

Applied Issues in the Construction and Expert Assessment of Photo Lineups

JOHN C. BRIGHAM,* CHRISTIAN A. MEISSNER
and ADINA W. WASSERMAN

Florida State University, USA

SUMMARY

Issues surrounding lineup fairness have been explored scientifically for over two decades. The present study investigates the applied/external validity of this line of research. First, several factors leading to bias in the construction of photo lineups are examined, and results of a preliminary survey on current law-enforcement practices are presented. Several statistics that have been developed to assess the fairness of lineups are reviewed and the application of these techniques to lineups used in 18 criminal cases is discussed, including the mixed agreement that sometimes occurs between estimates. Finally, we address the usefulness of lineup fairness assessment for expert testimony in the courtroom, and the dilemma that may be faced by the expert witness who is asked to testify by the defence. It is suggested that a useful and empirically justified index of overall lineup fairness can be created by combining a single estimate of bias (Functional Size) and a single estimate of lineup size (Effective Size) into a four-point index. Copyright © 1999 John Wiley & Sons, Ltd.

The 'fairness' of a lineup is a crucial piece of information that is taken into account when evaluating the likely validity of a disputed eyewitness identification. The concept of fairness may be encountered at the time of construction by law-enforcement officers, or subsequently in court when the defence may raise it as an issue of justice. Before the 1970s, the only way that lineup fairness could be assessed was by 'eyeballing' and making a purely subjective, global judgement. The first empirical test of lineup fairness was proposed by Doob and Kirschenbaum in 1973. Since that time, several additional estimates of lineup fairness have been developed. Much debate has surrounded the usefulness of these statistics, especially in relation to their use by expert witnesses in actual cases. This paper will begin by reviewing factors that introduce bias in lineups at the construction phase, and preliminary data on a survey of current police procedures will be discussed. Following this, the empirical assessment of fairness indices will be examined, after which we will describe the application of these techniques to 18 criminal cases that involved disputed eyewitness identifications. Finally, the dilemma faced by the expert witness will be addressed.

*Correspondence to: John C. Brigham, Department of Psychology, Florida State University, Tallahassee, FL 32306-1270, USA. E-mail: brigham@psy.fsu.edu

THE CONCEPTS OF LINEUP SIZE AND LINEUP BIAS

In theory, a lineup is seen as fair to the suspect when it contains a sufficient number of distractors (foils) who are similar in appearance to the general description of the criminal (Doob and Kirschenbaum, 1973). Two dimensions of lineup fairness have been proposed (Malpass, 1981; Malpass and Devine, 1983). First, the concept of *lineup size* suggests that a lineup should be large enough that the probability of a chance identification of an innocent suspect is relatively low. In the USA lineups containing six members (the suspect and five foils) have generally been viewed as the minimally acceptable lineup size. But not just any six-person lineup is of acceptable size. It is critical that the five foils be reasonably plausible alternatives for the choice task (i.e. similar in appearance to the criminal's description). For example, if three of the foils are so dissimilar from the prior description of the perpetrator that they do not represent viable alternatives, then it is as if the eyewitness were in reality facing a three-person lineup containing the suspect and the two remaining foils.

Lineup bias, in contrast, refers to the extent that the suspect is distinctive from the other lineup members. Anything which causes the suspect to stand out from the other lineup members compromises the validity of the lineup, as this distinctiveness may be used by an uncertain eyewitness as a cue on which to base the identification decision. These can involve personal characteristics of the target as well as characteristics of the lineup itself. Examples might include aspects of the suspect's physical appearance (e.g. a blond suspect among five dark-haired foils, a heavyset suspect among five thin foils, etc.), or of his photo (e.g. its size, texture, background). Any such factor that sets the defendant apart from other lineup members in a systematic way compromises the validity of the lineup.

One way to describe this issue is to distinguish between positive and negative bias. *Positive bias* represents a situation in which the suspect is likely to be selected from the lineup by an eyewitness even if he or she has a poor memory of the suspect, because the suspect or his photo is distinctive in some way. Conversely, *negative bias* occurs when the suspect is relatively unlikely to be selected by an eyewitness, even if the eyewitness has a good memory of the event, because the other lineup members are so similar to him (i.e. 'clones'). Presumably, the lineup task will be most veridical as a measure of memory when neither type of bias is present. Research to date has focused most upon ways in which to assess the degree of positive bias in lineup construction, administration, and assessment.

The concepts of lineup size and bias may also overlap with each other. For example, if the suspect is the only lineup member who fits the description of the criminal, the lineup is of inadequate size because the suspect is the only viable choice, in terms of the description of the culprit. In a psychological sense, the size of this lineup is not six members, it is one. From the eyewitness's perspective, this situation is akin to the showup procedure wherein a suspect is shown alone to the eyewitness who is asked, 'Is this the guy?' US courts have consistently ruled that showups are impermissibly suggestive because the chance of misidentification is too great (Brigham, 1989).

CONTROLLING BIAS IN LINEUP CONSTRUCTION

The genesis of bias in any lineup is at the point of construction. Researchers have noted many factors that appear to influence the subsequent veridicality of a lineup

administration. As mentioned previously, optimal fairness would be represented in a photo array consisting of neither positive nor negative bias toward the suspect. In other words, what is desired is a 'reasonable' test of the witness's memory for the target; a test in which the suspect is neither too distinctive nor one of several 'clones' in the array.

Investigator bias and the effect of race

One source of bias may involve the sentiments of the investigator in charge of the case, as often it is this individual who constructs the photo array. While the importance of creating an unbiased array is undoubtedly considered, pressures to obtain a positive identification from the witness may counter this notion, and thereby influence the degree of difficulty created in the lineup. While there is no empirical evidence as to the validity of this supposition, variation in the bias of lineups used in real cases indicates the possibility of such a factor.

Additionally, law officers' race appears to impact their evaluations of lineups in a systematic way. In a study in which Brigham and Brandt (1992) asked officers to rate the 'usefulness' of 23 lineups ranging on a continuum from fair to unfair, White officers found the lineups more useful as a whole (68 per cent of the time they were classed as useful) than did Black officers (56 per cent useful; z for proportions = 3.86, $p < 0.001$). The effect of officers' race was most evident on evaluations of Black lineups, which were rated as useful more often by the White officers (64 per cent of the time) than by the Black officers (49 per cent of the time; z for proportions = 3.22, $p < 0.01$).

This finding is somewhat consistent with an earlier finding by Brigham and Ready (1985), who found that individuals made finer distinctions in evaluating lineups containing members of their own race compared with lineups containing other-race individuals. Both Whites and Blacks found it easier to locate 'similar' foils for other-race lineups than for same-race lineups, presumably because the 'own-race bias' that occurs in facial memory (Bothwell *et al.*, 1989) had an effect here. The stereotypic perception that 'they [other-race persons] all look alike' apparently operates in lineup construction as well as in recognition memory.

Recommendations for guarding against investigator bias in lineup construction appear fairly straightforward. An investigator who feels a strong sense of immediacy to obtain a positive identification from an eyewitness should not participate in construction of the photo array. Likewise, an officer of the same race as the suspect should construct the lineup to guard against the selection of dissimilar foils that may occur in cross-race comparisons.

