Applied Cognition Psychology, **13**, S59-S72 (1999)

An evaluation of the fairness of police line-ups and video identifications.

Tim Valentine and Pamela Heaton
Goldsmiths College, University of London,

Running head: Line-ups and video identifications

Address for correspondence: Professor Tim Valentine,
Department of Psychology,
Goldsmiths College,
University of London,
New Cross
London SE14 6NW

email: t.valentine@gold.ac.uk

Acknowledgements

This research was supported by a project grant from the Nuffield Foundation. We gratefully acknowledge the contribution of Deputy Chief Constable Edward Marchant and Inspector Les Hall (Durham Constabulary) and of Dr Raymond Petre and Linda Baillie (West Yorkshire Police).

Abstract

Mistaken eyewitness identification is a major source of miscarriages of justice. In England and Wales, procedures for obtaining identification evidence are set out in legislation. The vast majority of identifications are obtained using a traditional 'live' identity parade (or line-up). However, in some circumstances video identifications are being used more frequently. Records of line-ups and video identifications used in actual criminal cases were obtained. The fairness of these two procedures was compared by use of a mock witness procedure. In a perfectly fair line-up the suspect would be chosen, by chance, by 11% of the mock witnesses. However, twenty-five percent of mock witnesses selected the suspect from 25 photographs of live line-ups, compared to 15% of mock witnesses who selected the suspect from video identifications. An analysis of covariance, taking the number of visual features mentioned in the original witness's first description as the covariate, showed that the proportion choosing the suspect was significantly smaller from video identifications. It is concluded that the video line-ups were fairer than the live-ups, and therefore that wider use of video identifications has the potential to improve the reliability of eyewitness identification evidence.

Introduction.

Analysis of known cases of wrongful imprisonment have repeatedly suggested that mistaken eyewitness identification is a major factor in miscarriages of justice (e.g. Brandon & Davies, 1973; Huff; Rattner & Sagarin, 1986). Huff et. al. (1986) found that nearly 60% of a database of approximately 500 cases of known wrongful convictions, mainly from the USA, involved errors of eyewitness identification. Wells, Small, Penrod, Malpass, Fulero and Brimacombe (1998) report 40 cases from the United States (including 28 cases previously reported by Connors, Lundregan, Miller & McEwan, 1996) in which DNA evidence that was not available at the time of conviction has subsequently resulted in the exoneration of people who were wrongfully convicted of a crime. Ninety percent (36) of the cases involved eyewitness misidentification. One person was falsely identified by 5 eyewitnesses, seven people (17.5% of the sample) were wrongly identified by more than one person.

The fallibility of eyewitness identification has been recognised in England for some time. For example, Devlin (1976) pointed out the problems associated with identification evidence in a report to the Home Secretary. At about the same time a judgement in the case of R v Turnbull and others (1976) produced the 'Turnbull Guidelines' that aimed to distinguish between 'good' and 'poor' identification evidence. Although guidelines on the procedures for collecting identification evidence were issued in England by the Home Office as early as 1925, the first statute to cover this area was not passed until 1984 (Police and Criminal Evidence Act 1984, henceforth referred to as PACE). Procedures for the conduct of identification attempts are set out in a 'code of practice'. The current code came into force in 1995. Despite the development of legislation affecting identification evidence it remains the case that in England and Wales conviction can be secured on the uncorroborated evidence of a single eyewitness. See Davies (1996) for discussion of this point.

The vast majority of eyewitness identifications in England are obtained by asking a witness to view a 'live' line-up. (This procedure is referred to as an 'identity parade' in English legislation.). The use of identification suites in which the lineup is usually viewed by the witness from behind a one-way mirror is increasingly common, but is by no means universal. Many line-ups are staged in police stations without the use of one-way mirrors. The code of practice requires that the line-up comprises at least 8 volunteers who 'so far as possible resemble the suspect in age, height, general appearance and position in life'. The witness must be advised that the culprit 'may or may not be on the parade and if he (sic) cannot make a positive identification he should say so but that he should not make a decision before looking at each member of the parade at least twice' (Code D of PACE, annex A, paragraph 14). It is interesting to note that despite the warning that the culprit may not be present, identification of line-up volunteers is relatively common in line-ups conducted under the PACE code of practice. Slater (1994) analysed the outcome of identification attempts by 843 witnesses who viewed 302 suspects, and found that a foil was identified by 190 witnesses (22.5%). Wright and McDaid (1996) analysed the outcome of identification attempts made by 1561 witnesses who viewed 616 suspects, and found that 310 witnesses picked a foil (19.9%).

