Applied Cognitive Psychology (in press)
Manuscript received April 2002
Manuscript accepted December 2002

Characteristics of eyewitness identification that predict the outcome of real lineups

Tim Valentine, Alan Pickering & Stephen Darling
Goldsmiths College,
University of London,
England.

Short heading: Identification from real lineups

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

Tel: +44 20 7919 7871
email: t.valentine@gold.ac.uk

Acknowledgements

We gratefully acknowledge the contribution of Chief Inspector David Wilkins, Inspector Chris Jones and Inspector Peter Douglas of the Metropolitan Police for help in setting up and designing the study; Inspector Chris Jones, the identification officers at St Anne's Identification Suite and the investigating officers in the cases concerned for their considerable effort in collecting the data; Dr. Donald Hedeker for advice on use of his statistical software and Stuart Stiff for assisting with exploratory data analysis.

Abstract

Data were analysed from 640 attempts by eyewitnesses to identify the alleged culprit in 314 lineups organised by the Metropolitan Police in London. Characteristics of the witness, the suspect, the witness's opportunity to view the culprit, the crime and the lineup were recorded. Data analysis, using mixed effects multinomial logistic regression, revealed that the suspect was more likely to be identified if the witness is younger than 30, the suspect is a white European (rather than African - Caribbean), the witness gave a detailed description, viewed the culprit for over a minute and made a fast decision at the lineup. None of the explanatory variables were significantly associated with a mistaken identification of a foil. No independent, statistically reliable effects of weapon focus, cross-race identification or of the delay before the identification attempt were observed.

Introduction

The procedures used to obtain eyewitness identification have been intensively studied by psychologists over the last twenty years. There has been considerable development in both the methodology of the studies and in our understanding of the psychology of eyewitness identification. For reviews of the literature see Cutler and Penrod (1995), Lindsay (1999), Sporer, Malpass and Koehnken (1996), Ross, Read and Toglia; Wells (1993); Wells, Small, Penrod, Malpass, Fulero, & Brinacombe (1998)

Much progress has been made through carefully controlled laboratory studies. It has been recognised that this approach should be augmented by the study of identification attempts by real witnesses to crimes (e.g. Yuille, 1993). However, there are considerable practical and ethical difficulties in studying eyewitness identification in real criminal cases, which can make the results difficult to interpret. As a result there remain relatively few such studies. One of the major difficulties that an applied researcher faces is that one seldom knows with any degree of certainty whether a lineup in a criminal case genuinely contains the culprit.

Several researchers have used laboratory procedures to measure the fairness of lineups from real cases (Brigham Meissner & Wasserman*,* 1999; Brigham, Ready and Spier, 1990; Doob and Kirshenbaum, 1973; Wells and Bradfield, 1999). Typically, these studies involve relatively small numbers of lineups (up to approximately 20) which were referred to the researchers by lawyers who thought the lineups were unfair. Therefore, these samples are likely to reflect the worst cases rather than form a representative sample. All of the above studies were of cases from the USA and mainly involve use of arrays of photographs (photospreads) to obtain identification evidence. Valentine and Heaton (1999) and Valentine, Harris, Colom Piera and Darling (in press) analysed the fairness of live lineups and/or video lineups from criminal cases in England. These procedures were run under the provisions of English law that differs substantially from the legal position in the USA. The studies by Valentine and colleagues used a representative sample of lineups obtained from police records but the number of lineups evaluated was still relatively small (up to 41). Valentine and Heaton (1999) found that, based only on the first description that a witness gave of the culprit, 25% of participants were able to select the suspect from a photograph of a live lineup but only 15% selected the suspect from a video lineup. Thus video lineups were fairer than live lineups. Valentine et al (in press) found that video lineups were equally fair to white European and African – Caribbean suspects. The British Home Office has recently announced a change in the codes of practice that will facilitate greater use of video lineups in England and Wales.

An alternative strategy is to use either an archival study or a field study to record and analyse the outcome of a much larger sample of identification attempts. Slater (1994) recorded the outcome of identification attempts by 843 witnesses who inspected 302 live lineups in England. The suspect was identified by 36% of witnesses, a foil was identified by 22% and 42% made no positive identification. Wright and McDaid (1996) analysed the outcome of identification attempts by 1561 witnesses who inspected 616 live lineups and found very similar outcomes; the suspect was identified by 39% of witnesses, a foil was identified by 20% and 41% made no positive identification. Their analysis included explanatory variables divided into estimator variables (i.e. features of the particular incident) and system variables (i.e. features of the identification process). The estimator variables included whether or not the crime involved violence, whether a weapon was present, the race of the suspect and the delay before the identification attempt. The only significant effect was that suspects from ethnic minorities were more likely to identified than white suspects. Adding system variables that coded for lineups held in two identification suites rather than a police station significantly improved the fit of the statistical model to the data. Foils were more likely to be identified in the identification suites, in which the lineup is viewed through a one-way mirror. A brief report of recent survey of identification procedures commissioned by the British Home Office reported a somewhat higher rate of suspect identifications, 49%, from approximately 8,800 lineups (Pike, Brace & Kynan, 2002). The higher rate reported in this survey may reflect a 'positive' outcome being counted as a suspect identification from one of several witness who saw a single lineup in some cases. The survey did not distinguish identification of a foil from no identification.

Trollestrup, Turtle and Yuille (1994) carried out a detailed archival analysis of 98 robbery and fraud cases in a district of Vancouver, Canada, which examined the influence of a number of factors on witness descriptions and identifications. There were 170 identification attempts in the cases; 90% used photospreads. The data did not allow the distinction between identification of a foil and no identification to be made. Robbery victims identified 46% of suspects; robbery witnesses identified 33% of suspects; fraud victims identified 25% of suspects; there were no data available from witnesses to fraud. These differences between the role of the witness and the type of offence were not statistically significant. When the analysis was restricted to suspects who confessed (n = 13 robbery victims; 9 robbery witnesses; 22 fraud victims), there was a significant difference between robbery victims (85% identified the suspect) and fraud victims (23% identified the suspect). Unfortunately the comparison between victims of fraud and robbery was confounded with a shorter delay before the identification for robbery victims.

The group of witnesses for robbery (n = 18) who experienced the shortest delay prior to identification (range of 0 – 1; mean = 0.5 days) identified 71% of suspects. The group of witnesses (n = 18) for fraud who experienced the shortest delay (range of 7 – 62 days; mean = 33 days) identified 78% of suspects. The identification rate of robbery suspects showed a monotonic decrease with delay (46% after a mean of 4 days, 33% after a mean of 19 days, and 14% after a delay of 120 days). The identification rate of fraud suspects after a mean delays of 74 days, 132 days and 200 days were much reduced but did not show a monotonic trend (5%, 20% & 17% respectively). The authors do not report any statistical tests of the effect of delay.

The evidence from laboratory studies suggests that the probability of identifying a culprit is reduced by the presence of a weapon (e.g. Cutler, Penrod & Martens, 1987; Cutler, Penrod, O'Rourke & Martens, 1986; Loftus, Loftus & Messo, 1987). Steblay (1992) found reliable effect of the presence of a weapon from a meta-analysis of 19 studies. Trollestrup et al. (1994) found a complex effect of the presence of a weapon in robbery cases. Presence of a weapon did not have a detrimental effect on the detail or accuracy of the description of the culprit, but did reduce the likelihood of the suspect being identified. Forty four percent of suspects of robberies involving a weapon were identified compared to 71% of suspects of robberies without a weapon. The mean delay in the relevant samples was matched at approximately 5 days.

Trollestrup et al.'s (1994) study illustrates some of the complications and limitations of interpreting an analysis of data from real cases. The aim of the present study was to augment our knowledge of the factors affecting real eyewitness identification by analysis of data from a larger sample of lineups. The study was designed to evaluate the external validity of a number of empirical findings and theoretical predictions from the scientific literature and to test some assumptions about eyewitness testimony in the English legal system. The present study differs from previous work by collecting data on a wider range of explanatory variables in a large sample of lineups, which allow the independent effects of explanatory variables to be observed.