Further, as pointed out by Wells *et al.* (1998), the person who administers the lineup should not know which lineup member is the suspect. Thus, in experimental terms, a *double-blind procedure* should be used in which both the administrator of the lineup and the eyewitness are unaware of whether the suspect is present in the photo array. Such a procedure also appears to protect against 'confirmatory feedback' from the person administering the lineup, which Wells and Bradfield (1998) have demonstrated can dramatically affect a witness's reconstruction of the events and later testimony regarding their identification of the suspect.

Distinctiveness of suspect

A second concern revolves around the issue of suspect distinctiveness. As noted by Brigham *et al.* (1990), it can be an arduous task to assemble a lineup that contains a

suspect and an adequate number of fair foils. This task is made even more difficult when the suspect is distinctive or unusual in appearance. Indeed, this situation occurred in the Brigham *et al.* (1990) study, where the researchers were unable to compose a fair lineup for one very distinctive target person.

In *Neil v. Biggers* (1972), one of a handful of US Supreme Court cases that created case law on the utilization of eyewitness evidence, law officers used a one-person 'showup' as the identification task, rather than the usual six- or eight-person lineup. The police claimed that they had earlier checked the city jail and the city juvenile home for persons to serve as foils in a live lineup, but could find no one at either place fitting the suspect's unusual physical description. The US Supreme Court upheld the conviction, ruling that while the showup procedure may have been suggestive, under the 'totality of circumstances' standard the victim's identification of the suspect was reliable and was properly allowed to go to the jury. Among the circumstances enumerated by the Court was the fact that the victim earlier failed to make any positive identifications from 30 to 40 photographs that comprised several photo lineups and photo showups, and that she had expressed 'no doubt' of the correctness of her identification from the final live showup.

Target/foil similarity and bias

One final issue involves the degree of similarity between lineup foils and the target. It seems logical to assume that increasing the degree of target-foil similarity would decrease the chance of positive bias towards the suspect (as defined earlier), but that too much similarity would create negative bias, making the identification task too difficult. Interestingly enough, though, there is some evidence that increased target-foil similarity may not increase negative bias. Laughery *et al.* (1988) reported results from a study in which five computerized foils were created by varying only one facial feature from the original target. Results indicated that positive bias was still present when using these high-similarity foils, as the target was chosen by mock witnesses as the most 'familiar' of the faces at a rate greater than chance. Similar results were obtained using actual photographs of faces, in which the lineup was created by matching foils to the suspect (Wogalter *et al.*, 1992).

These results suggest that a positive bias, and not a negative bias, may occur from creating a lineup in which the foils are very similar to the target. Alternative methods to creating lineups in which foils were chosen based on similarities to both the target and other foils produced significantly less biased lineups (Marwitz and Wogalter, 1988; Wogalter *et al.*, 1991, 1992). Based upon this evidence, Koehnken *et al.* (1996) suggested that the similarity-fairness function appears to take on an inverted U-shape, such that both high and low extremes of similarity can create a positive bias. Fairness lies somewhere in between, and probably, according to Koehnken *et al.*, more toward the high-similarity end of the spectrum.

Luus and Wells (1991) argued for an alternative method in which foils should be chosen based upon their similarity to the witness's description of the suspect, and not the appearance of the suspect, *per se*. Such a method relies upon the knowledge the witness held at the time of the incident while allowing for variation on features unmentioned, later termed 'propitious heterogeneity' (Wells *et al.*, 1993, 1994). Luus and Wells argued that a lineup which includes foils that differ on these unmentioned

characteristics would increase the likelihood of correct identifications while not increasing misidentifications in a target-absent lineup.

However, Koehnken *et al.* (1996) pointed out several opportunities for error in the use of this match-to-description procedure. First, descriptions of faces are notoriously vague or exceedingly general, making the interpretation of the description a subjective task. Furthermore, if a witness is unable to specifically describe characteristics of a suspect, this may increase the chance that the description may be misinterpreted by the officer constructing the lineup. Another potential problem is the inclusion of foils that fit the description, but do not resemble the suspect in any way. Finally, complications may also arise if different eyewitnesses provide dissimilar descriptions of the criminal (e.g. see Table 2). This would entail constructing a separate lineup for each eyewitness's description, an additional step that law officers might be reluctant to take.

To compensate for the problems of using either method exclusively (i.e. matching-to-target or matching-to-description), Koehnken *et al.* (1996) argued for a two-part procedure in which foils are first selected on the basis of objective features that are generally stated in the witness's description (i.e. height, weight, race, build, hair style, etc.). Next, available foils meeting the first criterion should be selected for lineup membership based upon more subjective ratings of similarity. While foils that are nearly identical to the target person should not be selected (thus avoiding what Luus and Wells, 1991, labelled the 'clone paradox'), those with 'sufficient similarity' should be sought, such that no combination of features would make the target stand out conspicuously from the other lineup members.

Current law-enforcement procedures

While a great deal of research has addressed the notion of proper construction techniques, no data are yet available on the actual procedures investigators use in daily practice. To gather some preliminary data on this issue, we surveyed investigators ($N = 27$) in two urban police departments and one county sheriff's department in the state of Florida who create lineups for the cases on which they work. Officers responded to questions regarding their procedures in constructing photo lineups, and their beliefs concerning an appropriate level of difficulty for the eyewitness.

When asked 'How many foils [in a six-person lineup] should there be that closely resemble the suspect, in terms of facial similarity, for it to be a reasonable measure of the witness's memory?' 67 per cent responded 'five' and 22 per cent 'four'. While Brigham *et al.* (1990) contended that a lineup should contain a minimum of three viable alternatives in a six-person lineup, Malpass (1981) argued for a more conservative criterion of five acceptable members. Most investigators seemed to agree with this more conservative approach.

When asked the source on which they usually relied to choose lineup foils, two-thirds responded 'the appearance of the suspect' and one-third used some combination of 'the suspect's appearance and the witness's description' (the latter procedure is consistent with what Koehnken *et al.*, 1996, suggest). None relied solely on the eyewitness's description, although some researchers (e.g. Luus and Wells, 1991) have argued that this would be the best procedure.

We also asked 'When constructing a lineup, how difficult do you attempt to make the identification of the suspect?' Almost all respondents (96 per cent) responded 4 or

5 on a 7-point scale where 7 represented 'very difficult'. Finally, in an attempt to gauge feelings about a reasonable test of eyewitness memory, we asked 'Hypothetically speaking, if 100 witnesses with a good to very good chance at viewing the suspect were presented a six-person lineup, *what number should fail to identify the suspect* for it to be considered a *reasonable assessment* of the witness's memory?' Responses ranged from 5 per cent to 80 per cent, with modal responses of 20 per cent and 10 per cent. Hence, these law officers appeared cognizant of the fact that, if a measure is to provide a reasonable assessment of memory, there will be some associated 'cost' (i.e. cases in which an eyewitness does not identify the perpetrator when he is in the lineup).

