly associate the innocent person with the crime due to a failure to recollect the context in which the face was seen. Transference of familiarity from a face seen in a prior identification task leads to

source confusion (Deffenbacher *et al.*, 2006, Earles, Kesten, Curtayne & Perle, 2008; Memon *et al.*, 2002). A similar unconscious transference of familiarity can occur from an ‘innocent bystander’ present at the time of the incident (e.g., Deffenbacher *et al.*, 2006; Ross, Marsil, Benton, Hoffman, Warren, Lindsay & Metzger, 2006).

Blunt and McAllister (2009) and Goodsell *et al.* (2009) challenged the evidence for source confusion errors from repeated identification procedures, and argued that most eyewitness errors that result from a prior identification procedure are attributable to an effect of commitment. Referring to a culprit-absent mugshot album, Goodsell *et al.* summarise their results as follows: “Mugshot choosers will select their prior mugshot choice if given the opportunity and will reject a lineup that does not contain it [...] even when the opportunity to select the actual culprit is available” (p.798) and “mugshot non-choosers are likely to reject a subsequent lineup after failing to select from a mugshot search” (p.799).

The current research

The majority of research on showups has been conducted in the USA using procedures that differ from those used in the UK, and the majority of research on carry-over effects from a prior identification procedure has been based on mugshot albums. The aim of the current research was to compare the reliability of live showups and video lineups that follow the legal codes of practice in the UK. A further aim was to measure the influence of a live showup on the outcome of a subsequent video lineup. Experiments were designed to maximise ecological validity by use of a live staged incident, and live showups, with the video lineups constructed by police using the same procedure and database of foils used for criminal investigations in the UK. Generalisation of the results for different culprits and suspects, and across different delays was addressed by recruiting different actors for each experiment, and varying the delay between the repeated identification tests (Experiment 3). In each case, an innocent suspect who was as similar as possible in appearance to the culprit was

selected from volunteer actors. The similarity achieved differed substantially across experiments. However, this reflects the situation in real cases in which suspects are apprehended on the basis of their appearance. Reliable effects of relevance to the criminal justice system should generalise across different levels of similarity of an innocent suspect to the actual culprit.

Experiment 1 compared the sensitivity, in obtaining identifications of the culprit, and fairness, in minimising mistaken identifications of an innocent suspect, of live showups and video lineups. The experiment also examined the effect of a live or video showup on the outcome of a video lineup conducted 15 minutes later. Experiment 2 involved a further comparison of a live and video showup with a video lineup. The effect of a live showup on a video lineup held after a delay of 1-6 days or 9-30 days was explored in Experiment 3. The final experiment explored the effect of clothing bias on transference of familiarity from a culprit to an innocent bystander present at the scene.

An additional aim of the research was to evaluate the effects of the nature of identification procedures on witness confidence. Ratings were collected a short time after the staged scenario, and immediately after all identification decisions. Previous research has shown that confidence measured immediately after an identification procedure has a stronger relationship with accuracy than confidence measured immediately after the incident (Cutler & Penrod, 1988). In a meta-analysis of 30 studies, Sporer, Penrod, Read and Cutler (1995) found a stronger accuracy-confidence relationship in those who make an identification from a lineup (*choosers*) than those who reject the lineup (*non-choosers*).

Experiment 1

The first aim of Experiment 1 was to compare the reliability of a live showup with a video lineup to address the question of whether a live showup is a less reliable procedure. Based on the research literature, it was predicted that more selections would be made to

lineups than to showups but, because some selections from lineups would be of foils, there would be no difference in rates of accurate identification in culprit-present trials. It was also predicted that there would be more mistaken identifications from showups than from lineups in culprit-absent trials.

The comparison between live showups and video lineups confounds a comparison between a showup and lineup procedure with a change of test media. The staged incident was live, so a live showup involves a test in the same presentation mode, whilst a video lineup involves a change of media. For this reason a third condition, a video showup, was included to enable the effects of test method and a change of media to be separated. The video showup consisted of presentation of a video clip of the 'suspect', filmed at the same time as the video lineup. According to the Encoding Specificity Principle (Tulving & Thomson, 1973) memory performance should be better if more cues available at learning are available at test. This leads to the prediction for the first identification test of better recognition of the culprit in a live showup than in a video showup or video lineup, because the video procedures use a head-and-shoulders view rather than the full-length view available from a live presentation. However, previous eyewitness identification research has shown that the effects of different test media are limited. Egan, Pittner and Goldstein (1977) found higher rates of correct culprit-present identifications in live lineups than in photographs. Cutler and Fisher (1990) found no differences in correct identifications when the culprit was in a live, video or photograph lineup, but more false identifications were made with photo lineups. In a review of experiments with adults, Cutler, Berman, Penrod and Fisher (1994) concluded that the type of procedure used has a limited influence on outcomes.

The second aim of Experiment 1 was to investigate the influence of a showup on the outcome of a subsequent video lineup. The outcomes of video lineups when the witnesses had seen the suspect in a prior showup were compared with the outcome of lineups viewed by

a control group who did not participate in a previous identification procedure. It was predicted that witnesses would show commitment to their previous decisions. Suspects who were identified previously from a showup would be more likely than suspects viewed by the control group to be identified in a video lineup, whether or not they were the culprit or an innocent suspect. In addition, applying the source monitoring framework leads to the prediction that suspects are more likely to be identified by participants who had seen them in a prior showup, but did not identify them, than by control participants. Recent research has called this effect into question but a further opportunity to test the prediction is provided by Experiment 1. The accuracy-confidence relationship was expected to be stronger amongst choosers than non-choosers. No difference was predicted in the confidence of responses to live and video showups.

Method

Design: Experiment 1 employed a 2 (*culprit presence: culprit-present, culprit-absent*) x 3 (*identification procedure: live showup, video showup, control*) between-subjects design. Participants viewed a surprise argument between an actress (*the culprit*) and the experimenter. A third of the participants attempted to identify the culprit in a first identification phase with either the culprit (*culprit-present*), or an innocent suspect (*culprit-absent*) present in person (*live showup*). A third of the participants attempted to identify the culprit from a single video clip (*video showup*) which was either of the culprit, or an innocent suspect. A third of the participants did not participate in a showup trial (*control*). All participants were presented with a *culprit-present* or *culprit-absent* nine-person *video lineup* approximately 30 minutes after the original incident. The suspect seen previously in a showup (if applicable) was always a member of the lineup. Participants who saw a culprit-present showup, saw a culprit-present video lineup. Participants who saw a culprit-absent

showup, saw a culprit-absent video lineup. All other lineup members were novel foils. The dependent variables were the outcome and rated confidence of the identification procedures.

Participants: Participants were staff and students at Goldsmiths, University of London recruited in groups (mean group size = 25.2, SD = 8.3, max = 47, min = 16). The data from six participants were excluded for being familiar with one of the actresses, for providing ratings of 0% confidence in being able to recognise her or for being pre-warned of the experimental rationale. Participants who expressed no ability to recognize the culprit shortly after the staged incident, possibly because they did not pay attention to the scenario, were excluded on the grounds that police would not conduct an identification procedure if a witness said they did not see the perpetrator. Valid data were obtained from 283 participants (54 male; 229 female; aged 18 - 67: $M = 23.1$). Two hundred and seventy provided their ethnicity and nationality; 64.4% were from the UK, 15.9% from other European Union (EU) countries, 8.5% from the Indian sub-continent, with the remainder from other regions. The majority (58.9%) were white European, the largest minority (11.1%) were Asian.

The two undergraduate actresses were recruited by poster advert and rated as the most similar in appearance from photos of the 30 applicants (Figure 1). Both were white European, 21 years of age, 5 foot 4 inches tall, of slim build, and a pale olive complexion with dark brown hair. The innocent suspect had long hair which was tied up for all identification trials. The culprit had short hair. To obscure this difference during the live incident the culprit wore a black scarf. She also wore dull-white trainers, blue jeans and a black coat. For the live and video showup trials, both actresses wore similar trainers and jeans and a stripped black and grey long sleeve t-shirt. In the video lineup they both wore a plain blue top.

Figure 1 about here

Materials: A series of questionnaires were prepared as follows.

Post-incident description questionnaires 1 and 2: Two questionnaires were designed for distribution prior to the identification trials. The following data were requested: participants' age, gender, ethnicity and nationality, whether they were familiar with the actress and a scale measuring confidence in being able to recognise her later (0 – 100%). The first questionnaire required a free recall description of the culprit. However, analyses of description data will be reported elsewhere.

The second questionnaire was based on a Metropolitan Police booklet for use at the scene of a crime. This comprised 27, mainly cued multiple-choice description questions. These data will be subject of a separate report. However, some of the qualitative data is referred to in the discussion of pertinent results. Additional questions evaluated the realism of the staged plagiarism scene by requesting ratings of the participant's suspiciousness that it was a staged event, and the convincingness of the scenario (from 0 – 100%). They also identified the point at which participants realised the scenario had been staged on an eight-point scale. Participants provided ratings of emotional impact, anxiety, confusion, anger, suspiciousness, surprise and arousal on a series of four-point scales.

Showup questionnaire: Participants taking part in the showups were provided with an identification questionnaire and instructions. The critical question being: "Is the person who has now entered the room (being shown on the screen), the same person you saw arguing with the experimenter? Please circle one response (yes or no). If you cannot make a positive response you should respond 'no'." A subsequent question measured identification confidence (from 0 – 100%). Two additional scales examined whether participants felt under pressure to make a selection (yes/no) and whether they remembered the warning that the culprit may or may not be present (yes/no).

Video lineup questionnaire: A similar questionnaire was also created for the video lineup phase. If participants made a positive identification decision, a lineup number (1 - 9) response was required.

Video lineups: Separate video lineups were created for each suspect. These were constructed by the Metropolitan Police using the PROMAT™ system². Foils were selected by a police officer who was experienced in compiling video lineups, by entering keywords that matched the suspect's description (e.g., gender, age, ethnicity and hairstyle). A list of still images that matched the description was displayed, from the PROMAT™ database of approximately 23,000. The operator then selected those believed to be of similar appearance to the suspect. Three of the eight foils were selected for both culprit-present and culprit-absent lineups. All lineups consisted of 15-sec colour head-and-shoulders video clips of the suspect and eight distracters. In each clip the lineup member faces the camera, turns to a left profile view, rotates back to a right profile view and then returns to a full-face view. On playback, a lineup member number appears on the screen.

Procedure: All rooms had good lighting and participants were seated between 1.5 and 4m from the front with a clear view of the actor. Participants were always unaware of the true purpose of the experiment prior to the staged incident. The experimenter handed out a personality questionnaire to complete. (These were later destroyed.) He then pretended to have technical problems with a laptop. On a hidden signal, a 45-sec staged act was performed in which an actress entered the room carrying an essay with prominent red markings on it. Increasingly loudly, she criticized the experimenter for accusing her of plagiarism. The actress exclaimed that she was not a "cheat". Eventually she was directed to leave the room to discuss the issue in private. Both the actress and the experimenter left the room and after a brief interlude the experimenter returned and announced that the true purpose of the study

² There are two software systems used to produce video lineups in the UK. PROMAT™ and VIPER™. Each system uses its own national database of foils. The video lineups produced by both systems are of the same style and format. There is no functional difference between the systems for research use.

was to test event memory. Participants were provided with informed consent forms, asked to treat the procedure as if they had been witnesses to a crime, not to discuss what they had seen and to ensure no one could view their responses. They were then given approximately 5 minutes to complete the first post-incident questionnaire and then once collected, another 10 minutes to complete the second, giving a total delay of about 15 minutes.

The participants taking part in the showups were then given brief details of a typical scenario in which, the police may have asked an individual, who matches the witness' description to remain in the vicinity. They were then provided with the showup questionnaire and told that an actress would enter the room in the live showup condition (or that her image would be displayed on a screen in the video showup condition). They were warned that the 'suspect' actress 'may or may not' be the culprit they had seen earlier and that if they could not recognise her, they should make a negative response. For the live showup, either the culprit or the innocent suspect entered the room and for approximately 1-min stood in the same place as in the incident facing the centre of the room. For the video showup a single PROMAT™ lineup video of the actress was played on a loop on a large display screen.

Approximately 15 minutes later, all participants were provided with the video lineup questionnaire and given similar instructions as in the showups. All watched a PROMAT™ video lineup which included either the culprit or the innocent suspect. A video of each lineup member was shown sequentially in a pseudo-random order, with the constraint that the suspect was never first or last. A second single video clip was filmed of both actresses by the police for use as a video showup. Participants were asked to watch the entire video twice before making their response. This is a requirement of the PACE codes, followed by the police in England and Wales. The experimenter sat facing away from the screen so that he did not know which image was being shown, to avoid introducing any experimenter

influence, but could monitor compliance with instructions. Once the questionnaires had been collected, participants were thanked and debriefed³.

To test the fairness of the culprit-present lineup, a pooled description of the suspect was compiled from descriptions produced by 5 volunteers. Additional pilot participants ($n = 54$) were provided with this description and were simultaneously presented with still full face images of the lineup members projected onto a screen. All were asked to select a lineup member based on this description. If a lineup is completely fair all lineup members should be selected equally. In this case, the culprit was selected by 7.4% of participants, fewer than the 11.1% that would be expected by chance, suggesting that the culprit did not stand out in anyway. One lineup foil was selected by 29.6% of participants. The selection of lineup members was not altered in any way because it was compiled by the police using the procedure used in criminal cases. It represented the lineup that would have been used if this was a criminal case.

Results

Table 1 shows the outcome of showups and video lineups as a function of culprit-presence, condition (live showup, video showup, control) and whether the decision was participants' first or second decision. Data from video lineups are shown as a function of whether the participant made a selection (chooser) in a prior showup.