The data were collected from one of four identification suites that serve the London area. All lineups in the Metropolitan Police area are now conducted in an identification suite, in which the witness views the lineup through a one-way mirror. A questionnaire was designed to collect information on factors that are either predicted by psychological theory to affect eyewitness accuracy, have been found to affect face recognition accuracy in laboratory studies or are given explanatory value by the English legal system. These explanatory variables can be divided into witness characteristics (e.g. age, gender, race, role), suspect characteristics (e.g. age, gender, race, height, build), variables about the eyewitness situation (e.g. viewing conditions), the incident (e.g. offence, presence of weapon), the eyewitness' description (e.g. completeness, match to suspect appearance) and variables associated with the identification attempt (e.g. delay, witness' decision speed) . The questionnaire allowed the following effects found in laboratory studies to be evaluated: the presence versus absence of a weapon (e.g. Loftus, Loftus & Messo, 1987); same versus cross race identification (e.g. Chiroro & Valentine, 1995, Valentine & Endo, 1992); the speed of the witness' choice (e.g. Sporer, 1992)

In cases that depend wholly or substantially on disputed eyewitness identification, English case law requires the judge to caution the jury on the need for special caution in relying on the accuracy of the identification to convict. The case law (R v Turnbull, 1976) has become known as 'the Turnbull guidelines'. The judge should instruct the jury on the reason for the warning and refer to the possibility that a mistaken witness can be a convincing one and that a number of such witnesses can all be mistaken. The judge should also instruct the jury to examine closely the circumstances under which each witness came to make their identification. The following is an extract from the judgement in court of appeal:

> "How long did the witness have the accused under observation? At what distance? In what light? Was the observation impeded in any way, as for example by passing traffic or a press of people? Had the witness ever seen the accused before? How often? If only occasionally, had he any special reason for remembering the accused? How long elapsed between the original observation and the subsequent identification to the police? Was there any material discrepancy between the description of the accused given to the police by the witness when first seen by them and his actual appearance?" (R v Turnbull (1976), p. 552; Lord Widgery CJ)

In this way the judgement asserts that good identification evidence can be distinguished from poor identification evidence. It concludes:

"In our judgment, when the quality is good, as for example when the identification is made after a long period of observation, or in satisfactory conditions by a relative, a neighbour, a close friend, a workmate and the like, the jury can safely be left to assess the value of the identifying evidence even though there is no other evidence to support it; providing always, however, that an adequate warning has been given about the special need for caution." (ibid.)

The present study was designed to measure the influence of the factors mentioned in the Turnbull guidelines on the likelihood of a suspect identification.

For legal reasons it was not possible to ask the witness to complete a questionnaire. Therefore, the investigating officer who accompanied the witness to the identification suite provided details about the witness, their opportunity to view the culprit, and the suspect. English law prohibits the investigating officer from being present during the actual identification procedure. These officers were familiar with the case and would have been likely to have taken the witness' statement. Therefore they were able to complete the questionnaire while waiting at the identification suite for the witness to complete the identification procedure. The identification officer responded to questions about the identification procedure.

The questionnaire was designed in consultation with experienced identification officers who were involved in the management of the Metropolitan identification suites. The identification officer in charge of the identification suite trained the individual identification officers from instructions provided by the researcher.

Method.

The data were collected in one of four identification suites that serve the Greater London area. Data were collected from January – September 2000. Questionnaires relating to a total of 664 witnesses who viewed 323 parades were completed. Data from 24 witnesses were excluded due to missing data (i.e. the outcome of the parade was not recorded), leaving data from 640 witnesses and 314 parades for analysis. The investigating officer provided the following details about the witness, their opportunity to view the culprit, the suspect, the offence and the witness' description by checking the appropriate boxes: The witness' age (9 levels), gender, race (using police race codes, 7 levels); the witness' role (victim, friend or relative, bystander); the time for which the witness could view the culprit (from less than 10 seconds to more than 1 minute; 4 levels); the quality of the lighting (good, poor, bad); whether their view was obstructed or not; the distance of the witness from the suspect (less than 2 metres to more than 25 metres; 4 levels); whether the suspect was known to the witness and whether the culprit carried a weapon (firearm, knife, other, none); the suspect's age, gender, race, height (short, average, tall) and build (thin, average, heavily built); whether there was a mismatch between the witness' description and the appearance of the suspect (hair description, race, build, age, other); the completeness of the witness description (few details, average, detailed); the offence (murder, grievous bodily harm, actual bodily harm, rape, indecent assault, armed robbery, street robbery, theft, burglary, criminal damage, other), and the delay prior to the identification attempt (from less than 1 week to over 6 months; 8 levels). Many different investigating officers will have provided details for different lineups. Usually there will have been one investigating officer for each lineup. The identification officer, one of approximately 6 officers who worked at the identification suite, provided details of the identification procedure itself. The following details were coded: the outcome of the lineup (suspect identification, foil identification, no identification), the speed with which the witness made their decision (fast, average, slow) and how difficult it had been to construct an acceptable lineup for the suspect (difficult, average, easy). The last variable was designed as a measure of the difficulty in selecting foils who resemble the suspect. Thereby it was intended to provide a measure of the distinctiveness of an individual. This factor is included because distinctiveness has been found to be influential in laboratory studies face recognition (Valentine, 1991), and it is intended to capture at least one aspect of a factor which Lord Widgery referred to as a "special reason for remembering the accused" (R. V Turnbull, 1976). One form was completed for each witness. The questionnaires were read by an optical mark reading machine. Missing data were checked against the original form by hand.

Results

The data available for analysis consisted of the outcome for 640 witnesses who attempted to identify suspects in 314 lineups. Fifty-six witnesses knew the suspect. The outcome of the identification attempts for known and unknown suspects are shown in Table 1. Data from two previous studies of lineups organised by British police forces are included for comparison (Slater, 1994; Wright and McDaid, 1996). It is not clear whether Slater excluded identifications of suspects known previously to a witness, but Wright and McDaid did not. Nevertheless the results of the studies are remarkable consistent. The findings for suspects not previously known can be summarized as follows. Approximately 20% of witnesses identify a foil, and thereby make a known mistaken identification. Approximately 40% identify the suspect and approximately 40% make no identification. Unsurprisingly, suspects known to the witness are more likely to be identified than are unknown suspects.

Table 1: The outcome identification attempts expressed as the percentage of witnesses for suspects known and unknown to the witness.

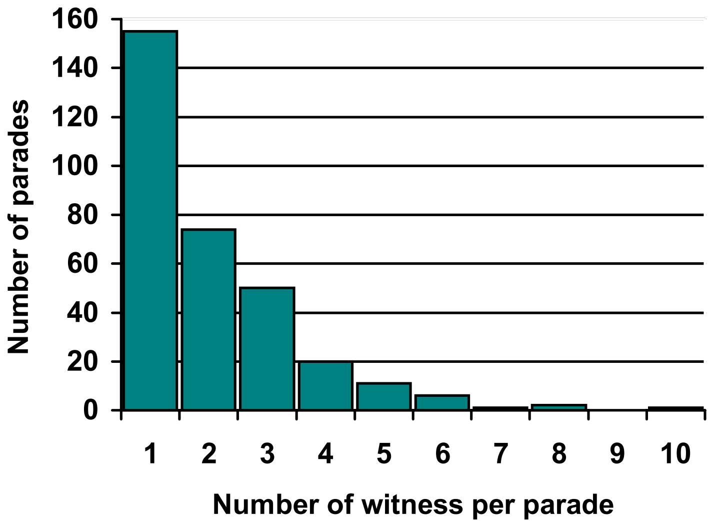
	Suspect ID	Foil ID	No ID	No. of witnesses	No. of lineups
Known	73%	5%	21%	56	25
Unknown	41%	21%	39%	584	295
All	44%	19%	37%	640	314
Slater 1994	36%	22%	42%	843	302
Wright & McDaid (1996)	39%	20%	41%	1561	616

Notes: Data from Slater (1994) and Wright and McDaid (1996) are included for comparison. It is not known whether Slater excluded suspects known to the witness, Wright and McDaid did not exclude known suspects.

The outcome of a lineup produces nominal data, with three possible outcomes (suspect identification, foil identification, no identification). Multinomial logistic regression is an appropriate model to use to assess the influence of explanatory variables if the data are independent. The present data, however, are clustered at the level of lineups. Many lineups were inspected by more than one witness. Lineups may differ from each other (e.g. in the fairness of the selection of foils). Therefore, the outcomes from two witnesses who inspected the same lineup are unlikely to be independent. Therefore, the use of the ordinary multinomial logistic regression model would be inappropriate in this case because it assumes independence of the data. Clustered data can be appropriately analysed by use of the mixed-effects multinomial logistic regression model, which allows different random terms to be included in the regression model for each cluster (or lineup). A random intercept reflects the inclusion in the regression model of a different intercept for each lineup. The software used for the analysis was developed by Hedeker (1999). Further discussion of statistical analysis of this type of data can be found in Hedeker (1999) and Wright and McDaid (1996).