This preliminary assessment of construction practices used by investigators in the field suggests that they are somewhat sensitive to issues of lineup fairness. However, as with any survey assessment, there remains the possibility that demand characteristics may have influenced individual responses. Additionally, it is yet undetermined whether these issues are actually considered by investigators during everyday lineup construction. Further training for officers on the issues of lineup fairness, and departmental adoption of the procedures discussed in this section, would greatly improve both the precision and fairness with which suspects are identified by an eyewitness.

ASSESSING LINEUP FAIRNESS

Lineup size

The concept of lineup size is based on the premise that lineups should be large enough to ensure that the probability of a chance identification of an innocent suspect is low (Malpass and Devine, 1983). A lineup's size is dependent upon the number of 'acceptable' members it contains (i.e. members who are similar in general appearance to the suspect's description). To date, lineup fairness has most often been assessed on the basis of responses from 'mock witnesses' who have not observed the crime, but who attempt to identify the target (suspect) from a lineup based solely on a description of the criminal's general appearance. A maximally fair six-person lineup would be one in which mock witnesses selected the suspect and the five foils equally often. Assessments of lineup size, then, analyse how many lineup members were acceptable, in terms of being selected relatively often by the mock witnesses (Table 1). However, increasing the nominal size (number of members) of a lineup does not necessarily reduce the risk of a false identification of an innocent suspect, if additional lineup members are not similar enough to the suspect to represent plausible alternatives (Malpass, 1981). The relevance of lineup size for investigating eyewitness accuracy is reflected in the two assessment techniques below.

Effective size technique

The Effective Size (ES) technique is based on the supposition that lineup foils who are selected by mock witnesses at a level below that expected by chance are unacceptable for inclusion in the lineup (Malpass, 1981; Malpass and Devine, 1983). In order to obtain the ES of a lineup, the nominal size of the lineup is first adjusted to reflect the removal of any zero cells (members not chosen by any of the mock witnesses). The chance expectation is then adjusted to reflect the new resulting nominal size. The choice frequencies of those lineup members chosen by mock witnesses less often than

Table 1. Measures of lineup fairness

Measures of lineup size (i.e. number of 'good' lineup members)	Measures of lineup bias (i.e. distinctiveness of suspect)
<p>Effective Size (ES):</p> <ul style="list-style-type: none"> Assumes that foils selected by mock witnesses at below-chance levels are unacceptable 	<p>Proportions technique</p> <ul style="list-style-type: none"> Can use z for proportions to detect deviations from expected chance rate of selection Significance affected by # of mock witnesses Also insensitive to addition to 'bad' foils
<p>Number of Acceptable Lineup Members (ALM):</p> <ul style="list-style-type: none"> Those lineup members selected at 75% (or 50%, 25%, etc.) or more of chance expectation by mock witnesses are seen as acceptable 	<p>Functional Size (FS):</p> <ul style="list-style-type: none"> Parameter estimation procedure: $\frac{\text{\#mock witnesses}}{\text{\#mock witnesses choosing suspect}}$ <p>Suspect bias:</p> <ul style="list-style-type: none"> Similar to proportions technique, but uses ES instead of nominal size as a basis for calculation

expected by chance are subtracted from the adjusted chance expectation. The differences are then summed, and divided by the adjusted chance frequency. The resulting figure is subtracted from the lineup's nominal size. A lineup is then considered fair if the ES is equal to or greater than half the original nominal size of the lineup (Brigham *et al.*, 1990; Brigham and Pfeifer, 1994). Examples of Effective Size values of lineups used in criminal cases are presented later in Table 2.

Number of acceptable lineup members technique

A potential disadvantage of the ES technique is that it may not be easily understood by laypersons or judicial officials because of its numerical transformations and distance from the 'raw data'. Hence, Malpass and Devine (1983) suggested that the Acceptable Lineup Members (ALM) technique (also termed the Acceptable Foils Technique—Brigham and Pfeifer, 1994) better meets these criteria than does ES. The ALM estimate is derived by counting the number of lineup members that were selected by mock witnesses with a frequency that exceeds the chance expectation, or exceeds some percentage of it. The minimum percentage of the chance expectation that is considered acceptable may be based on value judgements or empirical standards. Malpass and Devine (1983, p. 93) utilized three different levels (50 per cent, 75 per cent, 90 per cent of chance expectation) in their analyses. Brigham *et al.* (1990) adopted a criterion of 75 per cent of chance expectation, in which case lineup members chosen from a 6-person lineup by at least 13 per cent of mock witnesses (0.75×0.17) would be acceptable.

Lineup bias

Even when a lineup has an adequate size, it still may be biased against the suspect in that the suspect or his photo is distinctive in comparison to the other lineup members. Lineup bias has been examined via several different estimates.

Proportions technique

Doob and Kirschenbaum (1973) originally described an assessment of lineup bias which compares the proportion of time the suspect is selected by mock witnesses with the proportion of choices expected by chance alone (0.17 in a six-person lineup). If the suspect were selected at a rate significantly greater than chance (z -test for proportions), the lineup would be seen as biased against the suspect. Wells *et al.* (1979) noted several limitations of this hypothesis-testing procedure. First, the likelihood of obtaining a significant difference is affected by the number of mock witnesses as well as by the magnitude of the difference between the proportions. Thus, the larger the sample of mock witnesses, the smaller the difference needed to achieve statistical significance.

Second, the hypothesis-testing procedure is not sensitive to the addition of irrelevant lineup foils. For example, if another four foils were added to a six-person lineup, the expected proportion of mock witnesses who pick the suspect out of a ten-person lineup would be 0.10 (1/10). However, if the additional four foils did not draw any mock-witness choices because they were not similar to the description of the suspect, then the obtained proportion would be compared with an expected proportion of 0.10, rather than with the original proportion of 0.17. This would increase the chances that a significant difference between proportions would occur and that the lineup would be labelled unfair.

Functional Size

As a result of these potential problems, Wells and colleagues (1979) advocated the use of a parameter estimation procedure that they labelled Functional Size (FS). FS is derived by dividing the total number of mock witnesses by the number who chose the suspect from the lineup. Both the FS estimate and the Proportions technique can be seen as indicators of lineup bias, rather than lineup size, because they do not take into account the distribution of foil misidentifications (Malpass, 1981).

Suspect Bias technique

In contrast, the Suspect Bias technique¹ takes into consideration the overall distribution of foil choices, but uses the calculated Effective Size of the lineup, rather than the lineup's nominal size, as the basis for calculating chance expectancies (Malpass and Devine, 1983). From this perspective, a photo lineup is seen as biased if the observed choice frequency for the suspect is significantly different (z -test for proportions) from the adjusted expected choice frequency (1/ES).