PACE allows four means of identification to be used; an identity parade (line-up); a group identification (in which the suspect is viewed amongst an informal group of people); a video identification; or a confrontation. Use of a photospread (or photo line-up) to obtain identification evidence is not permitted. Under PACE, whenever a suspect disputes an identification a line-up must be held, if the suspect consents, unless it is not practicable to do so. For example, it may not be practicable to assemble sufficient people who resemble the suspect if (s)he has an unusual appearance. If it is not practicable to organise a line-up, a group or video identification can be held if it is the most satisfactory course of action in the circumstances. A confrontation can only be used if none of the other means of identification are practicable.

As PACE requires that the suspect must be offered the opportunity to stand in a line-up in all cases of disputed identification, there has been a sharp rise in the number of line-ups held in England and Wales. Slater (1994) reports that the number of line-ups held in 49 of the 52 police forces in the United Kingdom rose from 1,782 in 1990 to 6,853 in 1992 (284% increase). The projected total for 1993 was 13,652. Slater estimated the total cost of line-ups held in the UK in 1994 to be £14 million.

The rising number of line-ups held in the UK has led to an interest in developing video as the identification procedure of choice. Video offers a number of advantages. A large video library can be set-up which may allow the construction of a fairer line-up than would be possible from a smaller selection of volunteers who are available to stand in a 'live' line-up at a specific time and location. Costs can be controlled because it is not necessary to pay a volunteer every time their image is used in a video line-up. Furthermore, the cost of parades that have to be cancelled and re-scheduled is reduced. A common cause of cancelling a parade is that a suspect who has been bailed fails to keep an appointment for a line-up. Use of video eliminates this cost entirely. Video is a less threatening means of identification for victims of violent crime or in cases of witness intimidation. Use of video eliminates the need for the suspect and the witness to be present at the same location simultaneously. Ainsworth and King (1988) found that witnesses attending real identification parades who admitted to being nervous were significantly less likely to identify the suspect than witness who did not report being nervous. Therefore, any steps that reduce anxiety could improve the effectiveness of identification attempts, provided they did not increase the rate of false positives. It should be noted that PACE affords a legal representative of the suspect the right to be present when a witness attempts a video identification.

The present research has two aims. The first is to evaluate the fairness of archive records of live line-ups using a mock-witness procedure. This research will provide data to evaluate the effectiveness of the present PACE code of practice in producing fair and unbiased line-ups. The second aim is to compare the fairness of video identifications with the traditional live line-up. Thus, the research will provide some data to establish whether the practical benefits of video can also improve the quality of the identification evidence obtained.

Doob and Kirshenbaum (1973) introduced a paradigm to measure the fairness of a line-up based upon assessments by 'mock witnesses'. A mock witness is somebody who was not present at the crime scene. The participant is given the description that the witness gave and is asked to pick the person that they think is most likely to be the suspect from a photograph or video of a line-up. The purpose of a line-up is to obtain identification evidence that provides more information than that contained in a witness' verbal description of the culprit. A mock witness simulates the worst possible scenario of a witness who has no memory of the appearance of the culprit beyond that which they gave in a verbal description, but who chooses a person from the parade who best matches their description. In a perfectly fair line-up, all of the people in the parade would fit the witness' description to the same extent. Doob and Kirshenbaum argued that the probability of the suspect being chosen in a fair line-up would be at the level of chance expected if all choices were equally spread across all line-up members. However, if a parade is biased against the suspect, because some members of the parade do not fit the witness description, the suspect may be chosen at a rate above that which would be expected by chance alone. There is a large body of research on the mock witness paradigm, much of it is reviewed in the papers in this volume (also see Wells, 1993, and Wells *et al*, 1998 for reviews). A number of measures of line-up fairness have been developed. Tredoux (1996, this volume), Brigham and Pfeifer (1994), Brigham, Meissner and Wasserman (this volume), Lindsay, Smith and Pryke (this volume) all discuss the relative merits of these measures, their analysis and interpretation.