The present data have a further restriction. The number of witnesses who inspected each lineup is displayed in Figure 1. The median value is one. This distribution restricts the number of random effects that can meaningfully be estimated in the model. The models calculated are therefore restricted to models that include no

Figure 1: The number of witnesses per lineup



more than one random factor (the intercept) for each response variable category. ‘No identification’ is treated at the reference category against which the probabilities of identifying the suspect, or identifying a foil, are compared. Effects are calculated separately for the likelihood of suspect identifications and foil identifications.

There are a large number of potential explanatory variables that could be entered into a model. A number of models with a small number of explanatory variables were used to explore the effect of these variables. In all cases the independence model was calculated first. A model with a single random intercept for both response variable comparisons was then calculated and the fit to the data of the models evaluated. If the model with the random effect provided a significantly better fit, a model which allowed the random intercept to vary for each response variable contrast (identify suspect vs. no identification; identify foil vs. no identification) was calculated. Again the statistical fit to the data was compared and the appropriate model was selected for report of the results.

In the first analysis the effect of the suspect being known to the witness was explored in isolation. The interest in exploring the remaining variables is restricted to cases in which the suspect is unknown.

<u>Suspects known to the witness</u>

The effect of the suspect being known to the witness was the sole variable included in a model of data drawn from 640 witnesses and 314 lineups. The fit of a model assuming that all data points are independent (log likelihood = -664.7) was compared to one in which a single random intercept was included (log likelihood = -641.2). The intercept in the latter model was constrained to be the same for both comparisons between the nominal categories of the response variable (i.e. the comparison between suspect ID and no ID, and between foil ID and no ID). Therefore this model has only one extra degree of freedom. The difference in fit of the models shows that inclusion of the random intercept significantly improved the fit of the model (χ^2 (1) = 47.0, p<.005). (See Appendix A for an explanation of the comparison between models). A further model allowing a different random intercept for each of the two contrasts between nominal categories of the response variable was estimated. The fit of this model (log likelihood = -632.3) showed a further statistical improvement relative to the model with only one random intercept (χ^2 (1) = 17.9, p<.005). For brevity the results of only the best-fitting of these models is shown in Table 2. Models with more random parameters were not estimated. The small number of witnesses per lineup was judged to make calculation of more complex models inadvisable. This view was confirmed by Hedeker, (personal communication; also see Wright & McDaid, 1996). Inspection of Table 2 reveals that significantly more suspects are identified if they are known to the witness, but that there is no effect of the suspect being known on the number of foil identifications.

Table 2: The model parameters for a mixed effect multinomial model with separate random intercepts for each contrast of the response variable (outcome).

Variable	Estimate	SE	z	*p*
Fixed effects				
$Intercept_s$	-1.70	0.50	-3.40	<.01
$Intercept_f$	0.26	0.71	0.36	0.72
$Known_s$	1.64	0.42	3.89	<.01
$Known_f$	-0.66	0.67	-0.98	0.33
Random effects				
$Intercept_s$	2.00	0.29	6.88	<.01
$Intercept_f$	0.55	0.26	2.12	0.02

Notes: Subscript s denotes parameters from the contrast between choosing the suspect and no identification; subscript f denotes parameters from the contrast between choosing a foil and no identification. Probability levels of fixed-effects are for two-tailed tests. The random effects intercept parameter is a standard deviation for the random normal distribution of intercept values. A standard deviation must be positive, therefore all random effects are evaluated by a one-tailed test.

For all of the following analyses the data were restricted to those cases in which the suspect was unknown to the witness (584 witnesses, 295 lineups). This restriction is appropriate because the aim was to identify the factors that affect eyewitness identification accuracy in cases where the suspect is unknown. Furthermore, the restriction allows a better comparison between our data and experimental field and laboratory studies.

The same procedure that was adopted for the witness known/unkown analysis was also adopted for all of the analyses reported below. Models assuming independence were calculated first, followed by models with one intercept for both response variable contrasts, and finally models assuming separate intercepts for each contrast. The log likelihood of each model was evaluated against the preceding model (see Appendix A). For brevity, these log likelihood contrasts are not reported. In all comparisons the less parsimonious model provided a statistically better fit to the data. However, in some cases the models with more parameters did not meet the criterion for convergence (0.0001) and therefore could not be calculated. Only the results of the best-fitting model are reported. In some cases variables were recoded in order to maintain a sufficient number of observation in each cell of the analysis. It was attempted to maintain a minimum n in each cell of approximately 15. However, this wasn't possible in all cases. Analyses that were motivated by a directional a priori prediction, included a smaller n in some cells if the models converged satisfactorily. In some other analyses a smaller n was retained if the models converged rather than exclude a substantial number of cases. The n in each cell is shown in the appropriate tables.

Witness variables.

A model with four variables was calculated: Age of witness, recoded into 4 levels (< 20 years, 21-30, 31-40, >40); gender of witness (male, female); race of witness (restricted to 4 categories: white European, dark European, African-Caribbean, Asian[1]; and the role of the witness (3 levels: victim, friend or relative of the victim, bystander). Data on these variables were available from 546 witnesses and 289 lineups. The outcome of the lineups as a function of these variables is shown in Table 3. The results of the model are shown in Table 4.

Table 3: The outcome of identification attempts expressed as the percentage and number of witnesses (in parentheses) as a function of witness variables (N = 546).

	Suspect ID	Foil ID	No ID
Age of witness			
<20	47.6% (91)	15.7% (30)	36.6% (70)
21-30	45.7% (75)	19.5% (32)	34.8% (57)
31-40	35.6% (37)	25.0% (26)	39.4% (41)
>40	27.9% (34)	26.2% (32)	45.9% (56)
Gender of witness			
male	41.1% (172)	17.5% (73)	41.4% (173)
female	40.5% (66)	28.2% (46)	31.3% (51)
Race of Witness			
White European	41.6% (160)	21.3% (82)	37.1% (143)
Dark European	52.2% (24)	10.9% (5)	37.0% (17)
African – Caribbean	29.7% (19)	25.0% (16)	45.3% (29)
Asian	31.6% (18)	24.6% (14)	43.9% (25)
Role of witness			
victim	39.8% (134)	20.2% (68)	40.1% (135)
friend/relative	41.8% (38)	24.2% (22)	34.1% (31)
bystander	41.8% (64)	19.6% (30)	38.6% (59)

Note. Rows sum to 100%.

The age of the witness has a significant effect on the number of identifications of the suspect. 48% of witnesses under 20 years identified the suspect, compared to only 28% of witnesses over 40 years. However, age did not have a significant effect on the likelihood of identifying a foil. Inspection of the original data suggests that a

1. These categories are based on English Police race identification codes. The term 'Asian' refers to people from the Indian sub-continent. The categories Oriental (far-eastern countries) and Arab were excluded due to the small numbers in the dataset.

slow decline in the number of positive suspect identifications begins as early as the 31-40 age group. The proportion of suspect identifications are as follows. (The number of witnesses in each age group is shown in parentheses.) 0-10: 50% (8); 11-20: 47% (183): 21-30: 46% (164); 31-40: 36% (104); 41-50: 31% (52); 51-60: 32% (37); 61-70: 27% (11); 71-80: 18% (11); 81+: 9.1% (11).

Table 4: The model parameters for a mixed effect multinomial model of witness variables with separate random intercepts for each contrast of the response variable (outcome).

Variable	Estimate	SE	z	*p*
Fixed effects				
Intercept_s	0.34	0.64	0.53	0.60
Intercept_f	-1.54	0.54	-2.83	<0.01
Age_s	-0.31	0.14	-2.30	0.02
Age_f	0.04	0.10	0.37	0.71
Gender_s	0.46	0.33	1.40	0.16
Gender_f	0.67	0.28	2.35	0.02
Race_s	-0.19	0.14	-1.41	0.16
Race_f	0.03	0.10	0.31	0.76
Role_s	0.00	0.16	-0.04	0.97
Role_f	0.03	0.14	0.21	0.82
Random effects				
Intercept_s	1.86	0.34	5.45	<0.01
Intercept_f	0.36	0.27	1.35	0.09

Notes. *Subscript s denotes parameters from the contrast between choosing the suspect and no identification; subscript f denotes parameters from the contrast between choosing a foil and no identification. Probability levels of fixed-effects are for two-tailed tests. The random effects intercept parameter is a standard deviation for the random normal distribution of intercept values. A standard deviation must be positive, therefore all random effects are evaluated by a one-tailed test.*

Witness gender had a significant effect only on the number of mistaken identifications of foils ($p = 0.02$; see Table 4). Almost identical proportions of males and females identified the suspect. However, males were less likely than females to identify a foil (17% vs. 28% respectively) and correspondingly more likely to make no identification (41% vs. 31%). Thus, females were more likely to choose from the lineup but no more likely to identify the suspect. This effect may reflect a tendency for females to be more compliant to the task demands of the Police by making a choice from the lineup. Neither witness race nor the role of the witness showed any statistically significant effect on the outcome of the lineups.