Sensitivity and discriminability of lineup fairness estimates

While the comprehensibility of a lineup fairness calculation technique is important, the *accuracy* of the technique in assessing lineup fairness is also a crucial factor. Brigham *et al.* (1990) asserted that fairness assessments should be evaluated in terms of both their sensitivity and discriminability. *Sensitivity* refers to an absolute standard of acceptability that is inherent in, or attributed to, a statistic. As such, the concept involves the cutoff point, in terms of the calculated size or degree of bias, at which a

¹Malpass and Devine (1983) called this the *defendant bias* technique, but we prefer *suspect bias* (cf. Brigham and Pfeifer, 1994) which more accurately reflects the fact that when a lineup identification is attempted, the suspect has usually not yet been indicted, and hence, is not yet a defendant.

given lineup should be classified as unfair. Wells *et al.* (1979, p. 289) proposed that researchers should not attempt to specify appropriate levels of sensitivity for fairness indices, arguing that these value-related distinctions should be left to the courts. However, other researchers (cf., Brigham *et al.*, 1990; Brigham and Pfeifer, 1994; Malpass and Devine, 1983) have suggested that appropriate levels of sensitivity can be estimated. In contrast, *discriminability* has been defined as the ability of a lineup fairness estimate to distinguish between lineups that are fair and those that are unfair, as established by some independent criterion.

Empirical assessment of lineup size estimates

In order to assess the sensitivity and discriminability of lineup size estimates, Brigham *et al.* (1990) created five 'selected' photo lineups (ostensibly fair lineups, as based on ratings of high foil similarity to the target and on police judgements of foil appropriateness) and five less-fair 'random' lineups (created by randomly matching with each target photo five foils that had been rated in the middle one-third of similarity ratings). Results indicated that the Effective Size (ES) technique was able to statistically discriminate between the selected (fair) lineups and the random (less fair) lineups, with the selected lineups showing higher ES values. The Acceptable Lineup Members (ALM) technique, however, did not significantly discriminate between the selected and random lineups. Based on these results, Brigham *et al.* suggested that the ES technique appeared to be the more useful size statistic.

In terms of sensitivity, Brigham *et al.* argued for a criterion that an acceptable ES (denoting a fair lineup) should be more than half of the nominal size (i.e. greater than 3.0 for a six-person lineup). Four of their five selected lineups and three of the five random lineups were classified as fair by this criterion. Malpass (1981), however, argued that a six-person lineup should have an ES of at least 5.0 in order to be considered fair. According to this more stringent criterion, none of the ten lineups would have been classified as fair according to the ES estimate. The application of these criteria to actual lineups will be discussed later.

Empirical assessment of lineup bias estimates

As with lineup size, Brigham *et al.* (1990) also investigated the appropriateness of assessment techniques for evaluating lineup bias, utilizing the experimental paradigm described previously. Both bias estimates, Proportions and Suspect Bias, were able to discriminate between the selected and random lineups. In terms of sensitivity, the Proportions statistic classified three of the five selected lineups as fair, while all five of the random lineups differed significantly from chance expectancy. Overall, the Suspect Bias was able to discriminate between the selected and random lineups. However, subsequent analyses (see Table 2) have indicated that the Suspect Bias measure does not have adequate sensitivity, an issue that will be addressed further in the next section.

Based on this pattern of results, Brigham *et al.* (1990) concluded that the Proportions technique appeared to be the most useful estimate of lineup bias, showing considerable discriminability and sensitivity. As discussed previously, however, with this method the statistical significance of a given difference between proportions is directly related to sample size (McNemar, 1969; Wells *et al.*, 1979). Brigham *et al.* (1990) used subsamples of 18 mock witnesses each in calculating proportions, finding it to be both practical and somewhat reliable. Additionally, a significant degree of bias

($p < 0.05$ by the z for proportions test) will be shown whenever the target person in a six-person lineup is selected by at least 34 per cent of the mock witnesses, thereby producing a Functional Size of 2.94 or less. In practical terms, this statistical criterion could be approximated by using a Functional Size of 3.00 or less ($N = 18$) as the criterion for meaningful bias in a six-person lineup.

Relationship of fairness estimates to evaluations made by law-enforcement personnel

The use of mock witnesses's responses as a source of lineup fairness assessments is often criticized by attorneys and law-enforcement personnel who fail to understand, or decline to accept, the rationale underlying the various indices. These observers are critical of a technique that utilizes respondents who have no expertise in lineups and are not evaluating the entire lineup's acceptability in any direct way. In an attempt to validate the use of these statistics, Brigham and Brandt (1992) constructed a sample of lineups that could be evaluated by three sets of subjects performing different tasks: college students acting as mock witnesses, and college students and law officers providing personal evaluations of the lineups.

A set of 23 photo lineups were created and categorized into one of three groups: fair (eight lineups), moderately fair (eight lineups), and least fair (seven lineups). The law officers and students first subjectively assessed the global fairness of each lineup and then evaluated the acceptability of each of the five foils on a 6-point scale. Finally, a question designed to elicit 'estimated proportions' scores was asked. Four general types of lineup fairness indices were derived from the three samples: global fairness, lineup bias (Proportions) and two estimates of lineup size (ES and ALM).

Across the 23 lineups, law-enforcement evaluators were significantly more willing to use the lineups than were the college-student evaluators. With respect to the most central question, college-student mock-witness responses *were* predictive ($r(21) = 0.42$) of how law officers subjectively evaluated the same lineups based on the overall fairness estimate (a composite of the three mock witness indices). Two of the mock witness estimates, Proportions and ES, were consistently related to fairness indices derived from the other two samples. However, the other mock witness size estimate, ALM, was not consistently related to the evaluators' responses. This is consistent with results of Brigham *et al.* (1990) discussed earlier.

The overall pattern of results indicated that estimates of lineup fairness derived when college students play the role of mock witnesses are related to direct evaluations of lineup fairness made by people (law officers or college students) who evaluate entire lineups in a more straightforward manner. If one assumes that, on average, law officers' direct evaluations of lineup fairness are somewhat valid but subject to various biasing influences, then the moderate degree of relationship found by Brigham and Brandt (1992) seems appropriate. That is, we would predict only a moderate relationship if the samples are responding in terms of the same concept but are affected by different types of external biasing factors.

THE LINEUP EXPERT'S DILEMMA: ESTIMATING FAIRNESS IN REAL CASES

The validity and utility of lineup fairness estimates is more than an issue of purely academic interest. For over two decades, some researchers have been utilizing these

estimates to assess the fairness of lineups used in actual criminal cases, and some have tried to present their findings to judges and jurors via expert testimony. Below we will discuss several aspects of this contentious issue. We begin by analysing the applicability of these estimates to a set of 18 real cases, after which we present estimates of the frequency with which such analyses have been admitted in court in the past, and discuss the likely admissibility of such analyses in the future.

Estimating fairness of real lineups: 18 cases

How well do lineup fairness statistics apply to lineups used in actual criminal cases? This is a difficult issue to assess, but we will present some data that are directly relevant, based on the analyses of photo lineups (or, in two cases, photos of a live lineup) that were used in 18 criminal felony prosecutions in the 1980s and 1990s (Table 2). These were cases in which a defence attorney contacted one of us (JCB) because the attorney felt that the identification procedures in the case were unsatisfactory. In 16 of the 18 cases there was at least one eyewitness identification of the perpetrator, totalling 23 eyewitness descriptions that could be used for match-to-description analyses. (One of those analyses involved two differing descriptions given by the same eyewitness at different times.) In three cases two different lineups were used by the police, so a total of 26 lineup fairness analyses, using the match-to-description paradigm, could be conducted across the 18 cases to assess lineup bias (Proportions, FS, and Suspect Bias) and lineup size (ES and ALM).