At this point it is necessary to briefly introduce two measures of line-up fairness. The first is simply the proportion of mock witnesses who chose the suspect. This measure was used by Doob and Kirshenbaum (1973). This proportion can be evaluated against that expected by chance under the assumption that the choices will be spread equally across all members of the line-up. A second measure of line-up fairness is given simply by taking the reciprocal of this proportion. This measure is known as 'functional size'. The functional size is simply calculated by dividing the total number of mock witnesses by the number who chose the suspect. For example, consider the situation in which 36 mock witnesses viewed a line-up that consisted of the suspect and eight volunteers, and four mock witnesses select the suspect. The functional size of the line-up would be 36/4 = 9. In this case the suspect was identified at the level expected by chance (36 x 1/9) and the functional size is equal to the nominal size of the line-up (i.e. the number of people present). If the suspect was the only person who remotely matched the witness description given to the mock witnesses, all 36 mock witnesses might select the suspect. In this case, the functional size would be 36/36 = 1. That is to say, functionally the line-up is equivalent to confronting the witness with the suspect alone, because people who were not present at the crime scene were able to select the suspect. The foils in the line-up might as well be absent because they do not lure mock witnesses into making any selections of the volunteers. Note that bias in favour of the suspect (i.e. the suspect is chosen less often than predicted by chance) produces a measure of functional size that is greater than the nominal size.

The mock witness paradigm has been used extensively in studies of simulated crimes using line-ups that are most frequently constructed by the researchers. However, there are only a few studies that have attempted to measure the fairness of real line-ups. Doob and Kirshenbaum (1973) reported an analysis of a single line-up, which they concluded was unfair. Brigham, Ready and Spier (1990) analysed the functional size of photo line-ups and photographs of live line-ups from six criminal cases. In some cases more than one witness was involved and one case used more than one line-up, therefore Brigham *et al.* (1990) measured the fairness of 12 combinations of line-ups and witness descriptions. The functional size of the line-ups ranged from 1.2 to 15.0. Six of 12 analyses yielded a functional size of 2 or less, suggesting that the probability of a mock witness choosing the suspect on the basis of the witness description alone was greater than 50%.

Brigham *et al.* (this volume) analysed the fairness of photo line-ups and photographs of two live line-ups used in 18 criminal cases. Multiple witnesses and the use of more than one line-up yielded a total of 26 line-up fairness analyses. Almost 60 per cent of these line-ups would be deemed unfair when adopting the criterion suggested by Brigham *et al.* (1990) that the functional size should be greater than half the nominal size (e.g. greater than 3 in a six person line-up).

Wells and Bradfield (this volume) report measures of the functional size of 10 line-ups used in real cases (five photo line-ups and 5 photographs of live line-ups). Using Brigham *et al.'s* (1990) criterion of fairness, 70% of these line-ups were unfair (with a functional size of less than 3.0).

It is difficult to know what to infer about the fairness of identification procedures used in real cases from these studies. The cases studied by Doob and Kirshenbaum (1973), and Brigham *et al.* (1990; this volume) were selected because they were referred to the researchers by lawyers who thought the line-ups were unfair. Therefore, these samples are likely to yield more unfair line-ups than a representative sample. Wells and Bradfield (this volume) do not state the basis on which they selected the line-ups studied. Furthermore, the provisions of PACE means that the situation in England and Wales may be very different. For example, the minimum nominal size of a line-up permitted under PACE is 9. The nominal size of the line-ups reported in the literature is as small as three in one case (see Wells & Bradfield).

The aim of the present research was to obtain as representative a sample of identification parades and video identifications that had been conducted under the PACE guidelines. The collection of material was made under the following practical and ethical constraints.

1. Legal proceedings in all cases were complete, allowing time for any appeal.

2. Records of identification attempts cannot be kept in cases that do not result in a conviction. Therefore, the defendant had been convicted in all cases included in the study.
3. PACE requires that the first description given by a witness is recorded. Therefore, this first description was used for the study.
4. The necessary photographs, videos and first description must be available to the research team.

The sample of video identifications was supplied by West Yorkshire Police who have developed a format for video parades known as VIPER. As the use of this system is relatively recent, the sample was an exhaustive set of the cases available at the time. The sample of photographs of identification parades was supplied by Durham Constabulary. They included line-ups held in Durham City Identification Suite using a one-way mirror and some line-ups held in police stations around County Durham. Again the sample was the entire set of relevant cases for which the necessary records were available.

We report three statistics for the line-ups that were evaluated. These three measures form the dependent variables in the study. First, the proportion of mock witnesses who chose the suspect was evaluated against the probability of the suspect being selected when each line-up member was chosen equally often using a Binomial test (as recommended by Tredoux (1996, p. 218). Second, functional size was calculated. A 95% confidence interval was calculated for functional size using the computational formula given by Tredoux (1996, equation 4).

Both functional size and the proportion of mock witnesses who select the suspect are measures of line-up bias. We also report a calculation of 'Effective Size' a measure of the extent to which mock witnesses choices are spread across all members of the line-up. When effective size is calculated, no account is taken of which member of the line-up is the suspect. Therefore, effective size is referred to as a measure of line-up size rather than line-up bias. The procedure for calculating effective size is given in Appendix A.