Variables relating to the witnesses' view of the culprit.

A model including four variables concerned with viewing conditions was calculated. The variables were recoded into the following form: Duration had 2 levels: less than 1 minute, more than 1 minute. Lighting quality had 2 levels: good and poor. Daylight or good indoor lighting was coded as 'good'. Poor indoor lighting, twilight or night was coded as 'poor'. Obstruction had two levels: partly obstructed or unobstructed. Distance had 2 levels: less than 2 metres, over 2 metres. Data for these variables were available from 570 witnesses and 290 lineups. The outcome of the lineups as a function of these variables is shown in Table 5. The results of the model are shown in Table 6.

Table 5: The outcome of identification attempts expressed as the percentage and number of witnesses (in parentheses) as a function of variables relating to the witness' view of the culprit (N = 570).

	Suspect ID	Foil ID	No ID
Duration			
< 1 minute	32.2% (67)	22.6% (47)	45.2% (94)
> I minute	45.3% (169)	19.6% (73)	35.1% (131)
Lighting Quality			
good	43.2% (190)	20.5% (90)	36.4% (160)
poor	33.1% (47)	21.8% (31)	45.1% (64)
Obstruction			
unobstructed	42.6% (218)	20.3% (104)	37.1% (190)
obstructed	27.3% (18)	24.2% (16)	48.5% (32)
Distance			
< 2m	41.3% (179)	19.6% (85)	39.0% (169)
> 2m	40.4% (59)	24.0% (35)	35.6% (52)

Note. Rows sum to 100%.

The only significant effect was that witnesses who had more than one minute to view the culprit were more likely to identify the suspect than witnesses who viewed the culprit for a minute or less ($p = 0.04$). Duration had no effect on the likelihood of identifying a foil. There was no significant independent effect of the quality of the lighting, obstruction of view or viewing distance.

Table 6: The model parameters for a mixed effect multinomial model of variables related to the witness' view of the culprit, with separate random intercepts for each contrast of the response variable (outcome).

Variable	Estimate	SE	z	*p*
Fixed effects				
Intercept_s	0.124	1.01	0.12	0.90
Intercept_f	-0.87	0.75	-1.17	0.24
Duration_s	0.68	0.32	2.09	0.04
Duration_f	0.19	0.25	0.77	0.44
Lighting_s	-0.64	0.40	-1.59	0.11
Lighting_f	-0.24	0.26	-0.91	0.36
Obstruction_s	-0.58	0.48	-1.20	0.23
Obstruction_f	0.03	0.36	0.086	0.93
Distance_s	0.07	0.32	0.23	0.82
Distance_f	0.29	0.27	1.06	0.28
Random effects				
Intercept_s	1.95	0.32	6.13	<0.01
Intercept_f	0.34	0.24	1.44	0.07

Notes. Subscript s denotes parameters from the contrast between choosing the suspect and no identification; subscript f denotes parameters from the contrast between choosing a foil and no identification. Probability levels of fixed-effects are for two-tailed tests. The random effects intercept parameter is a standard deviation for the random normal distribution of intercept values. A standard deviation must be positive, therefore all random effects are evaluated by a one-tailed test.

Offence

The number and outcome of identification attempts as a function of the alleged offence are shown in Table 7. Attempted offences are included but suspects known to the witness are excluded from the data set. Rape suspects were identified by 90% of witnesses. The data were recoded to avoid a small number of identification attempts or missing data in some cells. The offences were recoded into the following categories: Offences against the person (including, murder, grievous bodily harm, actual bodily harm, rape and indecent assault); robbery (armed robbery, street robbery); burglary and theft; criminal damage and other offences. A model including only the recoded variable of offence was calculated. Data on this variable were available from 583 witnesses and 294 lineups. The model with a random intercept that varied for each nominal response contrast gave the best fit but showed no significant effect of offence.

Table 7: The outcome of identification attempts expressed as the percentage and number of witnesses (in parentheses) as a function of the offence with which the suspect was charged (N = 583).

	Suspect ID	Foil ID	No ID
Murder	29.2% (7)	8.3% (2)	62.5% (15)
Grievous bodily harm	45.5% (70)	18.2% (28)	36.4% (56)
Actual bodily harm	58.2% (32)	12.7% (7)	29.1% (16)
Rape	90.0% (9)	0	10.0% (1)
Indecent assault	46.7% (7)	26.7% (4)	26.7% (4)
Crimes against person	48.6% (121)	16.5% (41)	34.9% (87)
Armed robbery	32.3% (10)	19.4% (6)	48.4% (15)
Street robbery	36.0% (67)	23.1% (43)	40.9% (76)
Robbery	35.2% (75)	22.1% (47)	42.7% (91)
Theft	32.0% (8)	40.0% (10)	28.0% (7)
Burglary	33.3% (14)	19.0% (8)	47.6% (20)
Theft & burglary	32.8% (21)	26.3% (15)	36.8% (21)
Criminal Damage	27.3% (6)	27.3% (6)	45.5% (10)
Other	40.8% (20)	20.4% (10)	38.8% (19)
Crim. Damage & Other	36.8% (21)	26.3% (15)	36.8% (21)

Notes. The data were recoded into 4 categories for analysis: Offences against the person (including, murder, grievous bodily harm, actual bodily harm, rape and indecent assault); robbery (armed robbery, street robbery); burglary and theft; criminal damage and other offences. Rows sum to 100%. Individual offences sum to more than the total in the summery statistic because some suspects were charged with more than one offence. Only the more serious offence is included in the summary statistics.

Table 8: The outcome of identification attempts expressed as the percentage and number of witnesses (in parentheses) as a function of the presence of a weapon (N= 584).

	Suspect ID	Foil ID	No ID
Firearm	30.8% (4)	15.4% (2)	53.8% (7)
Knife	44.3% (47)	20.8% (22)	34.9% (37)
Other	41.2% (49)	11.8% (14)	47.1% (56)
Weapon present	42.0% (100)	15.9% (38)	42.0% (100)
Weapon absent	40.2% (139)	23.7% (82)	36.1% (125)

Note. The data were coded as weapon present or absent for analysis. Rows sum to 100%.

Table 9: The model parameters for a mixed effect multinomial model including the variable weapon present vs. absent.

Variable	Estimate	SE	z	*p*
Fixed effects				
$Intercept_s$	-0.20	0.29	-0.71	0.48
$Intercept_f$	-0.74	0.22	-3.39	<0.01
$Weapon_s$	0.14	0.32	0.44	0.66
$Weapon_f$	0.53	0.23	2.25	0.02
Random effects				
$Intercept_s$	2.11	0.33	6.43	<0.01
$Intercept_f$	0.50	0.26	1.93	0.03

Notes. Subscript s denotes parameters from the contrast between choosing the suspect and no identification; subscript f denotes parameters from the contrast between choosing a foil and no identification. Probability levels of fixed-effects are for two-tailed tests. The random effects intercept parameter is a standard deviation for the random normal distribution of intercept values. A standard deviation must be positive, therefore all random effects are evaluated by a one-tailed test.

Table 10: The outcome of identification attempts expressed as the percentage and number of witnesses (in parentheses) as a function of the race of the suspect (N= 535).

	Suspect ID	Foil ID	No ID
White European	46.3% (130)	19.9% (56)	33.8% (95)
Dark European	42.2% (35)	22.9% (19)	34.9% (29)
African - Caribbean	29.2% (50)	23.4% (40)	47.4% (81)

Note. Rows sum to 100%.

Table 11: The model parameters for a mixed effect multinomial model including the variables relating to the suspect.

Variable	Estimate	SE	z	*p*
Fixed effects				
$Intercept_s$	1.40	1.77	0.79	0.42
$Intercept_f$	-0.81	1.31	-0.62	0.53
Age_s	-0.21	0.25	-0.85	0.39
Age_f	0.28	0.16	1.71	0.09
$Race_s$	-0.55	0.19	-2.95	<0.01
$Race_f$	-0.05	0.15	-0.32	0.75
$Height_s$	0.17	0.56	0.30	0.76
$Height_f$	-0.12	0.31	-0.39	0.69
$Build_s$	-0.16	0.56	-0.28	0.77
$Build_f$	0.01	0.50	0.03	0.97
Random effects				
$Intercept_s$	1.84	0.32	5.70	<0.01
$Intercept_f$	0.42	0.27	1.58	<0.01

Note. Subscript s denotes parameters from the contrast between choosing the suspect and no identification; subscript f denotes parameters from the contrast between choosing a foil and no identification. Probability levels of fixed-effects are for two-tailed tests. The random effects intercept parameter is a standard deviation for the random normal distribution of intercept values. A standard deviation must be positive, therefore all random effects are evaluated by a one-tailed test.