In addition, lineup fairness was also assessed for all 20 lineups using a no-description procedure in which mock witnesses were simply asked which person or photo looked most distinctive or most different from the others. The employment of 'blind' witnesses provided us with an indication of whether a witness, in the absence of a any memory for the suspect, would respond to the distinctiveness of the individual (feature-based) or of the photo (lighting, odd markings, pose, etc.).

It should be stressed that these lineups are not presented as a representative sample of lineups used in criminal cases. Presumably, they are located towards the unfair end of the continuum, since the attorneys felt it desirable to seek out a researcher to assess the lineup's fairness. That said, what do the lineup fairness indices tell us about these lineups? One relevant question is whether the estimates show an appropriate degree of sensitivity. If they are indiscriminant, such that all lineups are classified as fair or as unfair, then the statistics would not be useful in the criminal justice setting because they would not distinguish between more-fair and less-fair lineups. Recall that Brigham *et al.* (1990) argued that a lineup fairness estimate denotes a fair lineup when the value of FS or ES is greater than half of the lineup's nominal size (greater than 3.0 for a six-person lineup), or when the ALM is at least half of the lineup's nominal size. Statistical significance of the *z*-test for proportions can be used as the criterion for significant bias according to the Suspect Bias index.

Under these criteria, the match-to-description analyses found that almost 60 per cent were classified as biased according to FS, but only 15 per cent were classified as biased according to the Suspect Bias statistic. Looking at size, 50–60 per cent of the lineups were classified as unfair due to insufficient size according to the ES and ALM indices. The no-description procedure yielded a somewhat different pattern. Here the lineup bias estimates yielded a designation of unfair in only 5 per cent of the lineups according to FS, and 35 per cent of the lineups according to the Suspect Bias

Table 2. Assessing the fairness of lineups used in actual cases

Case #	Type of case	Assessment task	Nominal size	# of mock witnesses	Estimates of bias			Estimates of size		Overall fairness index
					Proportions technique	Functional Size	Suspect Bias	# of ALM	Effective Size	
1	Murder	EW 1	6 LIVE/C	18	0.50**	2.00	ns	3	2.89	1
		EW 2		15	0.33	3.00	ns	4	3.33	2
		EW 3		16	0.50**	2.00	ns	3	3.50	2
		ND		29	0.38**	2.64	ns	4	3.86	2
2	Murder	EW	6 B/W	66	0.26	3.94	ns	4	3.87	4
		ND		28	0.08	14.00	ns	1[#]	2.07	3
3	Murder	EW 1	6 B/W	16	0.66***	1.52	bias*	2	2.75	1
		EW 2		17	0.44**	2.27	ns	3	3.71	2
		ND		64	0.30**	3.37	ns	3	4.03	4
4	Murder	CD	6 B/W	58	0.66***	1.53	bias*	2	2.69	1
		ND		89	0.37***	2.70	ns	2	3.62	2
5	Bank robbery	EW	7 COLOUR	43	0.53***	1.72	ns	2	1.81	1
		ND		28	0.43***	2.33	ns	2	4.36	2
6	Bank robbery	ND	7 B/W	46	0.17	5.75	ns	4	3.52	4
7	Armed robbery	EW	6 B/W	31	0.23	4.43	ns	2	2.97	3
		ND		34	0.15	6.80	ns	3	2.91	3
8	Armed robbery	CD	6 B/W	22	0.41**	2.44	ns	2	2.36	1
		ND		34	0.03	34.00	ns	2[#]	2.74	3
9	Armed robbery	CD	6 COLOUR	47	0.23	4.27	ns	4	4.66	4
		CD	4 LIVE/C	46	0.50**	2.00	ns	2	2.61	2
		ND		33	0.15	6.60	ns	3 [#]	2.58	4
10	Rape	ND	8 B/W	88	0.53***	1.89	bias***	3	3.76	1
11	Sexual assault	EW	6 COLOUR	12	0.75***	1.33	ns	2	1.75	1
		ND		30	0.41***	2.45	ns	2	3.04	2
		EW	6 LIVE/C	12	0.83***	1.20	ns	2	1.33	1
		ND		30	0.33**	3.00	ns	4	3.44	2

12	Sexual battery	EW	8 B/W	56	0.68***	1.47	bias***	2	3.75	1
		ND		53	0.04	26.50	ns	3[#]	4.59	4
13	Sexual battery	EW	6 COLOUR	37	0.73***	1.37	bias***	2	2.35	1
		ND		34	0.29	3.40	ns	3	4.06	4
14	Sexual battery	EW 1a†	6 COLOUR	49	0.18	5.44	ns	4	4.35	4
		EW 1b†		28	0.21	4.67	ns	3	2.61	3
		ND		28	0.82**	1.22	ns	2	1.54	1
15	Burglary and assault	EW	6 B/W	25	0.88***	1.14	ns	1	1.37	1
		ND		34	0.26	3.78	ns	4	4.06	4
		EW	6 B/W	34	0.35**	2.83	ns	4	3.53	2
		ND		33	0.06	16.50	ns	4[#]	4.45	4
16	Burglary and assault	EW	6 B/W	22	0.32	3.14	ns	2	3.59	4
		ND		33	0.09	11.00	ns	1[#]	2.82	3
17	Burglary	EW 1	8 B/W	26	0.19	5.20	ns	3	3.69	3
		EW 2		27	0.37	2.70	ns	4	4.85	2
		EW 3		22	0.18	5.50	ns	4	5.87	4
		ND		34	0.03	34.00	ns	2[#]	3.23	3
18	Lewd and lascivious behaviour	EW 1	8 B/W	15	0.07	15.00	ns	5[#]	6.14	4
		EW 2		19	0.16	6.33	ns	4	5.21	4
		ND		33	0.24	4.13	ns	2	3.09	3

Notes: Measures in **bold** denote an unfair lineup according to the criteria suggested by Brigham *et al.* (1990).

EW: eyewitness description; CD: composite description; ND: no description.

LIVE/C: colour photograph of live lineup; COLOUR: colour photo array; B/W: black and white photo array.

ns: non-significant degree of bias against suspect; ALM: Acceptable Lineup Members.

*** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$; #target was not an ALM; †two descriptions given by the same eyewitness.

Overall fairness index values (1 = most unfair; 2 = unfair; 3 = somewhat unfair; 4 = fair).