The fairness of a line-up is a function of both the composition of the line-up and the witness description. Any differences in the completeness of the witness descriptions given for video line-ups and live line-ups are a potential confound in the comparison between the measures of fairness of live and video line-ups. It is possible that any difference in fairness of the two modes of presentation found could be attributed to more complete witness descriptions for one set of cases. Therefore, the number of visible features mentioned in the descriptions was counted and used as a co-variate in the analysis so that the variance attributable to the witness description could be partialled out in the statistical analysis of the results.

Method

Mock witnesses: 216 mock witnesses participated in the study. They were students and staff of Goldsmiths College, the Royal Ballet School and the Central School of Ballet, London. The mean age of participants was 28.7 years (range 16 - 62); 157 were female, 59 were male; 149 were students, 13 were manual workers (e.g. porter), 38 were non-manual (mainly clerical and administration) workers and 15 were professional workers (e.g. lecturers, accountants).

Stimuli: The following aspects were common to the records of both live line-ups and video line-ups. PACE requires a minimum line-up size of 9. Some line-ups originally consisted of more than nine members. The photographs of live line-ups were trimmed when necessary to remove extra foils from the end of the line-ups**.** Four photographs were trimmed in this way. In all videos the suspect was included in the first nine members of the line-up. Therefore, the video was terminated after the ninth person, so that all line-ups consisted of exactly 9 members. Fourteen videos were trimmed in this way. Thus, the minimum provisions of PACE were evaluated. Each line-up was evaluated using only one witness description. In cases were more than one was available, the description given by the first witness listed was used. This procedure ensures that all fairness measures were independent of each other. The 'first description' that the witness gave of the culprit as recorded by the Police and supplied to the defence was used. The number of visual features mentioned in the witness' description that could serve as a basis for selecting the suspect was counted and recorded for use in the analysis. Gender, description of clothing and non-visual features (e.g. description of the culprit's accent) were excluded from the count.

Live line-ups: Records of 25 line-ups were used in the study. All were of white males. There was at least one colour photograph of each line-up. For fourteen of the line-ups there were photographs taken from each end of the line-up, or two photographs that were cut and joined to provide a clear view of the entire line-up. There were single photographs of eleven of the line-ups. Fourteen photographs were of line-ups held in an ID suite and eleven of the line-ups were held in a police station.

Video line-ups: Records of 16 video identifications were included in the study. Each video consisted of a 10 second clip of the head and shoulders of nine people. Participants only saw one face at a time, shown one after another. A number indicating each person's position in the line-up was shown on-screen. The model was facing the camera at the start of the clip. They turned slowly to show their left profile, then to their right profile than back to looking full-face at the camera.. Fifteen video line-ups were of males, there was one line-up of white females in the sample. Seven video line-ups were of white males, 5 were of black males, 2 were of Asian males (e.g. Indian, Pakistani origin), one was of males of mixed race. There were no Orientals.

Procedure: The testing was carried out 'blind' (i.e. the experimenter did not know which person in the line-up was the suspect). No feedback was given to the participants for ethical reasons. The stimuli were divided into 3 sets of 8 – 9 photographs of line-ups, and 3 sets of 5 – 6 videos. Each set was viewed by 36 participants. Each stimulus set was shown in four different orders. Each order was seen by 9 participants. Thus, each participant either saw only photographs of line-ups or viewed only videotapes. The witness description was read to the participant who was instructed to pick the person in the line-up who they thought was most likely to be the suspect. Participants who viewed the photographs were asked to ensure that they looked at members of line-up carefully at least twice before they made their decision. Participants who viewed videotapes were instructed not to make a choice until they had viewed the entire tape twice. (This instruction is included in the PACE code of practice.) The task was a forced-choice, participants were required to select a suspect.

Results and discussion

The mean of the measures of line-up fairness, together with the mean number of visible features mentioned in the eye-witness description, are shown in Table 1. None of the mock-witnesses chose the suspect in 5 of the line-ups (3 photographs and 2 videos). Functional size is undefined for these line-ups because the calculation requires a divide by zero. Therefore these line-ups were excluded from the calculation of the mean functional size.

	Photos of line-ups	Video identifications	All identifications
n	25	16	41
Proportion choosing suspect	0.248 *(0.048)*	0.151 *(0.037)*	0.210 *(0.037)*
Functional size n	11.05 *(2.76)* 22	14.91 *(3.86)* 14	12.55 *(2.25)* 36
Effective size	4.24 *(0.31)*	4.46 *(0.32)*	4.33 *(0.22)*
Visual features in description (covariate)	4.72 *(0.21)*	5.44 *(0.41)*	5.00 *(0.21)*

Table 1: Mean values of line-up fairness measures, the mean number of visual features mentioned in the original witness description and the number of line-ups analysed shown as a function of line-up method (photographs of live line-ups vs. video line-ups). The standard error of the mean is in parentheses. Line-ups in which no mock-witnesses chose the suspect are excluded from the calculation of functional size.