Weapon presence

The number and outcome of identification attempts as a function of the presence of a weapon are shown in Table 8. The data were coded to identify the type of weapon. The weapons consisted mainly of knives or 'other' weapons. Few firearms were used in the offences. The presence or absence of a weapon was entered as a single variable in a model (see Table 9). Data on this variable were available from 584 witnesses and 295 lineups. The presence of a weapon had no effect on the likelihood of identifying a suspect, but it was significantly associated with witnesses being more likely to identify a foil. Witnesses were more likely to identify a foil if a weapon was not present.

Suspect variables

A model with four variables was calculated. The first was the age of suspect, recoded into 3 levels (< 20 years, 21-30, 31-40). Only 4 suspects were rated as being over 40 years, data from these lineups were excluded. The second variable was the race of the suspect. Only suspects of white European, dark European and African – Caribbean origin were included, as there were cells with few, if any, data for other race codes. The third factor was the build of the suspect coded into two categories, both thin and heavily built suspects were included in a 'non-average' category that was contrasted with 'average' build. The intention was to capture the distinctiveness of the culprit's appearance. Deviation from 'average' might provide a reason for the witness to be able to identify the culprit more easily. The fourth factor was height, which was coded in a similar manner that contrasted average with non-average. Data on these variables were available from 535 witnesses and 269 lineups. The only significant effect was an effect of suspect race on the likelihood of a positive identification of the suspect. The outcome of the lineups as a function of suspect race is shown in Table 10. The results of the model are shown in Table 11. African – Caribbean suspects were less likely to be identified than were European suspects ($p < 0.01$).

Race of witness and suspect

To test for an independent effect of cross-race recognition a model was calculated with witness race, suspect race and their interaction as explanatory variables. The cases were restricted to people of white European and African-Caribbean origin, because of small numbers in other categories. The effect of the interaction of race of witness and suspect from 558 witnesses and 285 lineups is shown in Table 12. The data show a trend in the predicted direction on suspect identification; numerically fewer suspects were identified by witnesses of a different race than by witnesses of the same race. The model confirmed a significant effect of suspect race on suspect identifications ($p<.02$) but no other significant effects.

Table 12: The outcome of identification attempts expressed as the percentage and number of witnesses (in parentheses) as a function of the difference in race of witness and suspect. (N= 558).

	Suspect ID	Foil ID	No ID
Different race	37.4% (120)	22.1% (71)	40.5% (130)
Same race	45.1% (107)	20.3% (48)	34.6% (82)

Notes. Cases are restricted to white European and African – Caribbean participants. Rows sum to 100%.

Witness Description

The effect of two variables relating to the witness' description of the culprit were analysed. One was a rating of the completeness of the description with three levels (few details, average, detailed). The other was a measure of whether the suspect's appearance did not match the witness's description on any aspect (all features match, at least one mismatch). A model with a single random factor that included both of these factors did not converge. Therefore, separate models were created for each variable. Valid data on the completeness was available from 530 witnesses and 281 lineups. The model for the completeness of description converged for a single random intercept and gave a better fit than the independence model. However, the model that allowed the intercept to vary between response category contrasts did not converge. The cross-tabulation of completeness of description with lineup outcome is shown in Table 13. The results of the equal intercept model for completeness of description is shown in Table 14. There was a significant effect on the likelihood of a positive identification of the suspect (p <.01). Two-thirds (66%) of witnesses who gave a detailed description of the culprit identified the suspect; 40% of witnesses who gave an average description identified the suspect, but only 14% of witnesses who gave few details identified the suspect. Valid data on whether there was a mismatch between the witness' description of the culprit and the appearance of the suspect were available from 569 witnesses and 286 lineups (Table 13). A separate model found no effect of the match of the description.

Table 13: The outcome of identification attempts expressed as the percentage and number of witnesses (in parentheses) as a function of the completeness of the witness description and the match with the suspect's appearance (N=569).

	Suspect ID	Foil ID	No ID
Completeness			
few details	14.3% (5)	20.0% (7)	65.7% (23)
average	40.4% (181)	21.4% (96)	38.2% (171)
detailed	66.0% (31)	12.8% (6)	21.3% (10)
Match to suspect			
mismatch	35.7% (50)	20.7% (29)	43.6% (61)
match	42.2% (181)	21.2% (120)	38.3% (218)

Note. Rows sum to 100%.

Table 14: The model parameters for a mixed effect multinomial model including the completeness of the description.

Variable	Estimate	SE	z	*p*
Fixed effects				
$Intercept_s$	-3.06	0.75	-4.09	<.01
$Intercept_f$	-1.75	0.80	-2.18	0.03
$Completeness_s$	1.59	0.37	4.33	<.01
$Completeness_f$	0.62	0.40	1.56	0.12
Random effects				
Intercept	1.84	0.32	5.70	<0.01

Notes. Subscript s denotes parameters from the contrast between choosing the suspect and no identification; subscript f denotes parameters from the contrast between choosing a foil and no identification. Probability levels of fixed-effects are for two-tailed tests. The random effects intercept parameter is a standard deviation for the random normal distribution of intercept values. A standard deviation must be positive, therefore all random effects are evaluated by a one-tailed test.

The identification procedure

The effects of three variables relating to the identification procedure itself were analysed. The first variable was the delay between the crime and the lineup (coded in 8 levels from 1 week to more than six months). The second variable was the speed with which the witness made their decision (fast, average, slow). The third variable was the difficulty that the identification officer reported in constructing the lineup (easy, average, difficulty). This factor was designed as a measure of the difficulty in selecting foils who resemble the suspect. Data were available from 558 witnesses and 282 lineups. The outcome of these identification attempts are shown as a function of the three identification procedure variables in Table 15. The result of the model is shown in Table 16. The only statistically significant effect was that of the witness' decision speed on the likelihood of making a positive identification of the suspect. Fast decisions are more likely to result in identification of the suspect (87%) than average or slow decisions (38% and 31% respectively)..

Table 15: The outcome of identification attempts expressed as the percentage and number of witnesses (in parentheses) as a function of the delay before the lineup, the speed of the witnesses decision, and the difficulty of constructing the lineup (N=558).

	Suspect ID	Foil ID	No ID
Delay			
< 1 week	65.5% (19)	10.3% (3)	24.1% (7)
< 1 month	34.1% (15)	25.0% (11)	40.9% (18)
< 2 months	41.8% (69)	18.2% (30)	40.0% (66)
< 3 months	39.3% (57)	22.1% (32)	38.6% (56)
< 4 months	35.2% (19)	24.1% (13)	40.7% (22)
< 5 months	43.8% (14)	15.6% (5)	40.6% (13)
< 6 months	46.7% (7)	13.3% (2)	40.0% (6)
> 6 months	39.2% (29)	27.0% (20)	33.8% (25)
Speed of decision			
fast	86.6% (33)	7.9% (3)	5.3% (2)
average	38.4% (179)	21.0% (98)	40.6% (189)
slow	31.5% (17)	27.8% (15)	40.7% (22)
Difficulty of constructing lineup			
difficult	60.0% (3)	40.0% (2)	0
average	39.8% (212)	21.0% (112)	39.2% (209)
easy	70.0 (14)	10.0 (2)	20.0 (4)

Note. Rows sum to 100%.

The global model

As a final stage of analysis explanatory variables that had a significant effect in any of the models described above were included in a single model. This final procedure was undertaken so that the variables which had independent effects could be identified. The variables included were witness age, witness gender, duration of witness view, completeness of witness description, weapon presence, suspect race, witness decision speed. Cases were restricted to suspects unknown to the witness, to white European and African – Caribbean suspects and to cases in which data were available on all explanatory variables (479 witnesses and 261 lineups). A model with an equal random intercept for both contrasts of the response variable gave a better fit to the data (log likelihood = -466.5) than the independence model with no random effect ($\chi^2(1) = 17.60$, $p < .005$). A model that allowed the random intercept to vary between response variable contrasts did not converge. The results of the equal random intercept model are shown in Table 17. There was a significant effect on the likelihood of a suspect identification of witness age, suspect race, completeness of witness description and decision speed. The effect of duration of view approached statistical significance (p = .09) with a two-tailed test. The effect is significant with a one-tailed test, which would be appropriate in view of the a priori nature of the predicted direction of the effect of this variable. In summary, the suspect is more likely to be identified if the witness is younger than 30, the suspect is a white European (rather than African - Caribbean), the witness viewed the culprit for more than a minute, gave a detailed description, and made a fast decision at the lineup. None of the explanatory variables was significantly associated with a mistaken identification of a foil.