Table 1 about here

³ **Incident realism and participant suspiciousness:** Incident realism, participant suspiciousness and emotional impact scores measuring the effectiveness of the plagiarism scenario differed across experiments, probably due to acting ability, but not across conditions within any experiment. Therefore the data from 1,158 participants in Experiments 1 – 4 were analysed together (descriptive statistics are presented in Appendix 1). Participants failing to respond on any measure were treated as missing data on that measure only. The act was rated as highly believable. On a scale from 0% (not at all) to 100%, suspiciousness was low, with 64.3% of participants responding at the extreme of the scale (0%). The majority (54.3%) were also at least 90% convinced that the act was real and most (80.3%) only realized after the actor had left the room. A few (7.6%) claimed to have realised this while the actress was speaking. Analyses conducted on the six emotional impact four-point scales (4 = very much so, 1 = not at all) found that the relative ranking of scores on these scales varied slightly across experiments. However, in all experiments the highest mean scores were given to surprise - a further indication of the successfulness of the staged act. This was generally followed by suspiciousness or confusion, and then anxiety or arousal with anger always the lowest. These values highlight the low emotional impact of the staged scene.

In culprit-present showups 65.3% of participants correctly selected the culprit. In culprit-absent showups 5.3% of participants incorrectly selected the innocent suspect. Combining all data from the control and second identification decision trials, in culprit-present video lineups, 67.6% of participants correctly selected the culprit, 9.2% a foil, and 23.2% made an incorrect rejection. In culprit-absent lineups, 12.1% of participants selected the innocent suspect, 29.8% a foil and 58.1% made a correct rejection.

First identification decisions: A 3 (live showup, video showup, control [video lineup]) x 2 (chooser, non-chooser) chi-squared analysis compared the first identification decisions made by participants in showups with the controls whose only decision was made to a video lineup⁴. The proportion who made a selection in culprit-present trials differed significantly across conditions, $\chi^2(2, n = 142) = 13.94, p < .01, \Phi = .31$. Post-hoc (2x2) Fisher's exact tests showed that fewer selections were made to live showups (51.1%) than to video showups (79.2%) and control lineups (83.0%; $p < .01$). The latter conditions did not differ ($p > .2$). In these analyses, selections could be of the culprit or a foil (i.e. correct or incorrect). A 3 x 2 chi-squared analysis of the accuracy of response (accurate vs. inaccurate) as a function of first identification procedure (live showup, video showup, control [video lineup]) was significant, $\chi^2(2, n = 142) = 9.28, p < .01, \Phi = .25$. Fisher's exact tests showed that fewer correct identifications of the culprit were made in live showups compared to video showups and control lineups ($p < .05$). There were no differences between video showups and control lineups ($p > .2$).

A 3 (live showup, video showup, control [video lineup]) x 2 (chooser, non-chooser) chi-squared test comparing selections from culprit-absent trials was significant, $\chi^2(2, n = 141) = 29.87, p < .01, \Phi = .46$. Post-hoc Fisher's exact tests showed that more incorrect

⁴ The comparison was made between video and live showups, held about 15 minutes after the incident, with video lineups held about 30 minutes after the incident. This analysis confounds the comparison to a video lineup with a small difference in delay. This issue is addressed in the discussion and in Experiment 2.

identifications were made to video lineups than to both types of showup ($p < .01$), which did not differ ($p > .2$). An identification from a culprit-absent video lineup could be of the innocent suspect or a foil. A chi-squared test showed that the proportion of innocent suspect identifications vs. other decision did not differ across the three conditions (live showup, video showup, control [video lineup]), $\chi^2(2, n = 141) = .35, p > .2, \Phi = .05$.

First decision confidence: Table 2 displays the mean confidence levels in the first identification decision made to showups or video lineups as a function of culprit presence, outcome, and condition.

Table 2 about here

The culprit-present and culprit-absent data were combined. A 2 (accuracy of response) x 3 (live showup, video showup, control [video lineup]) ANOVA revealed a significant main effect of accuracy, $F(1, 274) = 40.33, p < .001, \eta^2 = .128$, correct responses were associated with higher confidence than incorrect responses ($M = 74.9, SD = 20.6$ vs. $M = 57.6, SD = 20.7$). The main effect of condition was also significant, $F(2, 274) = 3.76, p < .05, \eta^2 = .027$. Tukey's tests found that confidence was higher in live showups than in video showups and video lineups ($M = 76.8, SD = 20.6$ vs. $M = 68.6, SD = 20.2$ vs. $M = 65.9, SD = 23.6, p < .05$). The latter conditions did not differ ($p > .2$). The interaction between accuracy and condition was not significant, $F(2, 274) = 3.01, p > .05, \eta^2 = .021$.

A 2 (accuracy of response: accurate vs. inaccurate) x 2 (decision: chooser, non-chooser) ANOVA, revealed a significant main effect of accuracy, $F(1, 276) = 35.05, p < .001, \eta^2 = .113$, correct responses were associated with higher confidence than incorrect responses. The main effect of decision, $F(1, 276) < 1, \eta^2 = .001$ and the interaction were not significant, $F(1, 276) < 1, \eta^2 = .003$.

Culprit-present video lineups: The proportion of the three outcomes from culprit-present lineups (correct culprit selections, foil selections and incorrect rejections) did not

differ across the three conditions (live showup, video showup, control, see Table 1), $\chi^2(4, n = 142) = 5.82, p > .2, \Phi = .14$. Fewer participants who had taken part in a live showup correctly selected the culprit (55.3%) than those who had taken part in a video showup (75.0%) or controls (72.3%). However, this difference was not significant, $\chi^2(2, n = 142) = 4.92, p = .085, \Phi = .19$.

Further analyses examined the effect on lineup outcomes of a previous selection from a prior showup. Data were collapsed across live and video showups to increase statistical power. A 2 (video lineup: choosing, non-choosing) x 3 (condition: showup chooser, showup non-chooser, control) chi-squared test was significant, $\chi^2(2, n = 142) = 70.81, p < .001, \Phi = .71$. Showup choosers made significantly more selections from the video lineup (100%) than controls (83.0%) who in turn made more selections than showup non-choosers (24.2%; $p < .001$ both analyses). A similar analysis compared the effect of a prior correct culprit identification from a showup with a prior incorrect response. A 2 (video lineup selection: correct culprit selection, other selection) x 3 (condition: showup chooser, showup non-chooser, control) chi-squared analysis was significant, $\chi^2(2, n = 142) = 65.92, p < .001, \Phi = .68$. More correct culprit selections were made by showup choosers (93.5%) than controls (72.3%) who in turn made more than showup non-choosers (12.1%; $p < .001$ both analyses).

Culprit-absent video lineups: The outcome of culprit-absent lineups (innocent suspect selections, foil selections and correct rejections) did not differ across the three conditions (prior viewing of a live showup, a video showup or no prior identification attempt - control), $\chi^2(4, n = 141) = 7.70, p > .1, \Phi = .17$.

There was a significant influence of choosing from a prior showup on the outcome of culprit-absent lineups. Because all lineups in this analysis were culprit-absent, all identifications were incorrect. The proportion of participants who made an identification from the video lineup (lineup chooser vs. lineup non-chooser) differed as a function of

whether they had made a prior identification from a showup (showup chooser, showup non-chooser, control), $\chi^2(2, n = 141) = 7.44, p < .05, \Phi = .23$. Showup choosers made more lineup selections (100%) than controls (42.6%) and non-choosers (38.2%; $p < .001$ both analyses). The latter conditions did not differ ($p > .05$).

The proportion of participants who made a mistaken identification of the innocent suspect from a video lineup differed as a function of condition (video lineup: innocent suspect selection, other decision) across the three conditions (showup chooser, showup non-chooser, control), $\chi^2(2, n = 141) = 24.47, p < .001, \Phi = .42$. Showup choosers made more innocent suspect selections (80%) than non-choosers (12.4%) and controls (4.3%, $p < .001$ both analyses). The difference between showup non-choosers and controls was not significant ($p > .05$).

Video lineup confidence: Table 2 displays the mean confidence levels in the decisions made to video lineups as a function of culprit presence, outcome, the media of a prior showup (if applicable) and the prior showup decision. Data were collapsed across culprit-present and culprit-absent trials to compare mean confidence of decisions to the video lineups in a 2 (accuracy) x 2 (lineup: chooser, non-chooser) x 3 (condition: prior live showup, video showup, control) ANOVA. The main effect of accuracy was significant, $F(1, 268) = 43.29, p < .001, \eta^2 = .139$; correct decisions were associated with higher confidence than incorrect decisions ($M = 73.4, SD = 22.1$ vs. $M = 54.8, SD = 22.2$). The main effect of decision was also significant, $F(1, 268) = 8.12, p < .005, \eta^2 = .008$; choosers had higher confidence than non-choosers ($M = 69.0, SD = 21.3$ vs. $M = 62.5, SD = 26.9$). The main effect of condition, $F(2, 268) = 1.12, p > .2, \eta^2 = .008$ and all interactions were not significant ($p > .1, \eta^2 < .015$ for all effects).

The final analysis examined lineup chooser and non-chooser confidence as a function of whether participants had made a previous identification to a showup. Data were collapsed

across live and video showup conditions. A 2 (accuracy) x 2 (lineup: chooser, non-chooser) x 3 (condition: showup chooser, showup non-chooser, control) ANOVA revealed a significant main effect of accuracy, $F(1, 270) = 25.58, p < .001, \eta^2 = .087$. Correct decisions were associated with higher confidence than incorrect decisions. The main effects of lineup choice, $F(1, 270) = 2.09, p > .1, \eta^2 = .008$; and condition, $F(1, 268) < 1, \eta^2 = .004$, and all the interactions were not significant ($p > .1$).

Predictive confidence, identification confidence and accuracy: Measures of confidence in being able to recognise the culprit, which were taken immediately after the staged incident, correlated significantly with accuracy of response to the video lineup, point-biserial $r(283) = .16, p < .01$, confidence taken after the showup, Pearson's $r(189) = .34, p < .01$, and confidence taken after the video lineup, Pearson's $r(280) = .32, p < .01$, but not with the accuracy of response to the showup, point-biserial $r(189) = .11, p > .05$. Using confidence rated immediately after the identification procedure, the accuracy-confidence correlation was significant for showups, Pearson's $r(189) = .41, p < .01$, and for lineups, Pearson's $r(280) = .38, p < .01$.

Self-report data: Most participants (showup: 74.5%, lineup: 87.9%) reported that they remembered the warning about the potential absence of the culprit. However, recalling the warning had no significant influence on showup accuracy, $\chi^2(1, n = 188) = 1.06, p > .1, \Phi = .07$ or on choosing (chooser vs. non-chooser), $\chi^2(1, n = 188) < 1, \Phi = .05$, or lineup accuracy, $\chi^2(1, n = 281) = 1.55, p > .1, \Phi = .07$, or choosing $\chi^2(1, n = 281) < 1, \Phi = .02$. In addition, 18.6% of showup participants and 29.9% of lineup participants felt under pressure to make a positive identification. However, pressure to select had no influence on showup accuracy $\chi^2(1, n = 188) = 2.15, p > .1, \Phi = .11$, or choosing, $\chi^2(1, n = 188) < 1, \Phi = .01$, or video lineup accuracy, $\chi^2(1, n = 281) = 2.29, p > .1, \Phi = .07$, or choosing, $\chi^2(1, n = 281) = 1.50, p > .1, \Phi = .09$.

Discussion

From an applied perspective, the main aim of Experiment 1 was to compare the reliability of a live showup with a video lineup as the first means of identification. When present, the culprit was identified by fewer participants (51%) from a live showup, than from a video lineup (72%) or a video showup (79%). The data suggest that identification is more effective from a video identification procedure, whether it is a showup or a lineup, than from a live showup. In contrast, previous research has found that correct culprit-present identification rates in showups and lineups are similar (Stebly *et al.*, 2003). However, in most of the previous research, photograph showups and photograph lineups have been used.

The overlap of cues available in a live showup (e.g., height, gait, build, and clothing) with the original live incident would be greater than those available from a head and shoulders video clip. By this reasoning, the Encoding Specificity Principle (Tulving & Thomson, 1973) would predict better recognition from the live showup condition. However, in the staged act, the culprit had been wearing a coat and scarf, which was removed for the showups and the video lineup. In the videos, only head and shoulders views were available. It is possible that the mismatch in clothing cues was more salient in the live showup than in the head-and-shoulders video clips. By this reasoning, the results could be seen as consistent with the Encoding Specificity Principle. An alternative interpretation is that a live procedure may inhibit identification, possibly due to social inhibition from closely examining the appearance of somebody physically present. This issue was explored further in Experiment 2.

The rate of mistaken identification of an innocent suspect was relatively low in all first identification procedures (4.2% from live showups, 6.5% from video showups and 4.3% from a video lineup). There was no evidence to support the contention that live showups are any less fair than video lineups. In Experiment 1, the culprit had short hair but the innocent suspect had long hair. Some effort was made to remedy this discrepancy as the innocent

suspect tied her hair back and in the staged act the culprit wore a scarf. However, a common description of the culprit was that her hair was short in a ‘boyish’ cut. Research has found that external facial features (Ellis, Shepherd & Davies, 1979), particularly hairstyle (O’Donnell & Bruce, 2001), are most salient in identification of unfamiliar faces. It is possible that if the actresses had a more similar hairstyle, error rates would have been higher. In further experiments, showups were conducted with other pairs of actresses and actors to assist evaluation of whether the effects reported here generalize across different people.

Video lineups were held after a delay of approximately 30 minutes compared to a delay of 15 minutes for a showup. Forgetting over the extra 15 minute delay cannot account for the pattern of results observed in either culprit-present or culprit-absent lineups.