Approximately 25% of mock-witnesses chose the suspect in the photographs of live parades. This suggests that a quarter of witnesses who remember the description of the culprit that they gave to the police, <u>but have no other memory of the culprit</u> and do not reject the line-up, would be expected to select the suspect from a line-up. In the video line-ups this proportion was only 15%. If all members of a line-up are chosen equally often the rate of false positive identifications of the suspect would be 11%. The probability of choosing the suspect in the video line-ups was not statistically significantly different from that expected under the assumption of equal distribution of choices across all line-up members ($t(15)=1.09$, $p=.29$). However, the proportion choosing the suspect in the photographs of live parades was statistically greater than that expected by chance ($t(24)=2.84$, $p<.01$).

Analyses of covariance, taking the number of visible features mentioned in the eye-witness description as the covariate, were used to compare the fairness of live and video line-ups. An analysis of covariance showed that the proportion of mock-witnesses who chose the suspect was significantly lower for the video identifications than for the photographs of the live line-ups ($F(1,38)=4.81$, $p<.05$). The number of visual features in the witness description was significantly related to the proportion of mock witnesses who chose the suspect ($F(1,38)=7.02$, $p<.05$).

The means of functional size of the line-ups shown in Table 1 give a rather different impression. The mean functional size of photographs of live line-ups is 11 and the functional size of video line-ups is 15. Both are greater than the nominal size of the line-ups (9). These data suggest that on average the line-ups are biased in favour of the suspect. The calculation of functional size excludes the 5 line-ups most biased in the suspect's favour in which none of the mock-witnesses selected the suspect. However, the value of the mean functional size does reflect a problem with the statistic. The maximum defined value of functional size is given by the number of mock-witness who participated in the study. If we had used fewer mock-witnesses the line-ups would appear to be less biased in the suspects favour, if we had used more mock-witnesses they may appear more biased. This highlights a problem in using functional size to assess line-up bias when the line-up is biased in favour of the suspect.

An analysis of covariance, taking the number of visual features in the description as the covariate, showed no significant effect of the mode of presentation of the line-up on its functional size. However, the number of visual features in the witness description was significantly related to the functional size of the line-up ($F(1,33)=4.76$, $p<.05$).

The mean effective size of both live line-ups and video line-ups was in the range of 4.2 – 4.5. On average the choices made by mock witness were spread across 4.3 of the 9 line-up members. This spread of choices was clearly unaffected by the mode of presentation of the line-up.

In view of the implications for potential misidentification, the range as well as the mean of the measures of line-up fairness needs to be considered. Table 2 shows the minimum and maximum values measured together with the limiting values of the three statistics. Individual statistics on all line-ups are included in Appendix B. First, it is important to note that the sample includes some line-ups which are very much biased against the suspect and therefore are a cause of considerable concern about the potential for misidentification. Seventy-eight percent of mock witnesses selected the suspect in one photograph of a live line-up. It is important to note that the 95% confidence interval for the functional size for the live line-up that was most biased against the suspect did not overlap with the 95% confidence interval for the most biased video parade. Therefore, the ‘worst’ video line-up was significantly fairer than the ‘worst’ live line-up. In terms of reducing the potential for miscarriages of justice, the comparison between worst cases is important because poorly constructed line-ups are most likely to generate the problem.

The sample of line-ups used in this research was limited by a number of practical constraints and ethical considerations. The sample of video line-ups was particularly restricted by these factors. Importantly there are also differences between the sample of video and live line-ups. Most notably, 7 of the 16 video line-ups were of African-Caribbeans and 2 were of Asians (e.g. Indian, Pakistani), whereas all of the live line-ups were of white Europeans. Can these differences in the samples account for the differences in line-up fairness observed between video and live line-ups? Comparisons between the mean proportion of mock witnesses choosing the suspect and the mean functional size of video line-ups involving white Europeans (0.145, 14.54 respectively), African-Caribbeans (0.233, 10.59 respectively) and all race codes other than white Europeans (0.156, 15.23 respectively) showed no significant differences (All t values <1.04). The trend is for line-ups of white Europeans to be fairer to the suspect than line-ups involving African-Caribbeans. Therefore, any effect of race on the measures of line-up fairness would have diminished the advantage in fairness observed for the video line-ups.