Table 16: The model parameters for a mixed effect multinomial model including the delay before the lineup, decision speed and the difficulty in constructing the lineup.

Variable	Estimate	SE	z	*p*
Fixed effects				
$Intercept_s$	1.31	2.16	0.60	0.54
$Intercept_f$	0.60	1.67	0.36	0.72
$delay_s$	-0.01	0.10	-0.09	0.93
$delay_f$	0.06	0.07	0.80	0.42
$speed_s$	-1.58	0.43	-3.69	<0.01
$speed_f$	-0.03	0.33	-0.08	0.93
$difficulty_s$	0.90	0.98	0.92	0.35
$difficulty_f$	-0.59	0.76	-0.77	0.44
Random effects				
$Intercept_s$	2.24	0.36	6.18	<0.01
$Intercept_f$	0.54	0.26	2.05	0.02

Note. Subscript s denotes parameters from the contrast between choosing the suspect and no identification; subscript f denotes parameters from the contrast between choosing a foil and no identification. Probability levels of fixed-effects are for two-tailed tests. The random effects intercept parameter is a standard deviation for the random normal distribution of intercept values. A standard deviation must be positive, therefore all random effects are evaluated by a one-tailed test.

Discussion

The data reported here provide an appropriate context in which to evaluate the external validity of empirical findings, theoretical predictions and legal assumptions on the outcome of real identification attempts conducted in the course of criminal investigations. The interpretation of these data raise four issues.

First, the identity of the suspect is not necessarily the same as the culprit. We do not know how many lineups contain the culprit. Indeed, the purpose of organising the lineup is to collect evidence of whether the suspect is the culprit. The large number of lineups in the sample will help to dilute the effect of ‘culprit-absent’ lineups if these are relatively few in number.

Second, there are a number of potential sources of error in the coding of the explanatory variables. For example, the witness may estimate viewing conditions inaccurately. The investigating officer may recall information incorrectly or make an error or omission in filling in the form. Errors of this nature are an inevitable feature of a study such as this. The effect of this issue and the first issue, described above, will be to reduce the effect of explanatory variables. The extent to which results are consistent with empirical findings from laboratory studies and confirm theoretically-derived predictions will contribute to confidence in the interpretation of the results.

Table 17: The model parameters for the global mixed effect multinomial model including explanatory variables that had a significant effect in the models reported above (N=479).

Variable	Estimate	SE	z	*p*
Fixed effects				
$Intercept_s$	-1.10	1.30	-0.85	0.40
$Intercept_f$	-3.45	1.51	-2.29	0.02
Witness age_s	-0.38	0.12	-3.20	<0.01
Witness age_f	-0.05	0.15	-0.35	0.73
Witness $gender_s$	0.28	0.31	0.91	0.36
Witness $gender_f$	0.47	0.34	1.37	0.17
Duration of $view_s$	0.48	0.28	1.69	0.09
Duration of $view_f$	0.11	0.31	0.35	0.73
$Weapon_s$	0.07	0.30	0.24	0.81
$Weapon_f$	0.29	0.35	0.82	0.41
Suspect $race_s$	-0.45	0.15	-2.90	<0.01
Suspect $race_f$	-0.12	0.18	-0.67	0.50
Description $completeness_s$	1.64	0.44	3.75	<0.01
Description $completeness_f$	0.84	0.48	1.74	0.08
Decision $speed_s$	-0.81	0.37	-2.19	0.03
Decision $speed_f$	0.27	0.40	0.66	0.51
Random effects				
Intercept	1.03	0.29	3.50	<0.01

Notes. Subscript s denotes parameters from the contrast between choosing the suspect and no identification; subscript f denotes parameters from the contrast between choosing a foil and no identification. Probability levels of fixed-effects are for two-tailed tests. The random effects intercept parameter is a standard deviation for the random normal distribution of intercept values. A standard deviation must be positive, therefore all random effects are evaluated by a one-tailed test.

Third, the analyses only show correlational relationships. The usual warning that correlation does not establish causality should be borne in mind. There will be many extraneous factors that influence the outcome of a lineup that was not coded in the questionnaire (e.g. the attention paid to the culprit by the witness, individual differences in face recognition ability, whether the culprit is present in the lineup, the fairness of the lineup).

Fourth, the appropriate caution needs to taken when interpreting negative results. The possibility cannot be excluded that a factor that fails to reach a conventional level of significant in the present study, does not genuinely affect eyewitness identification accuracy.

None of the explanatory variables was significantly associated with the identification of a foil in the final model calculated. This result could be interpreted as evidence that foil identifications, although fairly common, are driven by idiosyncratic factors that are not coded in the data set; for example the resemblance of a foil to a culprit in a culprit-absent lineup. It is noteworthy that, in most of the analyses reported, the model fit was significantly improved by including a separate random intercept term for the contrast between identifying a foil and making no identification (the relevant random intercept parameter was typically significant or near significant). This implies that there was significant variance in the identification of a foil that was associated with the lineup from which the identification was made. We cannot tell which lineup features are systematically associated with this variance except that it is not related to any of the lineup variables included in our questionnaire. Lineup fairness is a possible explanatory variable (Smith, Lindsay & Pryke, 2000).

A significant effect of witness gender, whereby female witnesses were more likely to mistakenly identify a foil than were males, and a significant effect of the presence of a weapon, whereby foils were more likely to identified if a weapon was absent, were found in the exploratory models but were not significant in the global model. There is some laboratory evidence consistent with the effect of witness gender observed. Shapiro and Penrod's (1986) meta-analysis of face recognition studies found that gender had a small effect on recognition accuracy. Females were more likely to make correct identifications (effect size = 0.10) but were also more likely to make false positives (effect size = 0.08). Thus the tendency for females to be more likely to make positive identifications in laboratory studies has been supported by the present analysis of real eyewitness identifications, although it should be noted that no effect of gender on suspect identifications was found in our data. Of course suspect identifications cannot be directly compared to correct identifications in laboratory studies for the reasons discussed above. The effect of the presence of a weapon that was observed was not the predicted effect. It should be noted that it was seldom the case the weapon was a gun. As weapon focus is usually operationalized by simulating an assailant with a gun, the data from the present study may address a rather different issue. As neither the effect of weapon focus or witness gender was significant in the global model, the evidence for these effects in the present study is inconsistent. The global model was based on fewer observations, due to the cumulative effect of missing data on a larger number of explanatory variables. It is possible that the effects were associated with a small number of cases that were dropped in the final analysis. Alternatively, it is possible that each of these explanatory variables was associated with other variables that were included in the global analysis and so the independent effect of the variables was reduced in the global analysis.

The completeness of the witness' description would be significantly associated with the identification of a foil by the criterion of a one-tailed test. Inspection of Table 13 reveals that the main driver of this effect is the high proportion of 'no identifications' when the witness gave few details in their description of the culprit. This result is entirely consistent with a 'common sense' prediction. We return to the effect of the completeness of the witness description below.

All but one of the factors that were significantly associated with a positive suspect identification were consistent with previous laboratory work. Suspects were more likely to be identified by young witnesses than by older witnesses (Table 3). Some laboratory studies have found that older eyewitnesses make fewer correct responses in tests of face recognition (e.g. Bartlett & Fulton, 1991; O'Rourke, Penrod & Cutler, 1989). Older eyewitnesses have been found to be less likely to make accurate identification choices in an eyewitness identification context (e.g. Memon and Bartlett, in press). The effect of age in lineup choice has been found in terms of both reduced correct identifications and increased mistaken identifications. Indeed, research has been directed towards understanding the increased false identification rate in older people (Searcy, Bartlett & Memon, 1999; 2000; Searcy, Bartlett, Memon & Swanson, in press). The present data confirm the laboratory findings inasmuch as older witnesses in real cases showed a lower positive identification rate of the suspect, but there was no significant effect of the age of the witness on the number of mistaken identification of foils.