In summary, a live showup was a somewhat less effective means of obtaining identification evidence against a guilty suspect than a video lineup, but it was equally fair to an innocent suspect. This conclusion is rather counter-intuitive. The effect may be driven by a reluctance to identify somebody physically present in the room, even though no overt identification was required in the presence of the suspect. The effect might be attributed to social inhibition or to clothing change. The effect of clothing change was investigated in Experiment 4. An important issue for further applied research is to establish whether the same reluctance to identify a live suspect would be observed if the witness had recently been the victim of a street robbery or assault.

A second aim of the study was to examine whether a showup would influence a later video lineup decision. There was little difference in the pattern of lineup selections in culprit-present conditions between the controls, whose first decision was to the lineup, and those who had taken part in a previous showup. Furthermore, no differences were found across the three conditions in culprit-absent trials. As such, these results might suggest that there was no influence of a prior showup on identification accuracy. However, when the participants who

completed a showup were divided into choosers and non-choosers, it is apparent that their prior identification had influenced their behaviour at the lineup. All of the showup choosers in both culprit-present and absent conditions made a selection from the lineup. The majority selected the suspect seen approximately 15 minutes earlier in a showup (92.5%). These data demonstrate support for a commitment effect. The majority of showup non-choosers also rejected the lineup (65.6%) again providing further evidence of a commitment effect. In this case the commitment is to not identifying the suspect seen in the showup and lineup. However, some showup non-choosers made a positive selection of the suspect from the lineup (12.3%), despite having previously responded that they had not seen that actress during the staged act. In culprit-present trials, these decisions were correct. In culprit-absent trials these responses provided evidence of source confusion. Presumably these participants in culprit-absent trials mistakenly attributed a feeling of familiarity for the innocent suspect to having seen her during the staged incident. Although showup non-choosers made three times as many innocent suspect selections from the lineup than controls, this difference was not statistically significant. The trend in the data suggests some difficulty in being able to discriminate the source or context in which the innocent suspect had first been encountered. Source confusion was explored further in Experiment 3.

Consistent with a meta-analysis by Sporer *et al.* (1995) the relationship between accuracy and confidence was stronger amongst video lineup choosers than amongst non-choosers. However, no difference in confidence was found between choosers and non-choosers from the showups. Ratings of confidence in being able to make an identification of the culprit taken immediately after the incident were associated with video showup and lineup accuracy, but not with live showup accuracy. These effects were explored further with different actors in the other experiments reported here.

Experiment 2

In the culprit-present trials in Experiment 1, accuracy was higher in the video showup and video lineup trials than in the live showups. However, participants reported the highest levels of confidence in live showup decisions, suggesting that they may have been more conservative when making an identification from a live showup. No accuracy differences were found in culprit-absent trials, although performance was close to ceiling. Egan *et al.* (1977) found the opposite effect; with better performance in live than in photographic lineups. Reviewing the literature, Cutler *et al.* (1994) concluded that the test media has only a limited influence on outcomes. Choosing rates tend to be lower in photograph showups than in photograph lineups although this results in similar levels of correct culprit selections (Stebly *et al.*, 2003).

In Experiment 1 a psychological reluctance to select may have been experienced more acutely by participants in the live trials. If applied to real cases, this suggests that a video showup might be more effective than a live showup. The aim of Experiment 2 was to establish whether the lower choosing rate from a live showup could be replicated using a different culprit. In this instance a male actor served as the culprit. Participants took part in a single identification procedure; live showup, video showup or video lineup after a delay of 15 minutes. The culprit was present in all trials. Following the results of Experiment 1, it was predicted that accuracy would be lowest and confidence highest in live showups.

Method

Design: Experiment 2 employed a between-subjects design and a similar staged scenario to Experiment 1. Participants took part in a single culprit-present identification trial, with the actor present in person (*live showup*), displayed on a single video clip (*video showup*) or as a member of a nine-person lineup (*video lineup*). The dependent variables were identification outcome and confidence.

Participants: Participants, recruited in groups (mean size = 21.8, max = 30, min = 11) were first year psychology undergraduates (34 male; 119 female: aged 17 - 54: $M = 20.8$) at the University of Aberdeen, attending as a requirement of their course. Additional participants were excluded for being familiar with the actor, for being pre-warned of the experimental rationale or for not responding to critical questions ($n = 13$). Of the 149 participants who provided their ethnicity and nationality, 79.2% were from the UK and 18.1% from other EU countries. The majority (95.3%) were white European.

The actor was a male Scottish white European student, aged 21, 5' 7" tall with reddish brown hair. During the staged act he wore a plain white t-shirt, blue jeans and trainers. In the identification trials, he wore a black sweatshirt.

Materials and procedure: Experiment 2 closely replicated Experiment 1 except a male actor was recruited, all trials were culprit-present and there was no second identification phase. Grampian Police constructed a video lineup using the VIPERTM system. The video clip of the culprit was extracted from the video lineup and used for the video showups. The fairness of the lineup, based on a description of the culprit was measured using the same procedure and participants as described for Experiment 1. The culprit was selected by 7.4% of participants, fewer than 11.1% that would be expected by chance, suggesting that the culprit did not stand out. One lineup foil was selected by 25.9% of participants.

Results

All trials in Experiment 2 were culprit-present. An identification was made by 66.0% of participants. Table 3 depicts the percentage of each identification outcome as a function of condition. The first analysis compared choosing rates across the conditions. A smaller proportion of participants made an identification, of the culprit or a foil, from the video showup (65.5%) than from the live showup (79.6%) and video lineup (80.5%), but the difference was not significant, $\chi^2(2, n = 153) = 3.99, p > .1, \Phi = .16$. However, there was a

significant difference in the accuracy of decisions across conditions, $\chi^2(2, n = 153) = 7.96, p < .05, \Phi = .23$. Post hoc Fisher's exact tests found that fewer correct culprit identifications were made to video lineups (52.2%) than in the live and video showups ($p < .01$ for both analyses), which did not differ ($p > .1$).

Table 3 about here

Confidence levels: Table 4 displays the mean confidence levels as a function of condition, accuracy and decision. A 2 (decision: chooser, non-chooser) x 3 (condition) ANOVA examined the effect of choosing and condition on confidence. There was a significant main effect of decision, $F(1, 147) = 18.04, p < .001, \eta^2 = .109$. Choosers had higher confidence than non-choosers. The main effect of condition was also significant, $F(2, 147) = 3.50, p < .05, \eta^2 = .045$. Tukey's post-hoc tests found that live showup confidence was significantly higher than the other two conditions ($p < .05$) which did not differ ($p > .1$). The interaction was not significant, $F(2, 147) < 1, \eta^2 = .011$.

A second analysis examined the effect of accuracy and condition on confidence in a 2 (accurate vs. inaccurate) x 3 (live showup, video showup, video lineup) ANOVA. This revealed a significant main effect of accuracy, $F(1, 147) = 22.70, p < .001, \eta^2 = .134$, correct responses were associated with higher confidence than incorrect responses. The main effect of condition, $F(2, 147) = 2.54, p > .05, \eta^2 = .033$; and the interaction were not significant, $F(2, 147) < 1, \eta^2 = .010$.

Table 4 about here

Predictive confidence, identification confidence and accuracy: Confidence in being able to recognise the culprit, which was taken immediately after the staged incident, correlated significantly with confidence taken after the identification procedure, Pearson's $r(153) = .41, p < .01$, but not with identification accuracy, point-biserial $r(153) = .12, p < .01$.

The confidence after the identification procedure was significantly correlated with accuracy, Pearson's $r(153) = .41, p < .01$.

Self-report data: Most (73.7%) participants reported remembering hearing the potential absence of the culprit warning. However, recalling the warning had no significant influence on identification accuracy, $\chi^2(1, n = 152) < 1, \Phi = .05$ or choosing, $\chi^2(1, n = 152) < 1, \Phi = .07$. In addition, 23.5% of participants felt under pressure to make a positive identification. However, pressure to select had no influence on accuracy, $\chi^2(1, n = 153) = 2.30, p > .1, \Phi = .12$. or choosing, $\chi^2(1, n = 153) < 1, \Phi = .04$.

Discussion

Experiment 2 partly replicated Experiment 1 by comparing culprit-present identification accuracy in a live showup, a video showup and a video lineup. In Experiment 1 accuracy was lowest but confidence was highest when the actress was in a live showup. In Experiment 2, the highest rates of accuracy and confidence were in the live showup trials, although accuracy did not significantly differ from the video showup condition. One explanation for these results was found in some of the descriptive data given by participants. The actor in Experiment 2 was described as possessing a distinctive 'swagger' or gait. Gait recognition of familiar people is possible from extremely impoverished images such as dynamic point-light displays (Cutting & Kozlowski, 1977) and it is possible to learn and identify idiosyncratic stride patterns of previously-unfamiliar people from this type of display (Stevenage, Nixon & Vince, 1999). There was no attempt to disguise gait when the actor re-entered the room for the live showup trial and it is possible that decisions were partly made on this basis. Nevertheless, even with these cues, 20% of participants failed to recognise the actor in person even though the incident had taken place only a few minutes earlier. In conclusion, the relative sensitivity of a culprit-present live showup compared to a video

showup appears to depend on the idiosyncratic cues of the culprit, such as gait, that may be revealed in a live showup but not in a video showup.

A methodological refinement of Experiment 2 was that it eliminated the difference in delay prior to a showup or a lineup that was present in Experiment 1. The data are consistent with this small difference in delay having had no effect on identification performance.

Experiment 3

Long delays between a crime being committed and a witness viewing a lineup are common in criminal investigations. The modal delay for a sample of police lineups conducted in London was between 1 and 2 months (Valentine, Pickering & Darling, 2003). An archival study of criminal cases in the USA found that identification of the police suspect was more likely if the lineup occurred within 7 days (55%) rather than after 7 days (45%; Behrman & Davey, 2001). Similarly, Valentine *et al.* (2003) found that 65% of witnesses identified the police suspect from live lineups held within 7 days but only 40% identified the suspect if the lineup was held more than 7 days after the incident. In archival studies it cannot be established independently how many suspects are the culprit. Therefore it is important to consider the experimental literature.

The experimental evidence on the effect of delay can appear surprisingly ambiguous. (See Dysart & Lindsay 2007b for a review.) Yarmey, Yarmey and Yarmey (1996) found accuracy of identification from culprit-present trials was reduced by 15% with an increase in delay from 2 hours to 24 hours. In culprit-absent trials, accuracy in lineups was slightly reduced. However, delay had a larger negative impact on correct showup rejections. Cutler, Penrod, O'Rourke and Martens (1987) found that culprit-present and culprit-absent lineup accuracy was lower after 28 days than after 7 days. When participants were asked to re-read the description they had originally provided before viewing the lineups, no effect of delay was found. Other researchers have found no effect of delay on mugshot identifications

between 4 min and 1 week (Laughery, Fessler, Lenroovitz & Yoblick, 1974) or between 30 min and 2 days (Mauldin & Laughery, 1981). Dysart and Lindsay (2007b) concluded that delay adversely influences showups after 24 hours, and probably increases errors in both culprit-present and culprit-absent lineups, although other factors may have more impact. In a meta-analysis of 57 studies, Deffenbacher, Bornstein, McGorty and Penrod (2008) found a reliable association between longer retention intervals and lower identification accuracy of faces seen once. The effect was equally strong in face recognition paradigms (which usually required discrimination of many faces previously seen once from new faces), and in eyewitness identification paradigms, which usually require identification of a single person from a lineup.

Most live showups occur shortly after an offence. Davis *et al.* (submitted) found that only 1.2% of live showups in real cases in the UK occurred with a delay of more than 2 hours. In 38 cases a live showup was followed by the same witness attempting to identify the same suspect in a video lineup, with 10.5% of lineups conducted on the day of the offence, 68.4% conducted within a week and 11.5% between one week and a month later ($M = 19$ days). Some occurred 112 days later. Most studies of the commitment effect have used short delays (e.g., 20 min – 2 days) between the first and second identifications. However, a commitment effect from a showup (Godfrey & Clark, 2010) and from a mugshot search (Goodsell *et al.*, 2009) has been found to persist after 7 days. Pezdek and Blandon-Gitlin (2005) found no effect of commitment from a first lineup, which followed 2 hours after viewing a target face, on identification from a second lineup after 1 month. The effect of delay found in archival studies of police lineups suggests that a delay of more than 7 days may have a significant impact on identification ability (e.g., Valentine *et al.* 2003). Therefore, the aim of Experiment 3 was to examine the effect of delay after taking part in a live showup on the outcomes of a video lineup held 1 - 6 days, or 9 - 30 days later. Half of the

participants, designated as controls only viewed a video lineup. Live showup choosers were expected to demonstrate a commitment effect by selecting the same suspect in the video lineup, regardless of accuracy. Evidence of source confusion was expected to be observed in culprit-absent trials, in that showup non-choosers would select the innocent suspect from the lineup more often than controls. Lineup accuracy was predicted to be greater if conducted within a week, rather than more than a week later.

Method

Design: Experiment 3 employed a 2 (*culprit presence*) x 2 (*identification procedure: prior live showup vs. control*) x 2 (*delay before video lineup*) between-subjects design.

Participants viewed a staged act involving the culprit. Later half of the participants attempted to identify her in a live showup featuring either the same actress (*culprit-present*), or an innocent suspect (*culprit-absent*). The remainder did not take part in a showup (*control condition*). Video lineups were presented in a second phase, 1 - 6 days (*less than one week*), or 9 - 30 days (*more than one week*) later. The primary dependent variable was identification outcome.