		Most biased against suspect or minimum size	Binomial probability or 95% confidence interval	Most biased in favour of suspect or maximum size	Binomial probability or 95% confidence interval	Minimum value of statistic	Maximum value of statistic
Proportion choosing suspect	photos	.78	<.001	0	<.001	0	1
	videos	.44	<.001	0	<.001		
Functional size	photos	1.28	1.15-1.55	36	8.64-161.4	1	36
	videos	2.25	1.72-3.16	36	8.64-161.4		
Effective size	photos	1.33	-	8.00	-	1	9
	videos	2.42	-	7.25	-		

Table 2: Minimum and maximum values of line-up fairness as measured and the possible range of each statistic, as a function of line-up method (photographs of live line-ups vs. video line-ups). Line-ups in which no mock-witnesses chose the suspect are excluded from the calculation of functional size, which is undefined in these cases. Therefore, the maximum value of functional size is given by the number of mock witnesses. Ninety-five percent confidence intervals (Tredoux, 1996, equation 4) or binomial probability (based on Z approxiamtion) are included as appropriate.

Whenever the fairness of a sample of line-ups used in real cases is measured, a question of considerable applied interest is 'how many of the line-ups were unfair?' The definition of an unfair line-up is fraught with difficulties because different measures of line-up bias can lead to different conclusions for the same line-up (see Tredoux, 1996). Notwithstanding these problems, we suggest the following approach. A naive juror might reasonably expect the probability of identifying the suspect by chance in a line-up with 8 foils to be 1/9. The functional size statistic has the advantage that a confidence interval can be computed for each line-up. Therefore, we shall define an unfair line-up as one for which the entire 95% confidence interval of functional size is below 9. The rationale for this criterion is that a juror's reasonable expectation would be excluded by a criterion based upon the error of the measurement of line-up fairness. Using this definition 48% of the 'live' line-ups and 31% of video line-ups in our sample are unfair. Relaxing the criterion to classify only line-ups with a 95% confidence interval of functional size below 6 (i.e. two-third of the nominal size) results in only a modest reduction of the proportion of line-ups classified as unfair (44% of live line-ups and 25% of video line-ups).

These figures suggest that there is a considerable problem of bias against the suspect in line-ups, and that the bias can be reduced by use of video. It should be borne in mind that the line-ups were limited to 8 foils for the purposes on this study. Some of the line-ups actually used in the cases could have been fairer as they included extra foils. However, the following points should be noted. First, all line-ups as tested complied with the PACE code of practice. Second, foils were not removed from the least fair live line-up or but one foil was removed from the least fair video line-up. Third, more videos than live line-ups were trimmed. The last two points suggest that, if there is any effect of trimming, it would have led to an underestimate of increase in line-up fairness that can result from the use of video.

Our analyses reveal that in nine live line-ups (36%) and eight video line-ups (50%) the suspect was chosen by mock-witnesses less often than would be predicted under the assumption that each line-up member would be chosen equally frequency. The implication is that these line-ups may be biased in favour of the suspect. However, the 95% confidence interval of functional size included the nominal size (9) for all line-ups for which functional size could be determined. Therefore, bias in favour of the suspect can only be considered to have been reliably demonstrated for the 3 live line-ups and the 2 video line-ups for which none of the mock witnesses chose the suspect. The mock witness procedure is likely to reveal bias in favour of the suspect whenever the suspect does not fit the witness's description of the culprit**.** Inspection of the 5 relevant line-ups revealed that the suspect did not match the witness description on at least one aspect, and one or two foils who matched the description rather better drew a high proportion of mock-witness' responses.

Bias in favour of the suspect will not lead to wrongful convictions, but may be seen as leading to failures to convict guilty suspects. However, it can be argued that 'bias' in favour of a suspect is not a problem[1]. If the suspect is guilty, the witness can rely on recognition memory to make an identification. The fact that a witness selects the suspect in a line-up from which mock witnesses could not select the suspect, on the basis of the witness's description alone, provides evidence that it is very likely that the witness has genuinely recognised the suspect. Alternatively if the suspect is innocent, 'bias' in their favour will reduce the number of 'correct' guesses made by unreliable, compliant witnesses. This outcome is desirable.