Table 3. Percentage of lineups used in actual cases classified as 'unfair'

	Match-to-description paradigm ($N = 26$)	No-description paradigm ($N = 20$)
<i>Lineup bias</i>		
Functional Size ^a	58.8%	35.0%
Suspect Bias ^b	15.4%	5.0%
<i>Lineup size</i>		
Effective Size ^c	50.0%	40.0%
# of ALM ^{d1}	60.0%	55.0%

^aCriterion = FS is half or less of the lineup's nominal size

^bCriterion = Statistical significance of z for proportions test

^cCriterion = ES is half or less of the lineup's nominal size

^dCriterion = # if ALM is less than half of the lineup's nominal size

¹ALM: Acceptable Lineup Members

technique. The no-description analyses found lineups to be of insufficient size 40 per cent of the time according to the ES estimate, and 55 per cent of the time according to the ALM estimate (Table 3).

If we look at the degree of agreement for the fair/unfair dichotomous classifications, a somewhat different picture emerges. The two bias estimates yielded different categorizations almost half the time for the match-to-description analyses and about one-third of the time for the no-description analyses. On the other hand, the degree of agreement in fairness categorizations between the two size estimates (ALM and ES) was considerably greater, with 80–90 per cent agreement across the cases (Table 4). The lack of agreement between the two estimates of bias was mostly a reflection of the fact that the Suspect Bias statistic was much less sensitive to lineup unfairness than was Functional Size, as illustrated in Table 3. It appears that when a given lineup's size is relatively low (i.e. less than half its nominal size), chance frequency increases such that the proportion of mock witnesses selecting the suspect must be extremely high (i.e. 50–60 per cent) to yield a significant z for proportions difference.

What about agreement across the fairness estimates: did two statistics relevant to lineup size (ES and ALM) and an estimate of lineup bias (Proportions, on which FS is based) yield roughly equivalent scores across the set of lineups? (The Suspect Bias

Table 4. Within-lineup agreement for estimates of lineup size and lineup bias

	Lineup bias statistics (FS and SB)			Lineup size statistics (ES and ALM)		
	Lineup fair (both)	Mixed	Lineup unfair (both)	Lineup fair (both)	Mixed	Lineup unfair (both)
Match-to-description paradigm ^a	38.5%	46.2%	15.4%	46.2%	11.5%	42.3%
No-description paradigm ^b	65.0%	30.0%	5.0%	40.0%	20.0%	40.0%

Notes: ^a26 analyses; ^b20 analyses.

FS: Functional Size; SB: Suspect Bias.

ES: Effective Size; ALM: Number of Acceptable Lineup Members.

Table 5. Relationship (correlations) between measures of lineup size and bias

<i>Eyewitness's description used</i>		
Actual cases: 26 descriptions		
	<i>Effective Size</i>	<i># of ALM</i>
Proportions	-0.77***	-0.76***
Effective Size	-	0.80***
Research-created fair and less-fair lineups: 23 lineups and descriptions (Brigham and Brandt, 1992)		
	<i>Effective Size</i>	<i># of ALM</i>
Proportions	-0.83***	-0.80***
Effective Size	-	0.85***
<i>No description used</i>		
Actual cases: 20 lineups		
	<i>Effective Size</i>	<i># of ALM</i>
Proportions	-0.07	-0.04
Effective Size	-	0.59***

Note: *** $p < 0.001$; # of ALM: number of Acceptable Lineup Members.

technique does not yield an equivalent score that can be correlated with the others.) As illustrated in Table 5, correlational analyses indicated that the answer is 'yes' when the match-to-description procedure was used, but 'no' when the no-description format was utilized. The three estimates correlated strongly with each other (r 's above 0.75) across the actual cases and also across the 23 made-for-research lineups used by Brigham and Brandt (1992), when descriptions were used. Under the no-description procedure, however, the bias estimate (Proportions scores) was not significantly related to the two size estimates, ES, $r(20) = -0.07$, and the ALM score, $r(20) = 0.04$. Because the Proportions score focuses solely on mock witness responses to the target person (bias), while the other two indices evaluate responses across all the lineup members (size), it is entirely possible for a lineup to appear fair in terms of Proportions because few mock witnesses chose the target person, but unfair in terms of ES and ALM because responses focused mostly on one or two *other* lineup members.

Given that size and bias statistics assess different features of a lineup, how should these estimates be used together in coming to an overall forensic evaluation of a lineup's worth? Ideally, a lineup should be both of sufficient size and without bias. But, from the perspective of the accused, positive bias is the most critical issue. If the lineup is positively biased (i.e. the suspect is distinctive in some way) then it will probably be deficient in size as well – this was the case for 52 per cent of the lineups that were assessed as positively biased in Table 2. For the remaining 48 per cent of the biased lineups, one would still argue that the lineup is fundamentally unfair to the suspect, even though insufficient size itself was not a problem. But what about lineups that have inadequate size but are not positively biased, because mock witness choices were focused not on the suspect but on one or two lineup foils (e.g. Case 18 based on the eyewitness's description, and Cases 2, 8, 16, and 17 based on no-description in

Table 2)? It is less clear what one might argue in this situation. The lineup is not ideal in that there are less than three viable choice options, but the suspect is *not* one of the viable choice options, according to the pattern of mock witness responses.

It seems to us that overall lineup fairness may best be estimated by combining size and bias estimates. Given that bias appears to be the most important factor, the order from most unfair type of lineup to fair would be: *most unfair* – both positive bias and inadequate size; *unfair* – positive bias only; *somewhat unfair* – inadequate size only; and *fair* – no bias and adequate size. In order to utilize such a scale, one would need to specify a single bias statistic and a single size estimate. We suggest the use of Functional Size as the estimator of bias; it does not rely on sample size as the Proportions statistic does, and it has greater sensitivity than the Suspect Bias statistic. As a size estimator, we suggest the use of Effective Size, which previous research (Bringham and Brandt, 1992; Bringham *et al.*, 1990) found to be a more discriminant tool for detecting unfairness due to size than was the ALM. A potential disadvantage of ES is that it is not easily understandable to laypersons and could prove confusing if an expert attempted to explain it in detail to a judge or jury (cf. Malpass and Devine, 1983). Nevertheless, Effective Size appears to be the most appropriate estimator of lineup size based upon its statistical properties.² Utilizing these two estimates, FS and ES, produces the overall fairness categorizations in the far-right column of Table 2.

In multiple-eyewitness cases, where there is more than one description on which mock-witness judgements can be based, there is the possibility that overall fairness evaluations may differ depending on which descriptions the mock witnesses are given. Four of the cases (1, 3, 17, and 18) in Table 2 involved more than one eyewitness's description. In two of these cases the overall fairness estimate was the same regardless of which description was analysed, while Case 3 yielded somewhat different estimates in the degree of unfairness (very unfair; unfair). The only major between-witness difference involved Case 17, which yielded estimates of unfair, somewhat unfair, and fair depending upon which of the three descriptions was used.