Participants: Four hundred and sixty-nine students, staff and visitors to Goldsmiths, University of London, recruited in groups (mean group size = 12.0, max = 31, min = 1) viewed the plagiarism scenario used in Experiment 1. All who returned to view a video lineup were compensated £10. Data were excluded from participants if after the scenario they responded that they would definitely not recognise the actress, were familiar with her, or were forewarned as to the design ($n = 63$). Data from 406 participants (122 Male, 284 female, aged 18 – 56: $M = 24.0$ years) were retained, including 38 participants who took part in a showup but did not return for a lineup. Of the 368 participants who provided their ethnicity and nationality, 65.8% were from the UK, 16.6% from other EU countries, 6.2% from North

America, with the remainder from elsewhere. The majority (62.5%) were white European, with substantial black (11.4%) and Asian (8.7%) minorities.

Two undergraduate students played the part of the culprit and the innocent suspect. In the same manner as in Experiment 1, they were recruited from approximately 30 who responded to a poster advert. The two actresses selected (Figure 2), were independently rated to be the most similar in appearance. Both were from the UK, of white European ethnicity, with mid-length wavy dark brown hair, 5' 2" – 5'4" tall and of slim build. For the plagiarism scenario, the culprit wore a black mid-length coat over a black sweatshirt, blue jeans and white trainers, with her hair tied back. For the showups, the culprit removed the coat and untied her hair, whereas the innocent suspect had her hair tied back. These manipulations were included to simulate a situation in which the culprit took immediate and simple action to alter her appearance, and an innocent suspect who fitted the description was stopped. In the video lineups, both actors had their hair tied back and wore a plain navy blue fleece jacket with a collar. The appearance in filming the lineup is under control of the police, therefore an instruction to tie back her hair may be given.

Figure 2 and 3 about here

Procedure and materials: Experiment 3 closely replicated Experiment 1 except that participants booked individual video lineup sessions 1 - 6 days ($M = 3.3$, $SD = 2.1$), or 9 - 30 days ($M = 15.4$, $SD = 6.3$) later. Separate PROMAT™ video lineups were constructed for both actresses. The foils were selected by an experienced police officer. Even though the actresses were selected to be of similar appearance, the foils selected for the two lineups were different in each. The same group of pilot participants from Experiment 1 evaluated the fairness of the culprit-present lineup. In this case the actual culprit was selected by 22.2% of participants, more than the 11.1% that would be expected by chance, suggesting that based on

a description the culprit stood out from the foils. However, one foil was actually selected by slightly more participants (24.1%).

Results

Table 5 gives the percentage outcomes for showups and video lineups as a function of culprit-presence, condition, and delay. The outcome of second identification decisions to video lineups are expressed as a function of earlier showup decisions (showup chooser vs. non-chooser). In culprit-present showups, 45.6% selected the culprit. In culprit-absent showups, 35.9% selected the innocent suspect. Combining all data from the control and second identification decisions, in the culprit-present video lineups, 66.1% selected the culprit, 15.9% selected a foil and 18.0% rejected the lineup. In the culprit-absent video lineups 54.7% selected the innocent suspect, 12.9% selected a foil and 32.4% rejected the lineup.

Table 5 about here

First identification decisions: A 3 x 2 chi-squared test showed a significant effect of first identification procedure (showup, video lineup within 1 week, video lineup after more than 1 week) on the proportion of selections (choosing vs. non-choosing) in culprit-present trials, $\chi^2(2, n = 218) = 27.50, p < .001, \Phi = .36$. Post hoc Fisher exact tests showed that fewer selections were made to showups (45.6%) than to control lineups conducted within a week (82.0%) or more than a week (79.1%; $p < .01$). The latter conditions did not differ ($p > .2$). Selections from culprit-present showups are always of the culprit, but selections from a lineup could be of the culprit or a foil. For a subsequent 3 x 2 chi-squared test responses were categorised as correct culprit selections or incorrect responses. The nature of the first identification procedure (showup, video lineup within 1 week, video lineup after more than 1 week) had no significant effect on the accuracy of response to culprit-present procedures, $\chi^2(2, n = 218) = 4.04, p > .1, \Phi = .14$.

Similar analyses were conducted in culprit-absent trials. A chi-squared test that compared the proportion of mistaken identifications across conditions was significant, $\chi^2(2, n = 188) = 14.72, p < .001, \Phi = .28$. Fisher's exact tests found that fewer mistaken identifications were made from showups (35.9%) than from control lineups conducted within a week (60.8%) or after more than a week (66.7%; $p < .01$). The latter conditions did not differ ($p > .2$). A similar chi-squared test on the proportion of mistaken identifications of the innocent suspect compared to other responses showed no effect of the nature of the first identification procedure on the accuracy of responses, $\chi^2(2, n = 188) = .93, p > .2, \Phi = .07$.

First decision confidence: Table 6 displays the mean confidence levels of first decisions to showups and video lineups as a function of culprit presence, outcome, and condition. A 2 (accuracy: correct, incorrect) x 2 (decision: choosers, non-choosers) x 3 (condition: showup, lineup < 1 week, lineup > 1 week) ANOVA on the first decision confidence levels found that the main effect of decision was significant, $F(1, 394) = 17.77, p < .001, \eta^2 = .043$. Choosers ($M = 69.1, SD = 19.9$) had higher confidence than non-choosers ($M = 58.6, SD = 25.5$). The main effects of accuracy, $F(1, 394) = 2.79, p > .05, \eta^2 = .007$, and condition, $F(2, 394) = 1.30, p > .05, \eta^2 = .007$ were not significant. The interaction between decision and accuracy was significant, $F(1, 394) = 7.61, p < .01, \eta^2 = .019$. Simple main effects found that the effect of accuracy was significant for choosers, $F(1, 394) = 15.16, p < .001, \eta^2 = .037$ (correct $M = 75.8, SD = 18.6$; incorrect $M = 63.0, SD = 19.1$), but not for non-choosers, $F(1, 394) < 1, \eta^2 = .001$ (correct $M = 57.3, SD = 25.7$; incorrect $M = 60.0, SD = 25.4$). The remaining two-way interactions and the three-way interaction were not significant ($F < 1; \eta^2 < .004$).

Table 6 about here

Culprit-present video lineups: Further analyses examined the effect of delay and condition on the outcome of culprit-present video lineups from Table 5. A 3 (lineup outcome:

correct culprit ID, foil ID, lineup rejection) x 2 (condition: previous showup, control) x 2 (delay: within a week, more than a week) backward elimination hierarchical log-linear analysis (probability for removal $p < .05$) resulted in a final model involving the interaction between lineup outcome and condition with a likelihood ratio, $\chi^2(6) = 4.18$, $p = .65$. Delay had no effect on identification performance ($p > .05$). Post hoc chi-squared analysis on the interaction, $\chi^2(2, n = 189) = 15.11$, $p < .01$, $\Phi = .28$, revealed that showup participants (77.1%) were more likely than controls (54.8%) to select the culprit from the lineup. Lineup rejections did not significantly differ. Controls made more foil selections (25.8%) than the showup participants (6.3%).

Culprit-absent video lineups: A similar 3 (lineup outcome: suspect ID, foil ID, lineup rejection) x 2 (condition: street ID, control) x 2 (delay: within a week, more than a week) hierarchical log-linear analysis conducted on culprit-absent lineup responses produced a final model with a likelihood ratio, $\chi^2(6) = 2.73$, $p = .84$, that included the interaction between lineup outcome and condition. Delay had no influence on outcomes ($p > .05$). A chi-squared analysis, $\chi^2(2, n = 179) = 16.81$, $p < .01$, $\Phi = .31$, on the interaction found that showup participants made more innocent suspect selections (68.7%) from the lineup than controls (42.7%). Lineup rejections did not significantly differ. However, controls made more foil selections (20.8%) than the showup participants (3.6%).

Combined culprit-present and absent video lineup data: Data from culprit-present and culprit-absent lineups were analysed together. The greater number of observations allowed the influence of a previous showup decision and delay on the outcome of video lineups to be examined. A 2 (culprit presence: culprit-present, culprit-absent) x 2 (Lineup choosing: choosing, non-choosing) x 3 (condition: showup chooser, showup non-chooser, control) x 2 (delay: within a week, more than a week) hierarchical log-linear analysis produced a final model with a likelihood ratio, $\chi^2(16) = 10.23$, $p = .86$, that included firstly,

the interaction between lineup choosing and culprit presence and secondly, the interaction between lineup choosing and condition. A chi-squared analysis, $\chi^2(1, n = 368) = 10.15, p < .01, \Phi = .17$, examining the interaction between lineup choosing and culprit presence revealed that more selections were made to lineups in culprit-present trials (82.0%) than in culprit-absent trials (67.6%). A chi-squared analysis, $\chi^2(1, n = 368) = 25.77, p < .01, \Phi = .27$, examining the interaction between lineup choosing and condition was the result of significantly more showup choosers (97.3%) making a lineup selection than showup non-choosers (65.1%) and controls (72.0%). The latter did not differ ($p > .1$).

A further analysis examined the effect of the same factors on identifications of the suspect. A similar 2 (culprit presence: culprit present, culprit absent) x 2 (lineup outcome: suspect selection, other decision) x 3 (condition: showup chooser, showup non-chooser, control) x 2 (delay: within a week, more than a week) hierarchical log-linear analysis examined whether identification of the suspect was influenced by previous showup decisions and delay. The final model included the interaction between lineup outcome and culprit presence and the interaction between lineup outcome and condition, and had a likelihood ratio, $\chi^2(16) = 15.14, p = .51$. A chi-squared analysis, $\chi^2(1, n = 368) = 4.99, p < .05, \Phi = .12$, examining the interaction between lineup outcome and culprit presence revealed that the suspect was more likely to be identified in culprit-present trials (66.1%) than in culprit-absent trials (54.7%). A significant chi-squared analysis, $\chi^2(2, n = 368) = 38.45, p < .05, \Phi = .32$, and follow up Fisher exact tests examining the interaction between lineup outcome and condition was the result of significantly more showup choosers (90.4%) selecting the suspect from the lineup than did showup non-choosers (61.3%). There was a trend for the non-choosers to make more suspect selections than the controls (48.7%). However, this did not reach significance ($p > .05$), indicating that the effect of source confusion without an effect of commitment was not significant.

In the third analysis, the data provided by controls were excluded to specifically investigate the effect of a showup on video lineups as a function of delay. The control group data were not relevant to this comparison. A 3 (lineup outcome: suspect selection, other outcome) x 2 (culprit presence: culprit-present, culprit-absent) x 2 (showup: choosers, non-choosers) x 2 (delay: within a week, more than a week) hierarchical log-linear analysis produced a final model with a likelihood ratio, $\chi^2(10) = 12.89$, $p = .23$, that included the interaction between lineup outcome and showup decision and the interaction between lineup outcome and delay. A chi-squared analysis, $\chi^2(1, n = 179) = 18.64$, $p < .01$, $\Phi = .32$, examining the interaction between lineup outcome and showup decision revealed that showup choosers made significantly more suspect selections (90.4%) from the lineup than non-choosers (61.3%). The interaction between delay and lineup outcome was not significant when post hoc analyses were conducted, $\chi^2(1, n = 179) = 1.47$, $p > .1$, $\Phi = .09$, but was due to more participants selecting the suspect if the lineup occurred within a week (77.3%) than if it occurred after more than a week (69.2%).

Video lineup confidence: The culprit-present and culprit-absent video lineup confidence data from Table 6 were collated to examine the effects of accuracy, condition, video lineup choosing and delay on lineup confidence. It was not possible to separate showup choosers and non-choosers due to missing values in some conditions. A 2 (lineup accuracy: correct, incorrect) x 2 (condition: showup, control) x 2 (lineup decision: chooser, non-chooser) x 2 (delay: within a week, more than a week) ANOVA was conducted. The main effects of condition, $F(1, 352) = 7.25$, $p < .001$, $\eta^2 = .020$ and of decision were significant, $F(1, 352) = 12.60$, $p < .001$, $\eta^2 = .035$. Confidence was higher amongst showup participants than controls ($M = 72.4$, $SD = 20.8$ vs. $M = 63.5$, $SD = 22.9$), and choosers had higher confidence than non-choosers ($M = 70.2$, $SD = 20.3$ vs. $M = 60.7$, $SD = 26.5$). The main effect of accuracy, $F(1, 352) = 1.83$, $p > .1$, $\eta^2 = .005$ and delay, $F(1, 352) < 1$, $\eta^2 = .002$

were not significant. The interaction between decision and accuracy was significant, $F(1, 352) = 4.80, p < .05, \eta^2 = .013$. Simple main effects found that the accuracy effect was significant for choosers, $F(1, 352) = 13.07, p < .001, \eta^2 = .036$ (correct $M = 76.1, SD = 18.4$ vs. incorrect $M = 65.4, SD = 20.5$) but was not significant for non-choosers, $F(1, 352) < .1, \eta^2 = .001; M = 60.6, SD = 26.5$. The interaction between decision and delay was also significant, $F(1, 352) = 8.07, p < .005, \eta^2 = .022$. Simple main effects found that the effect of decision was significant when lineups were conducted within a week, $F(1, 352) = 19.71, p < .001, \eta^2 = .053$ (choosers $M = 73.3, SD = 20.6$ vs. non-choosers $M = 55.6, SD = 28.3$); but it was not significant after a week, $F(1, 356) < .1, \eta^2 = .001$ (choosers $M = 66.7, SD = 19.4$ vs. non-choosers $M = 65.5, SD = 24.0$). The remaining interactions were not significant ($p > .1, \eta^2 < .003$).

Predictive confidence, identification confidence and accuracy: Measures of confidence in being able to recognise the culprit, which were taken immediately after the staged incident, correlated significantly with showup accuracy, point-biserial $r(217) = .19, p < .01$, confidence taken after the showup, Pearson's $r(217) = .37, p < .01$, and confidence taken after the video lineup, Pearson's $r(368) = .19, p < .01$, but not with the accuracy of response to the video lineup, point-biserial $r(368) = .06, p > .05$. Using confidence rated immediately after the identification procedure, the accuracy-confidence correlation was significant for the video lineup, Pearson's $r(368) = .13, p < .05$, but not for the showup, Pearson's $r(217) = .11, p > .05$.