The study has highlighted some differences between measures of line-up fairness that has been noted elsewhere (Brigham *et al*., this volume; Tredoux, this volume). The question arises of which measure is preferable. When the <u>mean</u> fairness of a set of line-ups is calculated, we argue that the proportion choosing the suspect is a more useful statistic than functional size for two reasons. (a) The proportion choosing the suspect can be calculated for all possible outcomes. (b) The range and therefore the mean is unaffected by the number of mock witnesses. Neither of these properties is true of functional size. An additional reason to prefer the proportion choosing the suspect as a measure of line-up bias is given by Lindsay *et al.* (this volume). They found that the proportion choosing a line-up member, but not the functional size of the line-up, was a significant predictor of the proportion of witness who identify the line-up member in a culprit-absent line-up. In contrast, when considering the fairness of a <u>single</u> line-up, the 95% confidence interval that can be calculated for functional size (Tredoux, 1996) makes this statistic particularly useful as a readily understandable estimate of measurement error can be made. The problem associated with the upper bound of functional size does not arise when considering a single line-up that is biased <u>against</u> the suspect.

[1] 1. We thank Rod Lindsay for drawing our attention to this line of argument.

Conclusions

Evaluation of a sample of photographs of live line-ups used in criminal cases has shown that the code of practice introduced in England and Wales under the Police and Criminal Evidence act 1984 is not sufficient to prevent unfair line-ups from being used to collect identification evidence. Seventy-eight percent of mock witnesses were able to select the suspect in the most unfair of the line-ups studied. Forty-eight percent of live line-ups would be classified 'unfair' by the criterion that the 95% confidence interval of the function size excludes the nominal size of the line-up. Use of video produced fairer line-ups. The mean proportion of mock witnesses who chose the suspect was 15% for video line-ups compared to 25% for live line-ups. This difference was statistically significant despite the relatively small sample size, and cannot be attributed to a wider diversity of the ethnic origins of the people in the video line-ups.

References

Ainsworth, P. B. & King, E. (1988). Witnesses' perceptions of identification parades. In: M.M. Gruneberg, P. E. Morris & R. N. Sykes (eds.) Practical aspects of memory: Current research and issues, Vol. 1. Chicester: Wiley

Brandon, R., & Davies, C. (1973*). Wrongful imprisonment: Mistaken convictions and their consequences.* London: George Allen & Unwin

Brigham, J. C. & Pfeifer, J. E. (1994). Evaluating the fairness of line-ups. In: Ross, D. F., Read, J. D. & Toglia, M. P. (eds.) *Adult eyewitness testimony: Current trends and developments.* Cambridge University Press: Cambridge.

Brigham, J. C., Meissner, C. A. & Wasserman, A. W. (this volume). Applied issues in the construction and expert assessment of photo lineups. *Applied Cognitive Psychology.*

Brigham, J. C. Ready, D. J. & Spier, S. A. (1990). Standards for evaluating the fairness of photograph lineups. *Basic and Applied Social Psychology, 11*, 149-163.

Davies, G. M. (1996). Mistaken identification: Where law meets psychology head on. *The Howard Journal of Criminal Justice, 35*, 232-241.

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

Doob, A. N. & Kirshenbaum, H. M. (1973). Bias in Police line-ups - partial remembering. *Journal of Police Science and Administration, 1,* 287-293.

Lindsay, R. C. L., Smith, S. M. & Pryke, S. (this volume). Measures of lineup fairness: Do they postdict identification accuracy? *Applied Cognitive Psychology.*

Huff, C. R., Ratner, A. & Sagarin, E. (1986). Guilty until proven innocent: Wrongful conviction and public policy. *Crime & Delinquency, 32,* 518-544.

Her Majesty's Stationery Office (1995) *Police and Criminal Evidence Act 1984 (s.60(1)(a) and s.66) Codes of Practice.* London: HMSO Publications.

R v Turnbull and others (1976) 3 All ER 549

Slater, A. (1994). *Identification parades: A scientific evaluation.* Police Research Award Scheme. London: Police Research Group, Home Office.

Tredoux, C. G. (1996). Statistical inference on measures of lineup fairness. *Law and Human Behavior,* 22, 217-237.

Tredoux, C. G. (this volume). Statistical considerations when determining measures of lineup size and lineup bias. *Applied Cognitive Psychology.*

Wells, G. L. (1993). What do we know about eyewitness identification? *American Psychologist, 48,* 553-571.

Wells, G. L. & Bradfield, A. L. (this volume). Measuring the goodness of lineups: Parameter Estimation, Question Effects, and limits to the mock witness paradigm. *Applied Cognitive Psychology.*

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

Wright, D. B. & McDaid, A. T. (1996). Comparing system and estimator variables using data from real line-ups. *Applied Cognitive Psychology, 10,* 75-84.