O'Rourke et al. (1989) found that identification accuracy declined sharply at around age 50. For the purposes of statistical analysis age of witness in the present study was recoded from nine to four categories. The statistical analysis shows a decline in suspect identifications after age 40. Inspection of the original data (see text of 'Results' section above) suggests that a slow decline in the number of positive suspect identifications begins as early as age 30.

By the criterion of a one-tailed test witnesses who viewed the culprit for longer were more likely to make a positive identification of the suspect. In their meta-analysis of face recognition studies, Shapiro and Penrod (1986) found a significant effect of exposure duration at study on the number of correct identifications and no effect on false positive rate. Therefore, the effect of duration of view in the present lineup data is consistent with the findings of laboratory studies, despite marked differences in the circumstances of viewing and the absolute duration considered. Viewing conditions in the laboratory studies are likely to be good, the duration is considerably shorter than one minute and of course photographs would have served as stimuli.

Relatively few decisions were coded as 'fast' by identification officers, but fast choosers were much more likely to make a positive identification of the suspect (87% vs 38% of witnesses who chose with average speed). In a series of studies Sporer (1992, 1993, 1994) has found that fast identification decisions are associated with accurate choices. The present data confirms that fast choosers tend to be accurate in real world data.

Witnesses who gave relatively detailed descriptions of the culprit were more likely to make a positive identification on the suspect. This finding is intuitively appealing. It suggests that the ability to verbally describe a person is positively associated with the ability to recognise them in a lineup. The witnesses in the lineups will have viewed the culprits under a wide range of viewing conditions. Therefore, the completeness of their verbal description may reflect the opportunity that the suspect had to view the culprit, together with some individual differences in face recognition or visual attention.

African – Caribbean suspects were less likely to be identified than were European suspects. This result is in marked contrast to Wright and McDaid's (1996) survey of the outcome of London identity parades. They found that white European suspects were less likely to be identified than suspects from ethnic minorities. It is not clear why African – Caribbean suspects should be less likely to be identified than white European suspects or vice versa. The effect is not attributable to differences in the race of the suspect and the witness because the interaction between race of witness and suspect was analysed separately and showed no significant effect.

Bearing in mind the caution given above, it is interesting to consider the explanatory variables that were not significantly associated with suspect identifications. There was no significant effect of delay. Inspection of Table 15 reveals that the outcome of lineups organised from approximately 1 month to more than six months after the incident is remarkably consistent. Positive identifications of the suspect range from 34% - 46%, foil identifications range from 13% - 27% and no identification was made in 34% - 40% of cases. There is no hint of a decrease in suspect identification, a rise in foil identification or a rise in 'no identification' responses over increasing delay. However, lineups organised within a week of the offence yield more positive identifications of the suspect (65%). The present results are consistent with those reported by Shepherd (1983). He found that the proportion of correct identifications fell from 65% after 1 week, to 55% after a month, 50% after 3 months and 10% after 11 months. Shepherd found no effect of delay on false identifications (15%, 20%, 20% & 15% respectively). Our data also show little effect of delay on foil identifications. Shepherd found little effect of delay on either correct identifications or mistaken identifications in experiments that compared performance after a delay of 1 month with a delay of 4 months. Similarly, our data show a lack of effect of delay on either suspect or foil identifications over this period. Other studies that have shown an effect of delay have included a condition that has a short delay of less than one week. (e.g. Malpass & Devine, 1981; Krafka & Penrod, 1985; Davies, Ellis & Shepherd, 1978). Shapiro and Penrod (1986) found an effect of delay in a meta-analysis of 18 face recognition and eyewitness identification studies on both correct identifications (effect size = 0.43) and mistaken identifications (effect size = 0.33). The delay of the studies included in the analysis was a mean of 4.5 days with a standard deviation of 21 days. Therefore, most studies probably included a short delay of less than a week as one of the experimental conditions. In summary, a significant effect of delay found in experimental studies is consistent with the present data that are suggestive of a fall off in identification performance between a delay of a week and a month, notwithstanding the lack of a significant effect of delay in the present analysis.

Neither witness race nor the role of the witness showed any statistically significant effect on the outcome of the lineups. In contrast, Shapiro and Penrod's (1986) meta-analysis found that black participants made more correct identifications than did white participants (effect size = 0.17). There is no variable comparable to witness role in the experimental studies analysed by Shapiro and Penrod.

Laboratory studies have established that faces of the same-race as the observer are recognised less accurately than are same-race faces (e.g. Chiroro & Valentine, 1995; Valentine & Endo, 1992). Shapiro and Penrod (1986) and Bothwell, Brigham and Malpass (1989) both found a significant effect of cross-race recognition in meta-analyses of the literature. However, no effect of different vs. sane race of suspects and witnesses was found in the current data.

There was no significant effect of offence on suspect identifications. Offences that involved an assault on the person tended to result in a slightly higher proportion of suspect identifications (49% vs 41% mean for all offences). Within this group suspect identification for rape was exceptionally high (90% of cases in which the culprit was unknown). Identification of murderers was the lowest (29%) in this category for obvious reasons, although it should be noted that attempted offences are included in the sample. For all offences other than assaults, between 41% and 27% of suspects were identified. Thankfully, murders and rapes accounted for a relatively small number of offences in the sample. Therefore, the substantial differences in identification rates between these offences are not sufficient to produce a significant effect of offence on suspect identifications across the whole sample.

The presence of a weapon had no effect on the likelihood of identifying a suspect. This result contrasts with the literature in which the evidence suggests that number of identifications of the culprit is reduced by the presence of a weapon.

In summary, the results of this study are generally compatible with previous laboratory-based work and strengthen the external validity of some factors known to affect identification accuracy in experimental studies. These factors include the age of the witness, the duration for which the culprit was viewed, and the speed of their identification at the lineup. This outcome boosts confidence both in the relevance of the laboratory data and in the quality of the data collected from identifications carried out as part of criminal investigations. Equally the results serve to suggest that other laboratory findings, may well be reliable in the laboratory but are too weak to exert a significant influence in the body of cases analysed. Such factors include the presence of a weapon, the delay of identification beyond 1 month, cross-race identification. This conclusion may suggest that such factors may warrant less research attention from an applied perspective.

The current data were drawn from an original sample of over 300 lineups. Although this is a substantial sample it would not be prudent to base policy recommendations on single study. Therefore, one should look for converging evidence from a number of studies of data from real cases. Our results are mostly compatible with studies reported by Wright and McDaid (1996) and Tollestrup et al. (1994), although neither of these studies investigated such a wide range of explanatory variables. In a brief report of an analysis of 1776 identification attempts conducted by police in England and Wales, Pike et al. (2002) report results that are strikingly similar to those of the present study. They found a significant effect of the age of the witness, showing a decline in identifications of the suspect with age. Performance began to deteriorate after age 35. Similarly they found no effect of the presence of weapon, the use or threat of violence, or of cross-race vs. same-race identifications. The one result from the present study that was unexpected is the effect of suspect race. This variable had a strong effect on the probability of identifying the suspect but in the opposite direction to the effect found by Wright and McDaid (1996) in identification data collected in the London area. There is no clear reason for this difference and is worthy of further investigation.

Implications for policy.

The analysis of lineups reported has a number of practical implications for the development of evidence-based guidance for the courts, the law and the police. The courts and jury members may base their evaluation of eyewitness identification evidence on a number of assumptions. In English law the Turnbull guidelines have enshrined some factors as those that need to be considered. Only two of these factors had a significant effect in predicting a positive identification of a suspect: whether the suspect was known and the duration of view. Instead of the remaining factors our study suggests that jurors should consider the age of the witness, the completeness of the witness's description, and the speed with which the witness made the identification.

English law [Police and Criminal Evidence Act (1984), Codes of Practice (2002)], requires that witnesses inspect the entire lineup (whether live or on video) twice before making their choice. This requirement is intended to ensure that each member of the lineup is inspected before a decision is made. The mounting evidence that a fast identification is more likely to reflect an accurate choice, suggests that this requirement may be unnecessary or perhaps even counter-productive.

In English law, the suspect has a right for a representative to be present at a lineup. If no representative is in attendance the identification attempt must be video taped. The intention is to ensure that the identification officer does not lead the witness. If all identification attempts were recorded, the video could be used to record the speed of identification. This would allow a useful source of evidence to be available to the court.

If the police have a number of witnesses and cannot identify a suspect, they may ask one witness to attempt to construct a likeness using a face reconstruction system (e.g. e-fit), or search a file of photographs of people previously convicted for similar offences. A concern is that either of these procedures may contaminate the witness' memory. Therefore, the police often want to save their 'best' witness to view a subsequent identity parade and use another witness to construct an e-fit or search mug-shots. The results of the present study provide evidence on which to base the choice of the 'best' witness. Our results suggest that a positive suspect identification in a lineup is most likely to be made by a witness who gave a detailed description of the culprit, viewed the culprit for more than a minute and is under 30.