Self-report data: Most (showup: 74.6%; lineup: 94.6%) participants reported remembering the warning about the potential absence of the culprit. However, remembering the warning had no significant influence on showup accuracy, $\chi^2(1, n = 217) < 1, \Phi = .01$, or choosing, $\chi^2(1, n = 217) < 1, \Phi = .01$, or video lineup accuracy, $\chi^2(1, n = 367) < 1, \Phi = .05$. However, it did significantly influence lineup choosing, $\chi^2(1, n = 347) = 4.54, p < .05, \Phi =$

.11. Ninety-five percent of the few participants who did not remember the warning made a selection compared to 74% of those who *did* remember the warning. Although 31.5% of showup participants and 29.2% of lineup participants felt under pressure to make an identification, pressure to select had no influence on showup accuracy, $\chi^2(1, n = 216) < 1$, $\Phi = .03$, or choosing, $\chi^2(1, n = 367) < 1$, $\Phi = .01$, or video lineup choosing, $\chi^2(1, n = 281) = 1.50$, $p > .1$, $\Phi = .09$. However, there was a significant relationship between pressure to choose and video lineup accuracy, $\chi^2(1, n = 367) = 4.33$, $p < .05$, $\Phi = .11$. When participants felt under pressure to select, they made more incorrect decisions (58.9%) than correct decisions (41.1%). In contrast, when they felt there was no pressure to select, there were more correct decisions (53.1%) than incorrect decisions (46.9%).

Discussion

Only 46% of participants correctly identified the actress in a culprit-present, live showup a few minutes after seeing a staged incident. She had changed her appearance quite significantly by simply removing her coat and letting her tied-back hair loose. These results highlight the difficulty that may be encountered in a live showup used in a criminal case if the culprit takes simple steps to quickly change their appearance.

When the culprit was replaced by an innocent suspect of similar appearance and hairstyle, about 36% of participants made a false positive identification decision in culprit-absent showups. Even after such a brief delay, these results demonstrate the potential for misidentification if individuals have such similar features. It may seem an unlikely coincidence that an innocent suspect with such a close resemblance to the real perpetrator would be in the vicinity of a crime scene. However, the two actresses were selected from a small pool of only 30 applicants, in exactly the same manner as in Experiment 1. In Experiment 1, even though efforts were made to ensure that the innocent suspect resembled the culprit as closely as possible (e.g., wearing a scarf to obscure differences in hair length),

culprit-absent showup error rates were very low, meaning that the power to investigate carry-over effects to a video lineup was limited. In an attempt to increase statistical power for carry-over effects in Experiment 3, the two actresses were primarily selected for their highly similar hairstyles. Rates of misidentification were not anticipated to be as high as was observed. Nevertheless, when selecting the two actresses from photographs, the experimenters deliberated as to whether this pair, or another pair was more similar in appearance, suggesting that a close resemblance is not uncommon.

Controls made more identifications from a video lineup than did the participants in showups, despite the lineups being held after a delay of at least one day compared to a delay of approximately 15 min for the showups. However, many selections from the lineups were of foils, and so accuracy did not differ between showups and lineups in both culprit-present and culprit-absent conditions. These data suggest that a video lineup is not more reliable than a showup. The rate of mistaken identification of the innocent suspect from a lineup (43%) suggests that the lineup was unfair, and is probably of a consequence of the similarity in appearance of the culprit to the innocent suspect. In these circumstances, no identification procedure will provide adequate protection to an innocent suspect. The lineups, which contained a different set of foils in culprit-present and culprit-absent conditions, were selected by an experienced police identification operator using the method and database used in real cases. The innocent suspect had her hair tied back in the lineup video to match the appearance of the culprit from the original staged act. So did seven of the eight foils. Therefore, this is the lineup that would have been used had the scenario been a real case. It was retained as constructed to preserve the validity of the experiment.

As in Experiment 1, regardless of culprit presence, the majority of participants (97.3%) who made an identification of the suspect in a showup also made a selection from a lineup, further evidence of a commitment effect. Indeed, most showup choosers (90.4%)

selected the same suspect from the lineup. In culprit-absent trials, although more participants who had been correct non-choosers (60.0%) in the showup trials than controls (42.7%) selected the innocent suspect from the lineup, this difference was not significant. This result is consistent with recent studies which have failed to find a significant effect of source confusion, in the absence of commitment to an identification (Blunt & McAllister, 2009; Goodsell *et al.*, 2009).

One of the primary purposes of Experiment 3 was to examine the effect of delay prior to video lineup following a live showup. As expected from previous field studies there was a trend for participants to be more likely to select the suspect, in both culprit-present and culprit-absent lineups, if the lineup occurred within a week of the incident than if it occurred after more than a week. In culprit-present lineups, the suspect was the culprit. In culprit-absent lineups, the suspect was the innocent suspect who had appeared in a prior showup. Furthermore, regardless of accuracy, those making an identification decision within a week were more confident than non-choosers. Godfrey and Clark (2010) report similar effects of delay between a showup and a photograph lineup in a comparison of 20 minutes and seven days delay. After seven days delay, viewing the suspect in a photograph showup induced more correct and incorrect identifications of the suspect from a photograph lineup. In conclusion the effects observed in Experiment 3 are consistent with the literature, which demonstrates there is a moderate effect of delay found in a meta-analysis of many studies but the effect is not consistently seen in individual studies.

Experiment 4

In Experiments 1 and 3, the innocent suspect presented in an identification procedure was not at the scene of the staged incident. However a defendant may admit presence at the scene but deny involvement in a crime. Furthermore, the suspect may have deliberately exchanged clothing with an accomplice to reduce the likelihood of detection. A witness'

description usually includes a description of clothing (Davis *et al.*, submitted) and an individual wearing similar clothing may be asked to remain in the vicinity for a showup. The effect of a change of clothing on identification decisions has been evaluated in previous studies (Dysart, Lindsay & Dupuis, 2006; Lindsay, Wallbridge & Drennan, 1987; Yarmey *et al.*, 2006). Lindsay *et al.* examined culprit-present and culprit-absent lineup outcomes with all lineup members wearing a sweatshirt as worn in the original staged theft, all wearing a lab coat, all wearing different clothing which was also different to that worn in the offence or a *biased* lineup where only the culprit or an innocent suspect wore the same clothing as during the offence. No significant effects were found in culprit-present lineups. However, more innocent suspect selections were made in the biased culprit-absent trials.

Clothing bias effects have also been found in relation to culprit-absent showup identifications, but only when innocent suspects possessed a moderately similar appearance to the culprit (Dysart *et al.*, 2006; Yarmey *et al.*, 1996). Dissimilar suspects were not misidentified, suggesting a limitation to the strength of the bias. However, Dysart *et al.* also demonstrated that clothing distinctiveness mediates this effect, a clothing bias was only found with a distinctive t-shirt, but not with a more typical blue shirt.

Identification performance can be biased by the presence of an innocent bystander at the scene of a crime in the same way as it can be biased by a previous attempt to identify the suspect. For instance, Buckhout (1974) found that 40% of participants selected the culprit from six-person lineups, but 25% selected an innocent bystander. Thus, there is evidence of transfer of familiarity in this context. In their meta-analysis, Deffenbacher *et al.* (1996) reported a significant effect of unconscious transference of familiarity from a bystander present at the scene; although the effect size was smaller than the bias observed from previous mug-shot exposure.

Experiment 4 was designed to examine the scenario in which an innocent bystander may later become the focus of a live showup or a lineup. Using a similar scenario as the previous experiments, *two* actors of dissimilar appearance complained that they had been wrongly accused of plagiarism. During the staged incident one actor wore a *distinctive* rugby shirt, while the second actor wore a *plain* shirt (see Figure 4). Clothing was counterbalanced across trials. Participants took part in a culprit-absent showup with the *quiet* innocent bystander as the target. Participants were specifically asked to identify if he had been the *loud* actor. Clothing in the live showup was counterbalanced, so that the innocent bystander either wore the same clothing as during the staged incident or had swapped with the loud protagonist. In the video lineups the innocent bystander was depicted wearing a different plain top.

Figure 4 about here

Method

Design: Experiment 4 employed a 2 (*incident clothing*: distinctive, plain) x 2 (*clothing exchange*: same, exchange) between-subjects design using a similar procedure as the previous experiments, except that two actors were involved in the staged incident. One actor, the *loud culprit* stood close to the experimenter. The second actor, the *quiet innocent bystander*, was seen for the same period of time but contributed a single verbal line. The first factor, *incident clothing* was whether during the staged scene the innocent bystander wore a distinctive top, or a plain top. This was counterbalanced so that in half the trials the innocent bystander would wear one top while the culprit wore the other top. The *clothing exchange* factor was whether in the showup the innocent bystander wore the same top as during the incident or had exchanged with the culprit. All participants took part in a *culprit-absent* video lineup, containing the innocent bystander. The same two factors from the first phase were retained. However, participants who had not seen a showup served as control participants for

the video lineup. The clothing worn by the innocent suspect in the lineup was different from the clothing worn by either the loud culprit or the innocent bystander during the incident. The primary dependent variable was identification outcome.

Participants: Participants were 276 psychology undergraduate students (79 male, 197 female aged 17 – 40; $M = 20.0$) recruited in groups (mean group size = 27.6, $SD = 3.0$, max = 37, min = 13) at the University of Aberdeen participating for course credits. Additional participants were excluded for being familiar with one of the two actors, for having been pre-warned of the experimental rationale or for predicting that they would not be able to recognise the actors ($n = 23$). All participants reported their ethnicity and nationality; 90.9% were from the UK, 7.2% from other EU countries, with the remainder from North America. Most (97.5%) were white European.

The actors were white European postgraduate students from Aberdeen University (Figure 4). Both had short dark brown hair, styled similarly and a similar complexion. The culprit was 6 foot in height, athletic build with an American accent. The innocent bystander was 5 foot 6 inches with a Scottish accent. For the plagiarism scene and the showup they wore either a *distinctive* purple, white and green striped rugby shirt with a club badge, or a *typical* predominantly black sweat shirt with purple square patches on the shoulders. Both wore blue jeans and similar dull white trainers.

Figure 3 about here

Materials: A VIPERTM video lineup was used for Phase 2. The lineup consisted of eight distracters selected by the police operator from a national database and the innocent bystander. The questionnaires for Experiment 4 were similar to those used in Experiment 1 except on the first post-incident questionnaire participants were asked in four separate items whether they recognised both actors and how confident they would be in recognising them later. All other questionnaire items referred to the loud actor only.

Procedure: The procedure was similar to that of Experiment 1 except both actors entered the classroom and a staged performance was enacted in which the loud culprit would criticise the experimenter loudly for incorrectly accusing them of plagiarism. The quiet innocent bystander contributed a single vocal line. Once they had left the room the purpose of the session was revealed, and participants were continually reminded throughout each stage that the person who had been standing closest to the experimenter and arguing the loudest was the ‘loud’ actor. The innocent bystander standing further from the experimenter during the staged incident was described as the ‘quiet’ actor. No reference was made to appearance. After descriptions had been collected of the loud actor, participants were informed that they were to attempt to identify the loud actor who may or may not be present in the showups and later video lineups. The innocent bystander was always present in both phases. The video lineups were held approximately 15-min after the showup.

The same group of pilot participants from Experiment 1 evaluated the fairness of the culprit-absent lineup. In this case the innocent bystander was selected by 35.2% of participants, more than the 11.1% that would be expected by chance, suggesting that based on a description of the loud culprit, the innocent bystander stood out from the foils. However, throughout, the participants in the main study were also only asked to describe and identify the loud actor.

Results

All identification trials in Experiment 4 were culprit-absent in which the innocent bystander was always present. Table 7 displays the percentage outcomes for showups and video lineups dependent on condition (showup, control), incident clothing, showup clothing and, when it was participants’ second decision, video lineup outcomes are separated into those who made an identification (chooser) or not (non-chooser) from the prior showup. In the showups 7.7% of participants selected the innocent bystander, whereas across all control

and second identification video lineups 7.6% selected the innocent bystander, 21.7% a foil, and 70.7% made a correct rejection.

Table 7 about here

First identification decisions: Two chi-squared analyses compared the first identification decisions made from showups with video lineups in the control condition. The clothing factors were not included in these analyses. A 2 x 2 chi-squared test with the first identification decisions classified as chooser or non-chooser was significant, $\chi^2(1, \underline{n} = 276) = 15.82, p < .001, \underline{\Phi} = .091$. Fewer identifications were made to showups (7.7%) than to control lineups (26.8%). With identifications classified as innocent bystander selections versus any other decision, a similar chi-squared test was also significant, $\chi^2(1, \underline{n} = 276) = 4.61, p < .05, \underline{\Phi} = .091$, because none of the participants in the control lineup condition selected the innocent bystander.

Further analyses examined the effect of clothing distinctiveness and clothing change on identification of the innocent bystander. Data from the controls lineups were not included in this analysis. A 2 (showup selection: innocent bystander selection, correct rejection) x 2 (incident clothing: distinctive, plain) x 2 (clothing exchange: same, changed) hierarchical log-linear analysis produced a model that included the interaction between showup selection and incident clothing with a likelihood ratio, $\chi^2(4) = 1.63, p = .80$. Post hoc chi-squared analysis of the interaction, $\chi^2(1, \underline{n} = 220) = 5.08, p < .05, \underline{\Phi} = .15$, found that more innocent bystander selections were made when he had worn the distinctive clothing during the incident (12.8%) than when he had worn the plain top (4.5%). There were no significant effects of clothing exchange.