Comparing the match-to-description estimates with estimates from the no-description format, it is evident that a lineup may be judged unfair by both sets of analyses (e.g. Cases 1, 4, 5, and 11), or fair by one set of analyses but not the other. Interestingly, in no cases was the lineup categorized as fair by all the analyses. Focusing on a single estimate of bias (FS) and a single estimate of lineup size (ES) allows one to derive a meaningful overall estimate of lineup fairness. In contrast, if one attempts to assess *all five* proposed estimates of lineup fairness that are listed in Table 2, the situation becomes much more murky. For example, for Case 1 which involved three eyewitnesses who gave differing descriptions, all four analyses indicated significant unfairness (bias) in terms of FS, but none in terms of Suspect Bias. Turning to lineup size, only one of the four analyses indicated unfairness in terms of ES, and none of the four analyses indicated unfairness insofar as the ALM was concerned. What could a scientist or expert witness legitimately conclude about the fairness of such a lineup? Or consider Case 17: the lineup is classed as fair by all

²It is not the duty of the scientific community to choose statistical measures that are easiest for laypersons to comprehend. Rather, they should be chosen on a basis of fundamental reliability and validity. If ease of comprehension were the standard, DNA evidence, which has been considered quite confusing to some jurors, might not be admitted into evidence. We contend that as long as the ultimate goal of the statistic is made clear to the jurors, that of determining fairness due to size, the method of calculation is less relevant. When ES is used as a step towards obtaining a single, global fairness estimate, the details of its calculation may not seem as great an issue.

four statistics when the third eyewitness's description is used, fair in three quarters of the analyses using eyewitness #2's description, but fair in only two of the four analyses when eyewitness #1's description was used. Case 14 provides an instance in which using two descriptions given by the same eyewitness at different times produced somewhat inconsistent outcomes (see also Corey *et al.*, this issue).

Lineup fairness analyses of the 18 actual cases listed in Table 2 produced consistent outcomes for most of the cases. Based upon this non-random sample of 18 cases, it appears that in some cases, lineups are so unquestionably fair or unfair that they will be classified clearly, regardless of which of the fairness indices are used. If the researcher wishes to present these analyses to the court via expert testimony or affidavit, the expert's task seems relatively straightforward. But vexing issues remain when analyses of a case yields mixed results, as in Case 17. Should the expert report the inconsistent data and let the attorney, judge, or jury decide how to interpret it? Or should the expert decline to testify on the grounds that the data are too inconsistent from a scientific perspective, or because the resulting testimony would be confusing and unhelpful to jurors (thereby violating the 'helpfulness' prong of the Federal Rules of Evidence and many state evidence standards)? It seems likely that if such expert testimony is proffered before a judge, the judge will be less likely to admit it at trial if the results are inconsistent.

Current practice in applying lineup fairness indices

It has been widely documented that courts in most jurisdictions in the USA are usually unwilling to admit expert testimony from researchers concerning the accuracy of eyewitness evidence (e.g. Brigham, 1989; Cutler and Penrod, 1995; Lipton, 1996). While there is documentation that such expert testimony has been admitted in hundreds of cases over the past 25 years (Fulero, 1993, unpublished manuscript; Kassir *et al.*, 1989), the general consensus is that such admission is unlikely, except in a few jurisdictions (e.g. the states of Washington, Alaska, Arizona, California, and Ohio) in which appellate courts have ruled that exclusion of such testimony is an abuse of judicial discretion under some circumstances.

In order to get a crude estimate of the frequency of acceptance of such expert testimony, we conducted an informal survey of 11 of the leading eyewitness memory researchers in the USA and Canada, asking them to estimate the number of cases in which an attorney had attempted to have their testimony admitted before a jury at trial, and the number of times it had been admitted. The pattern of responses was bimodal, seeming to depend largely on the state or judicial district in which the expert worked. Three experts said that they had testified in one hundred or more cases, and estimated that their testimony had been accepted 75 per cent or more of the time. In contrast, seven other experts, also leading researchers, reported that their proposed expert testimony had been permitted at trial only occasionally – individual estimates ranged from 1 per cent to 35 per cent of the time.

Turning specifically to lineup fairness estimates, we also asked the experts to consider the cases in which they had consulted with an attorney about disputed eyewitness evidence, and to estimate the percentage of these cases in which they gathered mock witness data to assess the fairness of the lineup(s) used. The estimates ranged from 2 per cent to 60 per cent of the cases, with a median of 15 per cent across experts. We also asked them to consider all the cases in which they had attempted to

deliver expert testimony about eyewitness memory at a pretrial evidentiary hearing or at trial in front of a jury, and to estimate the proportion of times that this testimony included, or would have included, data on mock witness-based lineup fairness estimates. Responses varied widely here as well. Most experts reported that this data was almost never a part of their expert testimony. The median percentage of cases in which experts testified about eyewitness memory and also presented mock-witness-based lineup fairness data was 8.5 per cent for pretrial hearings, and 3 per cent at trial. Hence, although many psychologists have spent a great deal of time and effort in developing these statistics and speculating about their relevance and utility for actual cases, it appears that they are almost never presented in the courtroom.

CONCLUSION

In terms of lineup construction procedures, some of the investigators we surveyed reported behaviours consistent with Koehnken and colleagues' (1996) suggestions in which some combination of a match-to-description and match-to-appearance strategy was used. They also appeared somewhat cognizant of biasing factors in the construction of photospreads. However, the majority of construction methods utilized differed from the manner in which researchers later attempt to assess the validity of the lineup. This may account for some of the differences across lineup fairness estimates demonstrated both in the current data and in previous studies. For example, we have seen that different eyewitness's descriptions sometimes produced dissimilar patterns of mock witness responses (and subsequent fairness categorizations) across the same lineup. Were a match-to-description strategy used when constructing the lineup, as suggested by Luus and Wells (1991), such difficulties should be eliminated when later assessing the fairness of the lineup with the match-to-description technique. On the other hand, if one were to employ a match-to-appearance strategy (as do the majority of those investigators we surveyed), then alternative procedures for assessing the fairness of the lineup may need to be explored, so as to avoid inconsistent findings for a single lineup.

Jurors and judges are faced with a difficult task when they are asked to assess the fairness of a lineup at trial. In the vast majority of occasions, this judgement is made on a purely subjective basis. Most often, the triers of fact may listen to a defence attorney's arguments regarding the lineup's purported unfairness (in the attorney's opening or closing statements), and may be given a brief look at the set of photos used by law-enforcement officials. However, they are given no further guidance in estimating the fairness of the lineup. As we have noted, expert testimony that might provide the court with a scientifically based frame of reference for making this difficult judgement is seldom admitted by trial judges in most jurisdictions. Further, even in cases where such expert testimony is admitted, experts seldom present the results of an empirical analysis of the lineup's fairness. Does this puzzling state of events mean that mock witness-based lineup assessments are seen (by researchers or by judges) as irrelevant to actual criminal cases? Are the defence attorney's protestations and a quick 'eyeballing' of the photospread the only information jurors need to accurately assess the fairness of a lineup? We think not.

For example, many of the cases from the current data set (e.g. Cases 1, 3, 4, 5, 8, 10, 11, and 13 in Table 2) provide clear empirical evidence of overall unfairness: positive

bias and inadequate size. Such information could be presented to jurors in a way that gives them additional (helpful) information with which to evaluate that lineup. We suggest that an overall fairness evaluation, based on the concepts of Functional Size and Effective Size (but not Suspect Bias or number of Acceptable Lineup Members), can be presented meaningfully to judges and jurors, provided the expert is given the latitude to explain the fairness estimates carefully.