Appendix A

Effective size can be calculated by the following procedure:

1. Subtract the number of foils who received no choices from the nominal size of the line-up. The result is the adjusted nominal size. Foils who received no choices are ignored in all of the steps below.
2. Divide the number of mock witnesses by the adjusted nominal size to give an adjusted expected value of the number of times each remaining member of the line-up will be selected.
3. For all foils who were chosen less than the adjusted expected value, find the sum of the difference between the number of choices to each foil and the adjusted expected value.
4. Divide the sum calculated in step 3 by the adjusted expected value.
5. Subtract the value calculated in step 4 from the adjusted nominal size of the line-up (step 1). This value is the Effective Size.

Appendix B

No.	race	sex	N suspect	Prop choose	Binomial Prob	FS	FS_minus	FS_plus	ES	Visual Features
Photographs of 'live' line-ups										
1	1	m	10	.278	.002	3.60	2.42	5.76	5.13	4
2	1	m	3	.083	.396	12.00	5.22	29.67	5.65	5
3	1	m	2	.056	.214	18.00	6.45	54.06	2.97	5
4	1	m	8	.222	.032	4.50	2.83	7.70	5.75	6
5	1	m	28	.778	<.001	1.28	1.15	1.55	2.33	6
6	1	m	0	.000	<.001	-	-	-	3.00	3
7	1	m	1	.028	.093	36.00	8.64	161.37	5.22	4
8	1	m	15	.417	<.001	2.40	1.81	3.43	4.55	7
9	1	m	22	.611	<.001	1.63	1.36	2.11	2.94	7
10	1	m	1	.028	.093	36.00	8.64	161.37	5.67	5
11	1	m	19	.528	<.001	1.89	1.52	2.54	2.78	5
12	1	m	0	.000	<.001	-	-	-	1.33	4
13	1	m	11	.306	<.001	3.27	2.26	5.09	4.44	4
14	1	m	19	.528	<.001	1.89	1.52	2.54	3.75	4
15	1	m	5	.139	0.395	7.20	3.85	14.47	5.75	5
16	1	m	2	.056	.214	18.00	6.45	54.06	8.00	4
17	1	m	26	.722	<.001	1.38	1.21	1.70	2.67	5
18	1	m	6	.167	.212	6.00	3.43	11.29	6.55	4
19	1	m	1	.028	.093	36.00	8.64	161.37	2.67	3
20	1	m	17	.472	<.001	2.11	1.65	2.93	3.94	5
21	1	m	0	.000	<.001	-	-	-	3.39	5
22	1	m	11	.306	<.001	3.27	2.26	5.09	5.44	4
23	1	m	10	.278	.002	3.60	2.42	5.76	3.25	4
24	1	m	1	.028	.093	36.00	8.64	161.37	4.33	4
25	1	m	5	.139	.395	7.20	3.85	14.47	4.55	6
Video line-ups										
26	2	m	16	.444	<.001	2.25	1.72	3.16	3.25	6
27	1	m	3	.083	.396	12.00	5.22	29.67	6.11	5
28	1	m	6	.167	.212	6.00	3.43	11.29	4.33	9
29	1	m	1	.028	.093	36.00	8.64	161.37	4.67	2
30	2	m	8	.222	.032	4.50	2.83	7.70	5.67	7
31	1	m	12	.333	<.001	3.00	2.12	4.56	4.56	7
32	3	m	2	.056	.214	18.00	6.45	54.06	5.44	5
33	2	m	5	.139	.395	7.20	3.85	14.47	3.55	5
34	1	m	0	.000	<.001	-	-	-	2.42	4
35	1	m	13	.361	<.001	2.76	2.01	4.11	3.11	5
36	1	m	6	.167	.212	6.00	3.43	11.29	7.25	4
37	2	m	1	.028	.093	36.00	8.64	161.37	4.50	7
38	3	m	1	.028	.093	36.00	8.64	161.37	2.67	5
39	1	f	1	.028	.093	36.00	8.64	161.37	5.00	6
40	2	m	12	.333	<.001	3.00	2.12	4.56	4.75	6
41	4	m	0	.000	<.001	-	-	-	4.17	4

Table B1: Raw data for all line-ups evaluated. Key: Race; 1 = white European; 2 = African-Caribbean; 3 = Asian (e.g. Indian, Pakistani); 4 = mixed race. N suspect = the number of mock witnesses who chose the suspect; Binomial prob. = the binomial probability that the proportion of mock witnesses who chose the suspect differs from that expected by chance (1/9); Prop. Choose = the proportion of mock witnesses who chose the suspect; FS = functional size; FS_minus = the minimum endpoint of the 95% confidence interval of FS; FS _ plus = the maximum endpoint of the 95% confidence interval of FS; ES = effective size; Visual features = number of visible features mentioned in the witness's first description.