First decision confidence: Table 8 depicts the mean showup and video lineup confidence ratings as a function of condition, decision, identification outcome, incident clothing and clothing exchange. As none of the controls selected the innocent bystander, it

was only possible to compare confidence in incorrect choosing and correct non-choosing. A 2 (condition: showup, control lineup) x 2 (first decision choice: chooser, non-chooser) ANOVA on the first decision confidence levels revealed a significant effect of condition, $F(1, 272) = 10.84, p < .005, \eta^2 = .036$. Showup participants ($M = 81.4, SD = 23.0$) had higher confidence than controls ($M = 60.9, SD = 24.4$). The main effect of decision was also significant, $F(1, 272) = 16.97, p < .001, \eta^2 = .059$. Correct non-choosers ($M = 80.0, SD = 23.9$) had higher confidence than incorrect choosers ($M = 55.9, SD = 19.7$). The interaction was not significant, $F(1, 272) < 1, \eta^2 = .003$.

Table 8 about here

The next analysis examined the effect of incident clothing and clothing exchange on showup confidence levels. Due to low numbers in some cells it was not possible to include showup accuracy as a variable. A 2 (incident clothing) x 2 (clothing exchange) ANOVA revealed that the main effect of incident clothing was not significant, $F(1, 216) < 1, \eta^2 = .004$. However, the main effect of clothing exchange was significant, $F(1, 216) = 4.91, p < .05, \eta^2 = .022$. This effect was moderated by a significant interaction, $F(1, 216) = 7.39, p < .01, \eta^2 = .033$. Simple main effects found that when the innocent bystander wore the distinctive clothing during the incident, the effect of clothing change was significant, $F(1, 216) = 9.98, p < .01, \eta^2 = .044$. Confidence was higher when clothing was the same during the live showup ($M = 87.2, SD = 17.7$) than when it was changed ($M = 71.8, SD = 28.8$). However, when the innocent suspect wore the plain clothing during the incident (same clothing $M = 81.5, SD = 24.6$; exchanged clothing, $M = 83.0, SD = 19.0$) there was no clothing exchange effect, $F(1, 216) < 1, \eta^2 = .001$.

Video lineups: A series of hierarchical log-linear analyses examined whether incident clothing and clothing change influenced video lineup responses. With the controls excluded, three separate backward elimination hierarchical log-linear analyses found no effects of

incident clothing or clothing change as a function of whether participants had been a showup chooser or non-chooser on video lineup outcomes (bystander selections, foil selections, lineup rejections), lineup accuracy (incorrect selections, correct rejections) or innocent bystander selections (bystander selections, other selection; all analyses $p > .4$).

A further test was conducted to examine the effect of making an identification to a showup on the outcome of a subsequent video lineup. With the inclusion of controls, a 3 (condition: showup chooser, showup non-chooser, control) x 2 (lineup outcome: chooser, non-chooser) chi-squared test was significant, $\chi^2(4, n = 276) = 15.65, p < .001, \Phi = .17$. More showup choosers (70.6%) made identifications from the lineup than did showup non-choosers (26.6%) and controls (26.8%, $p < .01$ both analyses). The latter did not differ ($p > .2$). To examine this effect further, lineup outcomes were collapsed into innocent bystander identifications versus other decisions. A chi-squared test comparing the three conditions was significant, $\chi^2(2, n = 276) = 85.53, p < .001, \Phi = .56$. More innocent bystander identifications were made by showup choosers (64.7%) than both non-choosers (4.9%) and controls (0%; $p < .01$ both analyses). None of the controls selected the innocent bystander, but this did not significantly differ from showup non-choosers ($p > .05$).

Video lineup confidence: The first analysis examined the effect of incident clothing, clothing exchange and video lineup accuracy on the confidence of decisions made to a video lineup by participants who had taken part in a showup. A 2 (incident clothing) x 2 (clothing change) x 2 (lineup accuracy: correct vs. incorrect) ANOVA found no significant effects or interaction ($p > .05, \eta^2 < .011$). A 2 (incident clothing) x 2 (clothing change) x 2 (bystander selection, other selection) ANOVA also found no significant effects ($p > .05, \eta^2 < .015$). Finally, a 2 (lineup accuracy: correct vs. incorrect) x 2 (condition: showup, control) ANOVA compared confidence of decisions to the video lineup made by participants who had taken part in a showup with controls. The main effect of accuracy was significant, $F(1, 272) = 4.34,$

$p < .05$, $\eta^2 = .016$. Confidence of correct decisions was higher ($M = 60.0$, $SD = 25.9$) than of incorrect decisions ($M = 54.9$, $SD = 18.8$). The main effect of condition, $F(1, 272) < 1$, $\eta^2 < .001$, and the interaction were not significant, $F(1, 272) = 1.90$, $p > .1$, $\eta^2 = .007$.

Predictive confidence, identification confidence and accuracy: Predictive measures of confidence in being able to recognise the culprit and innocent bystander were taken shortly after the staged incidents. The confidence in being able to recognize the bystander was correlated with confidence in being able to recognize the culprit, although participants were more confident that they would be able to recognise the loud culprit ($M = 55.3$, $SD = 23.8$) than the innocent bystander ($M = 42.9$, $SD = 26.8$; paired $t(275) = 9.48$, $p < .001$). Neither confidence in being able to recognize the culprit nor the bystander was correlated with accuracy in the showup, point-biserial $r(220) = .12$, $p > .05$; innocent bystander, $r(220) = .09$, $p > .1$, or in the video lineup, point-biserial $r(276) = .07$, $p > .05$; innocent bystander, $r(276) = .09$, $p > .1$. However, both measures of pre-identification procedure confidence correlated with confidence in the showup decision, Pearson's $r(220) = .41$ and $.45$, $p < .01$ respectively, and with lineup decision confidence, Pearson's $r(276) = .30$ and $.28$, $p < .01$ respectively. Using confidence rated immediately after the identification procedure, the accuracy-confidence correlation was significant for showups, Pearson's $r(220) = .26$, $p < .01$, but was not significant for lineups, Pearson's $r(276) = .10$, $p > .1$.

Self-report data: Most (showup: 71.2%, lineup: 84.1%) participants reported remembering the warning about the potential absence of the culprit. However, remembering the warning or not had no significant influence on showup accuracy, $\chi^2(1, n = 219) = 2.60$, $p > .1$, $\Phi = .11$, or lineup accuracy, $\chi^2(1, n = 276) > 1$, $\Phi < .01$, or choosing, $\chi^2(1, n = 276) > 1$, $\Phi = .03$,

In addition, 14.2% of showup participants and 30.0% of lineup participants felt under pressure to make a positive identification. Pressure to select significantly influenced showup

accuracy, $\chi^2(1, n = 219) = 6.78, p < .05, \Phi = .18$, in that 54.5% of choosers reported feeling under pressure in contrast to 14.1% of non-choosers. Furthermore, pressure to select significantly influenced video lineup choosing, $\chi^2(1, n = 276) = 5.73, p < .05, \Phi = .14$, and accuracy, $\chi^2(1, n = 276) = 6.20, p < .05, \Phi = .15$. However, follow-up Fisher's exact tests found that there was no influence of pressure on identifications of the innocent bystander ($p > .05$).

Discussion

In Experiment 4 the actors were dissimilar in appearance. Nevertheless 7.7% of showup participants misidentified the innocent bystander a few minutes after the incident. No innocent bystander identifications were made from control lineups. Errors were higher if the innocent bystander had been wearing the distinctive shirt during the incident. However, there was no effect of an exchange of clothing on identification rates.

Confidence in showup decisions did not show the same pattern of results. Confidence was higher amongst participants who made a correct rejection than amongst those who made an innocent bystander misidentification. However, when the innocent bystander wore the distinctive clothing during the incident, confidence was higher when he also wore it in the showup. There was no effect of clothing change when the innocent bystander wore the plain clothing during the incident. The results illustrate that a clothing bias can influence confidence in a live showup. A possible explanation is that when wearing the plain shirt the actor did not attract the attention of participants during the incident. Ratings taken immediately after the staged event showed that participants believed they would be more likely to recognise the loud protagonist than the innocent bystander.

Neither of the two clothing factors was found to influence video lineup accuracy or confidence. However, none of the controls made an incorrect selection of the innocent bystander from the video lineup, suggesting there may have been a ceiling effect. A few (5%)

showup non-choosers selected the innocent bystander from the video lineup, although this did not significantly differ from the controls and therefore it is not possible to conclude that this provided evidence of source confusion. However, again, a commitment effect was demonstrated in that 65% of those making an incorrect identification decision in the live showup trials made the same mistaken identification of the innocent bystander in the lineup. Somewhat surprisingly, when measuring the fairness of the video lineup, over a third of pilot participants (35.2%) when given a written description of the loud actor, selected a photograph of the innocent bystander from the nine lineup members. This suggests that the lineup was biased, or that the innocent bystander stood out from the eight foils. However, a comparison of the actors shown in Figure 4 demonstrates the physical differences between them. Furthermore, even though more than a quarter of the control participants selected a foil from this lineup (26.8%), none selected the innocent bystander, perhaps an indication that some individuating aspects of facial appearance cannot easily be described.

General discussion

The aim of the experiments reported here was to provide an evidence base to address issues that arise from the use of live showups (street identifications) in criminal investigations. British police frequently use live showups, which have a strong influence on the outcome of cases (Davis *et al*, submitted). The procedure used in England and Wales was specifically addressed; live showups and video lineups were run in accordance with the PACE (1984; Code of Practice, 2008). Video lineups were compiled by the police using the same national databases used for criminal investigations. The lineups used in these experiments were identical in every respect to the lineups that would have been used had the staged incidents been real crimes. The fairness of culprit-present lineups was assessed for information, but the lineups were not altered in order to retain the authenticity of the lineups constructed by the police. In line with legal procedure, the sequential video lineups were

viewed twice by each witness before any identification could be made. Dysart and Lindsay (2007a) and Wolchover and Heaton-Armstrong (2004) have argued that showups are inherently suggestive. Therefore, the first question addressed was whether a (video) lineup should be arranged instead of a live showup, to afford protection for an innocent suspect.

The reliability of identification from showups and lineups compared.

In the initial identification procedures of all experiments the highest proportion of selections from culprit-present trials, whether accurate or not, were made from video lineups. The participants made a selection from 81.2% of all video lineups in the control condition; selections from 22.6% of lineups were mistaken identifications of a foil, and from 58.6% of lineups were correct identifications of the culprit. In real cases, a mistaken identification of a foil does not lead to a wrongful conviction; it serves only to demonstrate that the witness is unreliable. Therefore, for culprit-present trials the rate of correct identification is of most forensic relevance.

In both Experiment 1 and 3 the culprit was correctly identified in approximately 50% of live culprit-present showups held a few minutes after a live, staged incident. A meta-analysis of earlier research found that the rate of selection is lower in showups than in lineups, but result in similar rates of correct identifications in culprit-present trials (Stebly *et al.*, 2003). Collapsed across Experiments 1-3 the results are consistent with this conclusion. However, it is important to note considerable differences across individual culprits in different Experiments. In Experiment 3 there was no difference in accuracy between the live showups and the control video lineups, despite the latter being conducted at least one day later. However, in Experiment 1 accuracy was higher in the video lineups (72%) than in the live showups (51%). In contrast, accuracy in Experiment 2 was highest in the live showups (80%) and worst in the video lineups (52%). The effect observed in Experiment 2 is probably due to the enhanced gait cues of a distinctive “swagger” when the culprit was seen live. The

interpretation of the effect in Experiment 1 is not clear. In summary, there is no consistent evidence that the accuracy of culprit identification differs between a live showup and a video lineup. In Experiments 1 and 2 video showups were included to allow the effect of the test medium (video vs. live) to be distinguished from the identification task (showup vs. lineup). However, no consistent effect of the test media used for showups was found. Live presentation of a showup yielded better accuracy than video in Experiment 2 but worse accuracy in Experiment 1 as discussed above.

The greatest concern about the use of showups in criminal investigations is of potential misidentification of innocent suspects. Culprit-absent trials are relevant to this issue. In Experiments 1, 3 and 4 culprit-absent trials were conducted using similar procedures. Accuracy was high in rejecting the innocent suspect in Experiment 1 (95% correct) and the innocent bystander in Experiment 4 (92%). Nevertheless even a 5 – 8% error rate is of concern in practice because all of these identifications would be mistaken identifications of the police suspect. In Experiment 1, the rate of mistaken identification of the innocent suspect did not differ between live showups and control video lineups. The same result was found in Experiment 4 in which the suspect was an innocent bystander in the incident. But in both Experiments 1 and 4 the innocent suspects did not closely resemble the culprits. These results suggest that there should not be a great cause for concern if an innocent suspect who does not closely resemble the culprit becomes the focus of a live showup. In contrast, the innocent suspect did closely resemble the culprit in Experiment 3 and the results were very different. The proportion of mistaken identifications at the first procedure was similar for showups (36%) and video lineups (43%), but was much higher than the rates observed in Experiments 1 and 4 (5 – 10%). These results are a cause for concern. Error rates might be even higher under more difficult viewing conditions that may occur for a real victim of crime. In a real

case, the innocent suspect will have been detained because they fit the description of the culprit. It would be unknown how closely the suspect resembles the actual culprit.

In summary, across all experiments, the rate of mistaken identification of innocent suspects did not differ significantly as a function of the identification procedure. Thus, the data suggest that a showup is no less fair to an innocent suspect than a video lineup. All of the identification procedures run in the experiments reported here complied with current English law and were identical to the procedures that would have been used in real crimes. The results demonstrate that eyewitness identification, whether from a showup or a video lineup, is often unreliable even under the ideal conditions of a good view and a delay of only minutes before the identification procedure. The data illustrate the risk of basing prosecutions on uncorroborated identification evidence.