In contrast, some cases (e.g. Cases 2, 6, 16, and 18 in Table 2) yielded data that the lineup was basically structurally fair. This information would be valuable to jurors as well, although it is unlikely to reach them since the research was probably commissioned by the defence, and in such a case, the defence attorney is unlikely to call the researcher as an expert witness. If the researcher is called, the defence attorney is unlikely to ask the expert about lineup fairness analyses.

Most courts have taken a negative stance towards admitting expert testimony about eyewitness memory in general. Several experts told us that they seldom or never tried to present empirical evidence about lineup fairness because they believed that judges would be even more sceptical towards this type of testimony than towards more general testimony. However, it is not clear that this would necessarily be the case, particularly if an understandable empirically based and conceptually justified overall estimate of fairness is employed. Courts have often rejected general expert testimony on the basis that it would cover material already thought to be within the 'common knowledge' of jurors. However, it may be that empirically based lineup fairness information would be less susceptible to this criticism if it seeks to provide assistance to the court for those cases in which fairness analyses had shown the lineup to be unambiguously unfair or fair.

In any event, it seems to us that it remains an important quest to extend and clarify the conceptualizations of lineup fairness, to further develop and refine methods of lineup construction, and to seek assessment techniques that provide conceptually justified and empirically sound estimates of lineup fairness. If assessment techniques can be developed that provide consistent estimates of fairness, coupled with appropriate degrees of sensitivity, discriminability, and ease of understanding to laypersons, psychology may be able to furnish the courts with worthwhile testimony that would ensure a greater measure of protection against unfair practices in the use of eyewitness evidence.

ACKNOWLEDGEMENTS

We would like to thank Rebecca Williams for her assistance on this project. Portions of these data were presented at the Biennial Meeting of the American Psychology-Law Society in Redondo Beach, CA, in March, 1998.

REFERENCES

- Bothwell, R. K., Brigham, J. C. and Malpass, R. S. (1989). Cross-racial identifications. *Personality and Social Psychology Bulletin*, **15**, 19–25.
- Brigham, J. C. (1989). Disputed eyewitness identifications: Can experts help? *The Champion*, **8**, 691–695.

- Brigham, J. C. and Brandt, C. C. (1992). Measuring lineup fairness: Mock witness responses vs. direct evaluations of lineups. *Law and Human Behavior*, **16**, 475–489.
- Brigham, J. C. and Ready, D. J. (1985). Own-race bias in lineup construction. *Law and Human Behavior*, **9**, 415–424.
- Brigham, J. C., Ready, D. J. and Spier, S. A. (1990). Standards for evaluating the fairness of photograph lineups. *Basic and Applied Psychology*, **11**, 149–163.
- Brigham, J. C. and Pfeifer, J. E. (1994). Evaluating the fairness of lineups. In D. F. Ross, J. D. Read and M. P. Toglia (Eds.), *Adult eyewitness testimony: Current trends and developments* (pp. 201–222). Cambridge: Cambridge University Press.
- Cutler, B. L. and Penrod, S. D. (1995). *Mistaken identification: The eyewitness, psychology, and the law*. Cambridge: Cambridge University Press.
- Doob, A. N. and Kirshenbaum, H. M. (1973). Bias in police lineups—partial remembering. *Journal of Police Science and Administration*, **1**, 518–544.
- Kassin, S. M., Ellsworth, P. C. and Smith, V. L. (1989). The ‘general acceptance’ of psychological research on eyewitness testimony: A survey of the experts. *American Psychologist*, **44**, 1089–1098.
- Koehnken, G., Malpass, R. S. and Wogalter, M. S. (1996). Forensic applications of line-up research. In S. L. Sporer, R. S. Malpass and G. Koehnken (Eds.), *Psychological issues in eyewitness identification* (pp. 205–231). Mahwah, NJ: Erlbaum.
- Laughery, K. R., Jensen, D. G. and Wogalter, M. S. (1988). Response bias with prototypic faces. In M. M. Gruneberg, R. Sykes and P. Morris (Eds.), *Practical aspects of memory: Current research and issues* (pp. 157–162). Chichester: Wiley.
- Lipton, J. P. (1996). Legal aspects of eyewitness testimony. In S. L. Sporer, R. S. Malpass and G. Koehnken (Eds.), *Psychological issues in eyewitness identification* (pp. 7–22). Mahwah, NJ: Erlbaum.
- Luus, C. A. E. and Wells, G. L. (1991). Eyewitness identification and the selection of distracters for lineups. *Law and Human Behavior*, **14**, 43–57.
- Malpass, R. S. (1981). Effective size and defendant bias in eyewitness identification lineups. *Law and Human Behavior*, **5**, 299–309.
- Malpass, R. S. and Devine, P. G. (1983). Measuring the fairness of eyewitness identification lineups. In S. M. A. Lloyd-Bostock and B. R. Clifford (Eds.), *Evaluating witness evidence* (pp. 81–102). Chichester: Wiley.
- McNemar, Q. (1969). *Psychological statistics* (4th edn). New York: Wiley.
- Marwitz, D. B. and Wogalter, M. S. (1988). Bias in photo-spreads of faces: A comparison of two lineup construction methods. In *Proceeding of the Human Factors Society 32nd Annual Meeting* (pp. 541–543). Santa Monica, CA: Human Factors Society.
- Wells, G. L. and Bradfield, A. L. (1998). ‘Good, you identified the suspect’: Feedback to eyewitnesses distorts their reports of the witnessing experience. *Journal of Applied Psychology*, **83**, 360–376.
- Wells, G. L., Leppe, M. R. and Ostrom, T. M. (1979). Guidelines for empirically assessing the fairness of a lineup. *Law and Human Behavior*, **3**, 285–293.
- Wells, G. L., Rydell, S. M. and Seelau, E. P. (1993). On the selection of distractors for eyewitness lineups. *Journal of Applied Psychology*, **78**, 835–844.
- Wells, G. L., Seelau, E. P., Rydell, S. M. and Luus, C. A. E. (1994). Recommendations for properly conducted lineup identification tasks. In D. F. Ross, J. D. Read and M. P. Toglia (Eds.), *Adult eyewitness testimony: Current trends and developments* (pp. 223–244). Cambridge: Cambridge University Press.
- Wells, G. L., Small, M. A., Penrod, S. D., Malpass, R. S., Fulero, S. M. and Brimacombe, C. A. E. (1998). Eyewitness identification procedures: Recommendations for lineups and photospreads. *Law and Human Behavior*, **23**, 603–648.
- Wogalter, M. S., Marwitz, D. B. and Leonard, D. C. (1992). Suggestiveness in photospread lineups: Similarity induces distinctiveness. *Applied Cognitive Psychology*, **6**, 443–453.
- Wogalter, M. S., Van’t Slot, L. J. and Kalsher, M. J. (1991). Bias in police lineups and its reduction an alternative construction procedure. In *Proceedings of the Human Factors Society 35th Annual Meeting* (pp. 1561–1565). Santa Monica, CA: Human Factors Society.