Live lineups, which are the most common form of identification in the United Kingdom, are expensive to organise (c. £800 each). Many cases require an identification attempt for legal reasons. However, in some cases a decision must be made as to whether an identification should be attempted. Our results would be helpful to manage an expensive resource effectively by providing some indication of the cases that are more to result in a positive identification.

In the United States legislation on identification procedures is less proscriptive than in England and Wales, and differs between states (Lipton, 1996). The American Psychology - Law Association have recently proposed four rules of good practice for identification procedures (Wells et al. 1998). Davies and Valentine (1999) compare these rules with English legal provision. The major difference between identification procedures in the USA and England is the widespread use of photospreads rather than live lineups in the USA. Such differences in test media have relatively little influence of identification performance (Cutler, Berman, Penrod & Fisher, 1994). Although it is likely that the results reported here would generalise to American identification procedures, it would be prudent to attempt a replication in the appropriate jurisdiction.

References

Bartlett, J. C. & Fulton, A. (1991). Familiarity and recognition of faces in old age. Memory & Cognition, 19, 229-238.

Bothwell, R. K., Brigham, J. C. & Malpass, R.S. (1989). Cross-racial identification of faces. Personality and Social Psychology, 15, 19-25.

Brigham, J. C., Meissner, C. A. & Wasserman, A. W. (1999). Applied issues in the construction and expert assessment of photo lineups. Applied Cognitive Psychology, S73-S92.

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.

Chiroro, P. & Valentine, T. (1995). An investigation of the contact hypothesis of the own-race bias in face recognition. Quarterly Journal of Experimental Psychology, 48A, 879-894.

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

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

Cutler, B. L., Penrod, S. & Martens, T. K. (1987). The reliability of eyewitness identifications: The role of system and estimator variables. Law and Human Behaviour, 11, 223-258.

Cutler, B. L., Penrod, S. O'Rouke, T. E. & Martens, T. K. (1986). Unconfounding the effects of contextual cues on eyewitness identification accuracy. Social Behaviour, 1,113-134.

Davies, G. M. Ellis, H. D. & Shepherd, J. W. (1978). Face recognition accuracy as a function of made of representation. Journal of Applied Psychology, 63, 180-187.

Davies, G. M. & Valentine, T. (1999). Codes of practice for identification. Expert Evidence, 7, 59-65.

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

Fallshore, M. & Schooler, J. W. (1995). The verbal vulnerability of perceptual expertise. Journal of Experimental Psychology: Learning, Memory and Cognition, 21, 1608-1623.

Hedeker, D. (1999). MIXNO: A computer program for mixed-effects nominal logistic regression. URL: **http://www.uic.edu/~hedeker/mix.html**

Krafka, C. & Penrod, S. (1985). Reinstatement of context in a field experiment on eyewitness identification. Journal of Personality and Social Psychology, 49, 58-69.

Linsday, R. C. L. (1999). (guest ed.) Measuring lineup fairness. Applied Cognitive Psychology 10 (special issue),

Lipton, J. P. (1996). Legal aspects of eyewitness testimony. In Sporer, S., Malpass, R. S. & Koehnken, G. (eds.) Psychological issues in eyewitness identification. Mahwah, NJ: Lawrence Earlbaum Associates. (pp. 7 – 22).

Loftus, E. F., Loftus, G. R. & Messo, J. (1987). Some facts about "weapon focus". Law and Human Behavior, 11, 55-62.

Malpass, R. S. & Devine, P. G. (1981). Eyewitness identification: Lineup instructions and the absence of the offender. Journal of Applied Psychology, 66, 482-489.

Meissner, C. A. & Brigham, J. C. (2001). A meta-analysis of the verbal overshadowing effect in face identification. Applied Cognitive Psychology, 15, 603-616.

Memon, A. & Bartlett, J. C. (in press). The effects of verbalisation on face recognition in young and older adults. Applied Cognitive Psychology,

O'Rourke, T. E., Penrod, S. D. & Cutler, B. L. (1989). The external validity of eyewitness identification research: Generalizing across subject populations. Law and Human Behavior, 13, 385-397.

Pike, G., Brace, N. & Kyman, S. (2002). The visual identification of suspects: procedures and practice. Briefing note 2/02 Policing and Reducing Crime Unit, Home Office Research Development and Statistics Directorate. (**www.homeoffice.gov.uk/rds/prgbriefpubs1.html**)

Police and Criminal Evidence Act 1984 (s.60(1)(a)) Code of Practice (D) on identification of persons. Section 2: Identification by witnesses (2002). London: Home Office Communication Directorate.

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

Ross, D. F., Read, J. D. & Toglia, M. P. (eds.) Adult eyewitness testimony: Current trends and developments. Cambridge: Cambridge University Press.

Searcy, J. H., Bartlett, J. C. & Memon, A. (1999). Age differences in accuracy and choosing in eyewitness identification and face recognition. Memory & Cognition, 27, 538-552.

Searcy, J. H., Bartlett, J. C. & Memon, A. (2000). Relationships of availability, lineup conditions, and individual differences to false identification by young and older eyewitnesses. Legal and Criminological Psychology, 5, 219-236.

Searcy, J. H., Bartlett, J. C., Memon, A. & Swanson, K. (in press). Ageing and lineup performance at long retention intervals. Effects of metamemory and context reinstatement. Journal of Applied Psychology, Schooler, J. W. & Engstler- Schooler, T. Y. (1990). Verbal overshadowing of visual memories: some things are better left unsaid. Cognitive Psychology, 22, 36-71.

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

Shepherd, J. W. (1983). Identification after long delays. In: S. M. A. Lloyd-Bostock & B. R. Clifford (Eds.) Evaluating eyewitness evidence. Chichester: Wiley (pp. 173-187).

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

Smith, S. M., Lindsay, R.C. L. & Pryke, S. (2000). Postdictors of eyewitness errors: Can false identifications be diagnosed? Journal of Applied Psychology, 85, 542-550.

Sporer, S. (1992). Post-dicting accuracy: Confidence, decision times and person descriptions of choosers and non-choosers. European Journal of Social Psychology, 22, 157-180.

Sporer, S. (1993). Eyewitness identification accuracy, confidence and decision times in simultaneous and sequential lineups. Journal of Applied Psychology, 78, 22-33.

Sporer, S. (1994). Decision times and eyewitness identification accuracy in simultaneous and sequential lineups. In: Ross, D. F., Read, J. D. & Toglia, M. P. (eds.) Adult eyewitness testimony: Current trends and developments. Cambridge: Cambridge University Press. (pp. 300 – 327).

Sporer, S., Malpass, R. S. & Koehnken, G. (1996). Psychological issues in eyewitness identification. Mahwah, NJ: Lawrence Earlbaum Associates

Steblay, N. M. (1992). A meta-analysis review of the weapon effect. Law and Human Behaviour, 16, 413-424.

Tollestrup, P. A., Turtle, J. W. & Yuille, J. C. (1994). Actual victim and witnesses to robbery and fraud: An archival analysis. In: Ross, D. F., Read, J. D. & Toglia, M. P. (eds.) Adult eyewitness testimony: Current trends and developments. Cambridge: Cambridge University Press. (pp. 144 – 160).

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

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

Valentine, T. & Heaton, P. (1999). An evaluation of the fairness of police lineups and video identifications. Applied Cognitive Psychology, 13, S59-S72.

Valentine, T. Harris, N., Colom Piera, A. & Darling, S. (in press). Are police video identifications fair to African – Caribbean suspects? 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. (1999). Measuring the goodness of lineups: Parameter estimation, question effects, and limits to the mock witness paradigm. Applied Cognitive Psychology, 13, S27-S39.

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.

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

Yuille, J. C. (1993). We must study forensic eyewitnesses to know about them. American Psychologist, 48, 572-573.

Appendix

Selection of the appropriate model requires a comparison between the ability of two models M_1 and M_2 to account for the data. Assume that M_1 has n parameters more than M_2. The selection of the appropriate model is based upon the log-likelihood ratio statistic (λ).

$$\lambda = 2(LL_1 - LL_2)$$

in which LL_1 is the log-likelihood of M_1 and LL_2 is the log-likelihood of M_2. λ is distributed as χ^2 with n degrees of freedom. If λ is significant, then the more parsimonious model, M_2, can be rejected in favour of M_1.