Repeated identification procedures.

In Experiments 1, 3 and 4, the actors who were the targets in a showup were later presented to the same participants as a member of a video lineup. In all of the experiments, an effect of commitment was found. Most showup choosers made an identification from the lineup (95%), the majority selected the same suspect from a video lineup (88%), regardless of whether the suspect was the culprit or not. These results are consistent with real cases in which 82% of suspects identified in a showup were also positively identified in a video lineup by the same witness (Davis *et al.* submitted).

In the present study, the majority of showup non-choosers rejected the lineup (61%). Some of the showup non-choosers did identify the suspect from the lineups. In culprit-present conditions, 43% changed their decisions from showup to lineup to make a correct culprit selection. In culprit-absent conditions, 16% changed their decisions to make an innocent suspect selection. Similar carryover effects have been found in previous research (e.g., Deffenbacher *et al.*, 2006; Dysart *et al.*, 2001; Haw *et al.*, 2007; Memon *et al.*, 2002).

Identifications of the innocent suspect from the lineup by showup non-choosers in culprit-absent trials have been interpreted as evidence of source confusion between a previous identification procedure and the original incident. Deffenbacher *et al.* interpreted the results of their meta-analysis as providing evidence of misattribution of familiarity arising from confusion of the source of the memory. However, this interpretation has been challenged by Blunt and McAllister (2009) and Goodsell *et al.* (2009), who found evidence of a commitment effect but no effect of source confusion. These authors suggested that Deffenbacher *et al.*'s meta-analysis of source confusion may have been contaminated by failure to appropriately exclude the commitment effect from some studies. The experiments reported here did not show a significant effect of source confusion and therefore support the view that commitment, but not source confusion, from a prior identification procedure is a substantial cause of eyewitness error. Our results are consistent with the position taken by Blunt and McAllister (2009) and Goodsell *et al.* (2009). The data reported here show only a small, and statistically non-significant, effect size of prior exposure to a suspect who was not identified.

In real cases, a person identified in a showup is likely to be arrested. In such cases it could be argued from a defence point of view, that it might be advisable to request a subsequent video lineup because the identification is not repeated in all cases. An innocent suspect not identified in a showup is unlikely to be arrested at that time. However, additional evidence may become available at a later date and they may become the focus of a video lineup. The results of these experiments do not suggest that there would be a significantly enhanced likelihood of an incorrect lineup identification decision in these circumstances.

Delay.

In Experiment 3, the subsequent video lineups were held at least 24 hours after the showups. Slightly more selections of the suspect were made from both culprit-present and

culprit-absent lineups conducted within a week, than lineups held between a week and a month later but the effect size was small and not statistically significant in the experiment reported.

Clothing bias.

Sometimes suspects exchange or claim to have exchanged clothing after an offence has been committed. No effect of clothing bias from a dissimilar-looking innocent suspect wearing clothing worn by the culprit was found on the accuracy of identification in Experiment 4, but he was identified with more confidence if he was wearing the same clothing as the culprit wore during the incident. The innocent bystander tended to be selected more often in a showup if he had been wearing distinctive clothing during the incident, regardless of what was worn during the identification procedure.

Confidence

Analysis of confidence ratings were generally consistent with greater confidence of accurate than of inaccurate responses; greater confidence amongst choosers than amongst non-choosers and greater confidence amongst people who had seen the suspect in a prior showup than those who had not. Confidence expressed immediately after an identification procedure was associated with accuracy of response. Predictive confidence measured immediately after the incident was correlated with confidence measures taken after the identification procedure, but showed a weaker association with identification accuracy. These results are consistent with effects that have been reported previously in the literature.

Implications for policy

This research has examined the reliability of a live showup procedure, using the current codes of practice employed in the UK (PACE, 1984; Codes of Practice, 2008), and measured its influence on a subsequent video lineup. Some criticism could be directed at the procedures used for the showups in these experiments as the codes of practice specifically

warn against directing a witness's attention towards the suspect. However, the procedures used do comply with the code of practice which allows attention to be drawn to the suspect if it is not practical to avoid doing so.

The culprit-present identification rates in the showups were surprisingly low. Based on the results of the culprit-present trials it could be argued that unless the actual culprit possesses some particularly distinctive attribute such as the gait of the actor in Experiment 2, a correct identification of the culprit from a live showup is less likely than from a video lineup. Furthermore, the rate of mistaken identification of innocent suspects from live showups was no higher than the rate from video lineups. On the basis of the research literature and the experiments reported here we make two recommendations for good practice. First, repeated identification procedures should be avoided if they involve repeated presentation of a suspect who has already been identified by the witness concerned. Repeat identification in these circumstances run a high risk of repeating a mistaken identification if the witness has made a mistake during the first identification procedure. We acknowledge that conducting a second procedure may present a witness, who perhaps yielded to the suggestiveness of a showup to make an identification in the street, with an opportunity to 'nullify' the earlier identification by declining to identify the same suspect again. However, our data suggest that the protection offered by the safeguard of a second identification, is outweighed by the risk of a repeated mistaken identification.

Second, we recommend that a measure of confidence be taken immediately after any identification decision has been made before there is any opportunity for the witness to receive feedback. Confidence measured at this point is associated with identification accuracy. Confidence may provide some probative information but it needs to be used with caution. There are large individual differences: Confident but mistaken witnesses will be encountered.

One caution on the practical use of live showups for street crimes should be made. In the past a showup has been attempted with more than one suspect in a group. There is some evidence that this type of procedure may occur in up to a third of city centre showup attempts (Davis *et al.*, submitted). Each suspect acts as a foil to the others. In effect such a procedure is an all-suspect lineup, found empirically to increase the likelihood of misidentifications (Wells & Turtle, 1986). The experiments reported here showed robust evidence of greater choosing from lineups than live showups, but the higher choosing rate did not result in more identifications of the single innocent suspect because a substantial proportion of choices were of an innocent foil. By extension these data provide strong evidence that an all-suspect identification procedure, such as identification of suspects in a group on the street, should never be used because the risk of mistaken identification is substantially inflated.

References

- Behrman, B. W., & Davey, S. L. (2001). Eyewitness identification in actual criminal cases: An archival analysis. *Law and Human Behavior, 25*, 475-491.
- Blunt, M. R. & McAllister, H. A. (2009). Mug shot exposure effects: Does Size matter? *Law and Human Behavior, 33*, 175-182.
- Bogan, P. (2004). *Identification: investigation, trial and scientific evidence*. London: Legal Action Group.
- Brown, E., Deffenbacher, K. & Sturgill, W. (1977). Memory for faces and the circumstances of encounter. *Journal of Applied Psychology, 62*, 311-318.
- Brigham, J. C. & Cairns, D. L. (1988). The effect of mugshot inspections on eyewitness identification accuracy. *Journal of Applied Social Psychology, 18*, 1394-1410.
- Buckhout, R. (1974). Eyewitness testimony. *Scientific American, 231*, 23-31.
- Cutler, B.L., & Fisher, R.P. (1990). Live lineups, videotaped lineups, and photo-arrays. *Forensic Reports, 3*, 439 – 448.

Cutler, B.L., Berman, G.L., Penrod, S., & 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.), In *Adult Eyewitness Testimony: Current Trends and developments* (pp. 163 – 182). Cambridge: Cambridge University Press.

Cutler, B.L. & Penrod, S.D. (1988). Improving the reliability of eyewitness identifications: Lineup construction and presentation. *Journal of Applied Psychology*, 73, 281-290.

Cutler, B.L., Penrod, S.D., O'Rourke, T.E., & Martens, T.K. (1987). Improving the reliability of eyewitness identification: Putting context into context. *Social Behavior*, 1, 113-134.

Cutting, J.E., & Kozlowski, L.T. (1977). Recognising friends by their walk: Gait perception without familiarity cues. *Bulletin of the Psychonomic Society*, 9, 353 – 356.

Davis, J. P., Valentine, T., Memon, A. & Roberts, A. (submitted). Identification on the street: A comparison with other methods of eyewitness identification.

Deffenbacher, K.A., Bornstein, B.H., McGorty, E. K. & Penrod, S.D. (2008). Forgetting the once-seen face: Estimating the strength of an eyewitness's memory representation. *Journal of Experimental Psychology: Applied*, 14, 139-150.

Deffenbacher, K.A., Bornstein, B.H., & Penrod, S.D. (2006). Mugshot exposure effects: Retroactive interference, mugshot commitment, source confusion, and unconscious transference. *Law and Human Behavior*, 30, 287-307.

Devlin. P. (1976). *Report to the Secretary of State for the Home Departmental Committee on Evidence of Identification in Criminal Cases*, HC 338, London: HMSO.

Dysart, J.E., Lindsay, R.C.L., Hammond, R., & Dupuis, P.R. (2001). Mug shot exposure prior to lineup identification: Interference, transference, and commitment effects. *Journal of Applied Psychology*, 86, 1280-1284.

Dysart, J., Lindsay, R.C.L. & Dupuis, P.R. (2006). Show-ups: The critical issue of clothing bias. *Applied Cognitive Psychology, 20*, 1009-1023.

Dysart, J. E., & Lindsay, R. C. L. (2007a). Show-up identifications: Suggestive technique or reliable method? In R. C. L. Lindsay, D. R. Ross, J. D. Read, & M. P. Toglia (Eds.), *The handbook of eyewitness psychology, Vol. II, Memory for People* (pp. 137-153). Mahwah, NJ: Lawrence Erlbaum.

Dysart, J. E., & Lindsay, R. C. L. (2007b). The effect of delay on eyewitness identification accuracy: should we be concerned? In R. C. L. Lindsay, D. R. Ross, J. D. Read, & M. P. Toglia (Eds.), *The handbook of eyewitness psychology, Vol. II, Memory for People* (pp. 361-376). Mahwah, NJ: Lawrence Erlbaum.

Earles, J. L., Kesten, A. W., Curtayne, E. S. & Perle, J. G. (2008). That's the man who did it, or was it a woman? Actor similarity and binding errors in event memory. *Psychonomic Bulletin & Review, 15*, 1185-89.

Egan, D., Pittner, M., & Goldstein, A.G. (1977). Eyewitness identification: Photographs vs. live models. *Law and Human Behavior, 1*, 199 - 206.

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.

Godfrey, R. D. & Clark, S. E. (2010). Repeated eyewitness identification procedures: Memory, decision making and probative value. *Law and Human Behavior, 34*, 241-258.

Gonzales, R., Ellsworth, P., & Pembroke, M. (1993). Response biases in lineups and showups. *Journal of Personality and Social Psychology, 64*, 525-537.

Goodsell, C. A., Neuschatz J. S. & Gronlund, S. D. (2009). Effects of mugshot commitment on lineup performance in young and older adults. *Applied Cognitive Psychology, 23*, 788-803.

Gorenstein, G.W., & Ellsworth, P.C. (1980). Effect of choosing an incorrect photograph on a later identification by an eyewitness. *Journal of Applied Psychology*, 65, 616-622.

Haw, R.M., Dickinson, J.J., & Meissner, C.A. (2007). The phenomenology of carryover effects between show-up and line-up identification. *Memory*, 15, 117 -127.

Hinz, T., and Pezdek, K. (2001). The effect of exposure to multiple lineups on face identification accuracy. *Law and Human Behavior*, 25, 185-198.

Innocence Project (2011). <http://www.innocenceproject.org/understand/> Retrieved 29th January 2011.

Johnson, M. K., Hashtroudi, J. & Lindsay, D. S. (1993). Source Monitoring. *Psychological Bulletin*, 114, 3-28.

Laughery, K.R., Fessler, P.K., Lenorovitz, D.R., & Yoblick, D.A. (1974). Time delay and similarity effects in facial recognition. *Journal of Applied Psychology*, 59, 490-496.

Lindsay, R.C.L., Wallbridge, H., & Drennan, D. (1987). Do the clothes make the man? An exploration of the effect of lineup attire on eyewitness identification accuracy. *Canadian Journal of Behavioural Science*, 19, 463-478.

Mauldin, M.A., & Laughery, K.R. (1981). Composite production effects of subsequent facial recognition. *Journal of Applied Psychology*, 66, 351-357.

Memon, A., Hope, L., Bartlett, J., & Bull, R. (2002). Eyewitness recognition errors: The effects of mugshot viewing and choosing in young and old adults. *Memory & Cognition*, 30, 1219-1227.

Naughton, M. (2003). How big is the 'iceberg'? A zemiological approach to quantifying miscarriages of justice. *Radical Statistics*, 81, 5 -17.

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

Pezdek, K., & Blandon-Gitlin, I. (2005). When is an intervening line-up most likely to affect eyewitness identification accuracy? *Legal & Criminological Psychology, 10*, 247-263.

Police and Criminal Evidence Act (1984) Codes of Practice, Code D (2008).

Retrieved 29th January 2010 from:

<http://www.homeoffice.gov.uk/publications/police/operational-policing/pace-codes/pace-code-d?view=Binary>

R v Forbes (2001). 1 All ER 686.

Ross, D. E., Marsil, D. F., Benton, T. R., Hoffman, R., Warren, A. R., Lindsay, R. C. L. & Metzger, R. (2006). Children's Susceptibility to misidentifying a familiar bystander from a lineup: When younger is better. *Law and Human Behavior, 30*, 249-257.

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

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.

Stebly, N., Dysart, J., Fulero, S. & Lindsay, R.C.L. (2003). Eyewitness accuracy rates in police showup and lineup presentations: A meta-analytic comparison. *Law and Human Behavior, 27*, 523- 540.

Stevenage, S.V., Nixon, M.S., & Vince, K. (1999). Visual analysis of gait as a cue to identity. *Applied Cognitive Psychology, 13*, 513 – 526.

Tulving, E., & Thomson, D. (1973). Encoding specificity and retrieval processes in memory. *Psychological Review, 80*, 352-373.

Valentine, T. (2006). Forensic facial identification. In: Heaton-Armstrong, A., Shepherd, E., Gudjonsson, G. & Wolchover, D. (eds). *Witness Testimony; Psychological, Investigative and Evidential Perspectives*. Oxford: Oxford University Press. (pp. 281-307.)

Valentine, T., Pickering, A., & Darling, S. (2003). Characteristics of eyewitness identification that predict the outcome of real lineups. *Applied Cognitive Psychology, 17*, 969-993.

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., 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.

Wells, G.L., & Turtle, J.W. (1986). Eyewitness identification: The importance of lineup models. *Psychological Bulletin, 99*, 320-329.

Wolchover, & D. Heaton-Armstrong, A. (2004). Ending the farce of staged street identifications. *Archbold News, 3*, 5-7.

Yarmey, A.D., Yarmey, M.J. & Yarmey, A.L. (1996) Accuracy of eyewitness identification in show-ups and line-ups. *Law and Human Behavior, 20*, 459-477

Table 1: Number of participants and percentage of each identification performance outcome in Experiment 1 showups and video lineups as a function of culprit-presence, condition, and for participants taking part in a second identification trial, their decision to the showup

Culprit-present									
Condition		Total		Suspect ID		Foil ID		Rejection	
		n		n	%	n	%	n	%
First or only identification decisions: Showups and control video lineups									
Live showup		47		24	51.1	N/A		23	48.9
Video showup		48		38	79.2	N/A		10	20.8
Control video lineup		47		34	72.3	5	10.7	8	17.0
Second identification decisions (indicating previous decision): Video lineups									
Live showup	Chooser	24		22	91.7	2	8.3	0	0.0
	Non-chooser	23		4	17.4	3	13.0	16	69.6
Video showup	Chooser	38		36	94.7	2	5.3	-	-
	Non-chooser	10		-	-	1	10.0	9	90.0
Culprit-absent									
First or only identification decisions: Showups and control video lineups									
Live showup		48		2	4.2	N/A		46	95.8
Video showup		46		3	6.5	N/A		43	93.5
Control video lineup		47		2	4.3	18	38.3	27	57.4
Second identification decisions (indicating previous decision): Video lineups									
Live showup	Chooser	2		2	100.0	-	-	-	-
	Non-chooser	46		7	15.2	9	19.6	30	65.2
Video showup	Chooser	3		2	66.7	1	33.3	-	-
	Non-chooser	43		4	9.3	14	32.6	25	58.1

Table 2: Mean confidence (max = 100%) in Experiment 1 as a function of culprit-presence, condition, identification phase and for participants making a second decision, previous showup selection (SD in parentheses).

Culprit-present					
Condition		n	Suspect ID	Foil ID	Rejection
First or only identification decisions: Showups and control video lineups					
Live showup		47	77.5 (17.0)	N/A	58.7 (21.2)
Video showup		48	75.8 (19.3)	N/A	53.3 (20.8)
Control video lineup		47	70.9 (19.8)	52.0 (23.9)	55.0 (31.2)
Second identification decisions (indicating previous decision): Video lineups					
Live Showup	Choosers	24	79.6 (17.9)	40.0 (42.4)	-
	Non-choosers	23	72.5 (12.6)	46.1 (15.3)	38.8 (20.6)
Video Showup	Choosers	38	82.2 (15.7)	60.0 (14.1)	-
	Non-choosers	10	-	40.0 (-)	56.7 (20.6)
Culprit-absent					
First or only identification decisions: Showups and control video lineups					
Live showup		48	55.0 (21.2)	N/A	86.5 (14.9)
Video showup		46	50.0 (11.5)	N/A	67.7 (19.6)
Control video lineup		47	70.0 (14.1)	62.8 (19.9)	67.5 (28.1)
Second identification decisions (indicating previous decision): Video lineups					
Live Showup	Choosers	2	80.0 (28.3)	-	-
	Non-choosers	46	60.0 (14.1)	57.8 (18.6)	69.0 (25.5)
Video Showup	Choosers	3	80.0 (28.3)	50.0 (-)	-
	Non-choosers	43	62.5 (20.6)	54.3 (23.1)	69.6 (23.9)

Table 3: Number of participants and percentage of each performance outcome in Experiment 2 as a function of identification procedure.

Condition	Culprit-present						
	Total	Suspect ID		Foil ID		Rejection	
	n	n	%	n	%	n	%
Live showup	49	39	79.6	N/A		10	20.4
Video showup	58	38	65.5	N/A		20	34.5
Control video lineup	46	24	52.2	13	28.3	9	19.5

Table 4: Number of participants and mean confidence (max = 100%) in Experiment 2 as a function of identification procedure (SD in parentheses).

Condition	n	Suspect ID	Foil ID	Rejection
Live showup	49	81.0 (19.4)	N/A	69.0 (20.2)
Video showup	58	76.8 (19.0)	N/A	54.0 (23.5)
Control video lineup	46	74.6 (17.4)	60.8 (15.5)	54.4 (29.2)

Table 5: Number of participants and percentage of each performance outcome of identification procedures in Experiment 3 as a function of culprit-presence, condition and, for participants taking part in a second identification trial, the prior showup decision.

Culprit-present								
Condition	Total		Suspect ID		Foil ID		Rejection	
	n		n	%	n	%	n	%
First or only identification decisions: Showups and control video lineups								
Live showup	125		57	45.6	N/A		68	54.4
Video lineup < 1 week	50		31	62.0	10	20.0	9	18.0
Video lineup > 1 week	43		20	46.5	14	32.6	9	20.9
Second identification decisions (indicating previous decision and delay): Video lineups								
Choosers	< 1 week	25	22	88.0	2	8.0	1	4.0
	> 1 week	20	20	100.0	-	-	-	-
Non-choosers	< 1 week	26	19	73.1	1	3.8	6	23.1
	> 1 week	25	13	52.0	3	12.0	9	36.0
Culprit-absent								
First or only identification decisions: Showups and control video lineups								
Live showup	92		33	35.9	N/A		59	64.1
Video lineup < 1 week	51		22	43.1	9	17.6	20	39.2
Video lineup > 1 week	45		19	42.2	11	24.4	15	33.3
Second identification decisions (indicating previous decision and delay): Video lineups								
Choosers	< 1 week	12	10	83.3	1	8.3	1	8.3
	> 1 week	16	14	87.5	2	12.5	-	-
Non-choosers	< 1 week	25	17	68.0	-	-	8	32.0
	> 1 week	30	16	53.3	-	-	14	46.7

Table 6: Number of participants and mean confidence levels (max = 100%) in Experiment 3 as a function of culprit-presence, condition, identification phase and for participants making a second decision, previous showup selection (SD in parentheses).

Culprit-present					
Condition	n	Suspect ID	Foil ID	Rejection	
First or only identification decisions: Showups and control video lineups					
Live showup	125	76.7 (17.7)	N/A	60.3 (25.6)	
Control video lineup < 1 week	50	75.2 (22.2)	46.0 (27.6)	53.3 (33.2)	
Control video lineup > 1 week	43	74.5 (15.4)	57.1 (17.29)	64.4 (14.2)	
Second identification decisions (indicating previous decision and delay): Video lineups					
Choosers	< 1 week	25	86.8 (10.9)	85.0 (7.1)	50.0 (-)
	> 1 week	20	76.0 (19.0)	-	-
Non-choosers	< 1 week	26	74.7(17.8)	30.0 (-)	61.7 (23.2)
	> 1 week	25	64.6 (17.1)	46.7 (15.3)	72.2 (23.9)
Culprit-absent					
First or only identification decisions: Showups and control video lineups					
Live showup	92	67.6 (15.6)	N/A	58.8 (25.0)	
Control video lineup < 1 week	51	70.0 (16.9)	62.2 (15.6)	53.0 (30.5)	
Control video lineup > 1 week	45	65.8 (17.1)	53.6 (22.1)	57.3 (21.9)	
Second identification decisions (indicating previous decision and delay): Video lineups					
Choosers	< 1 week	12	81.0 (12.9)	70.0 (-)	100.0 (-)
	> 1 week	16	74.3 (18.7)	45.0 (21.2)	-
Non-choosers	< 1 week	25	74.7 (18.4)	-	55.0 (22.7)
	> 1 week	30	65.6 (20.0)	-	70.7 (30.3)

Table 7: Number of participants and percentage of each outcome in Experiment 4 of culprit-absent identification procedures as a function of innocent bystander clothing in the staged incident and whether this was the same or had been exchanged with the protagonist for a showup

Condition		Total	Suspect ID		Foil ID		Rejection	
Incident clothing	Showup clothing	n	n	%	n	%	n	%
First or only identification decisions: Showups (indicating clothing condition) and control video lineups								
	Same	46	7	15.2	N/A		39	84.8
Distinctive	Changed	40	4	10.0	N/A		36	90.0
	Control lineup	29	-	-	7	24.1	22	75.9
Plain	Same	66	2	3.0	N/A		64	97.0
	Changed	68	4	5.9	N/A		64	94.1
	Control lineup	27	-	-	8	29.6	19	70.4
Second identification decisions (indicating previous decision and clothing condition): Video lineups								
Choosers								
	Same	7	5	71.4	-	-	2	28.6
Distinctive	Changed	4	3	75.0	-	-	1	25.0
	Same	2	1	50.0	1	50.0	-	-
Plain	Changed	4	2	50.0	-	-	2	50.0
Non-choosers								
	Same	39	1	2.6	11	28.2	27	69.2
Distinctive	Changed	36	3	8.3	6	16.7	27	75.0
	Same	64	3	4.7	17	26.6	44	68.8
Plain	Changed	64	3	4.7	10	15.6	51	79.7

Table 8: Number of participants and mean confidence levels from Experiment 4 as a function of condition, identification decision; incident clothing and clothing change (standard deviations in parentheses)

Culprit-absent					
First or only identification decisions: Showups (indicating clothing condition) and control video lineups					
Incident Clothing	Showup clothing	n	Bystander	Foil	Rejection
	Same	46	67.1 (17.0)	N/A	90.8 (15.5)
Distinctive	Changed	40	65.0 (12.9)	N/A	72.5 (30.1)
	Control lineup	29	-	55.7 (21.5)	61.8 (24.4)
	Same	66	40.0 (14.1)	N/A	82.8 (23.8)
Plain	Changed	68	55.0 (23.8)	N/A	84.8 (17.5)
	Control lineup	27	-	46.3 (19.2)	67.9 (25.7)
Second identification decisions (indicating previous decision and clothing condition): Video lineups					
Showup choosers					
	Same	7	80.0 (12.3)	-	55.0 (7.1)
Distinctive	Changed	4	66.7 (5.8)	-	70.0 (-)
	Same	2	40.0 (-)	60.0 (-)	-
Plain	Changed	4	80.0 (-)	-	35.0 (49.5)
Showup non-choosers					
	Same	39	40.0 (-)	55.5 (14.4)	53.9 (19.4)
Distinctive	Changed	36	50.0 (17.3)	45.0 (16.4)	53.0 (24.6)
	Same	64	46.7 (20.8)	52.4 (19.9)	63.9 (26.8)
Plain	Changed	64	66.7 (23.1)	53.0 (17.7)	60.0 (26.2)

Appendix 1

Table A1: Mean incident realism and arousal data from each experiment

		Experiment 1	Experiment 2	Experiment 3	Experiment 4	Overall
		Mean (SD)	Mean (SD)	Mean (SD)	Mean (SD)	
How suspicious	0-100%	7.36 (1.47)	13.08 (23.83)	17.15 (26.28)	12.53 (24.04)	14.37
were you?						(25.27)
How convinced	0-100%	83.57 (27.87)	73.40 (33.12)	69.75 (32.41)	69.23 (34.49)	73.38
were you?						(32.53)
Realization event	1-8*	8	8	8	8	8
staged? (mode)						
I felt anxious	1-4	1.23 (0.53)	1.23 (0.48)	1.31 (0.62)	1.36 (0.61)	1.30 (0.59)
I felt confused	1-4	1.48 (0.77)	1.17 (0.46)	1.37 (0.68)	1.96 (0.84)	1.50 (0.77)
I felt angry	1-4	1.06 (0.33)	1.11 (0.44)	1.03 (0.21)	1.14 (0.42)	1.08 (0.33)
I felt suspicious	1-4	1.30 (0.71)	1.37 (0.75)	1.51 (0.81)	1.80 (0.96)	1.51 (0.84)
I felt surprised	1-4	1.66 (1.02)	1.40 (0.88)	1.56 (0.97)	2.53 (0.93)	1.79 (1.05)
I felt aroused	1-4	1.22 (0.57)	1.10 (0.32)	1.29 (0.64)	1.39 (0.65)	1.27 (0.60)

1. Before student entered; 2. When student entered, 3. When student spoke; 4. When student complained; 5. While lecturer spoke; 6. When student left; 7. Before informed of act; 8. At end (once informed act was staged)

Figures



Figure 1: Innocent suspect (left) and culprit (right) actresses in Experiment 1. Note: the innocent suspect is wearing the clothing both actresses wore in the Phase 1 street identification and video showup trial. The culprit is wearing the clothes she wore during the staged plagiarism incident.



Figure 2: Innocent suspect (left) and culprit (right) actresses in Experiment 3. Note: the innocent suspect is wearing the top both actresses wore in the Phase 1 street identification trial. The culprit is wearing the jacket she wore during the staged plagiarism incident.



Figure 3: Photograph of culprit with hair styled as in street identification trial.



Figure 4: Experiment 4 Innocent suspect (left - wearing distinctive top) and culprit (wearing plain top)