APPLIED COGNITIVE PSYCHOLOGY
Appl. Cognit. Psychol. **13**: S1–S7 (1999)

Measuring Line-up Fairness

ROY S. MALPASS[1]* and R. C. L. LINDSAY[2]
[1]*University of Texas, El Paso, USA*
[2]*Queen's University, Canada*

SUMMARY

The fairness of line-ups and photospreads is a traditional concern of research and policy development in the area of eyewitness identification. Quantification of fairness, the construction of fairness indices, and the development of evaluation procedures started in the 1970s and continues to this day. This paper reviews the historical development of the field as an introduction to the articles that follow. The entire set of articles addresses current questions and raises new issues of measuring the fairness of identification procedures. Copyright © 1999 John Wiley & Sons, Ltd.

The literature on the measurement of line-up fairness began with the report by Doob and Kirschenbaum (1973) of their attempt to measure the bias contained in an identification process brought to their attention by a defence attorney. The verbal description of the offender was extremely vague. She stated only that the criminal was 'rather good looking'. The possibility existed that the witness had simply chosen the best-looking person from the line-up, whether he was the offender or not. To test this conjecture, Doob and Kirschenbaum gave the description of the criminal and a photograph of the line-up to a set of persons who had never seen the criminal. They were asked to choose the suspect from the line-up. If these 'mock witnesses' (Wells *et al.*, 1979) were able to identify the suspect, never having seen him but solely on the basis of the verbal description, the line-up could be considered biased in the sense that the alternative line-up members were not full (effective, functional) alternatives to the suspect. The underlying assumption of Doob and Kirschenbaum's method was that the alternatives (distractors, fillers, or foils) in a line-up ought to protect innocent suspects from identifications based solely on their match to the witness's description of the criminal (recall). Instead, the witness should have to rely on the presumably more informative data provided by recognition of the face as a match of the suspect's appearance to the witness' memory image derived from having witnessed the original event (Wells, 1993). Thus line-up foils ought to be plausible choice alternatives: the suspect should not stand out in comparison to them. The *reductio ad absurdum* would be to place a suspect described as a European–American in a line-up with an African–American, an Asian–American, an elephant, and a refrigerator.

Doob and Kirschenbaum's (1973) procedure offered an attractive paradigm for *post hoc* evaluation of fairness in line-ups and photospreads, through the use of the

*Correspondence to Roy Malpass, Department of Psychology, University of Texas, El Paso, TX, USA, E-mail: rmalpass@utep.edu

Contract grant sponsor: Social Sciences and Humanities Research Council of Canada.

CCC 0888–4080/99/SI00S1–07 $17.50

choices made by mock witnesses. From a measurement point of view, the index of fairness used was the straightforward application of deviation from chance expectation, expressed as a proportion of mock witness choices of the suspect. Simple statistical tests on proportions were the medium of analysis. The identification procedure was deemed unfair if the suspect was selected by significantly more mock witnesses than would have been expected by chance, where chance expectation was $1/N$ and N was the number of people or photos presented to the witness.

FUNCTIONAL SIZE

Wells *et al.* (1979) introduced a quantitative index of fairness: Functional Size. Functional Size is defined as the total number of mock witnesses divided by the number of mock witnesses who chose the suspect (or 1 over the proportion of mock witnesses choosing the suspect). The intention was that this index represent the number of foils functionally present in the line-up. Functional Size provides some advantages in comparison with the significance testing approach. As sample size (of mock witnesses) increases, the likelihood of rejecting a line-up increases using significance tests even when the pattern of mock witness selections is constant. This potentially creates a situation in which the judged fairness of a line-up could be manipulated by employing too few mock witnesses to detect meaningful differences or so many that even trivial departures from randomness would be statistically significant. However, as sample size increases, the estimate of Functional Size simply becomes more reliable, not more likely to indicate bias.

A second advantage is that the Functional Size measure is appropriately insensitive to the presence of poor foils in a line-up. For example, consider 100 mock witnesses shown a line-up from which 20 selected the suspect and 20 selected each of four foils. Using the significance testing approach, this line-up would be considered fair if it had contained 5 members but unfair if it had contained 10 members with 5 foils receiving no mock witness selections. The Functional Size of the line-up would be 5.00 whether the line-up had included the 5 useless foils or not.

On the other hand, Functional Size is unbounded at the upper end. For example, if a suspect is chosen less frequently than expected by chance, the Functional Size will be greater than the number of people actually in the line-up; i.e. if the suspect is chosen at half of chance expectation, the Functional Size will be twice the number of people (suspect + foils) actually present. Even worse, if no mock witness selects the suspect, the Functional Size of the line-up is undefined ($N/0$). Also, the term Functional Size is a misnomer. The index actually takes no account at all of the individual foils and does not assess their individual adequacy as choice alternatives to the suspect. Hence, Functional Size does not address the size (number of adequate members) of a line-up. Because it focuses exclusively on choices of the suspect, the Functional Size index is a measure of bias towards or away from the suspect.

EFFECTIVE SIZE

Malpass (1981) distinguished between two related aspects of a line-up: its size as an aggregate (the number of effective members), and the degree of bias toward or away from the suspect. The Effective Size measure was offered as an index of the aggregate number of members present in the line-up on the basis of the distribution of mock witness choices. It is calculated in the following way:

(1) Find the adjusted nominal size (N_a) by subtracting from nominal size (N) the number of foils receiving no identifications (null foils). Use N_a to find the adjusted nominal chance expectation $(E)_a$. E_a equals $1/N_a(n)$, where n is the number of mock witnesses.
(2) For each non-null foil whose observed frequency of identification is less than E_a
  (a) Subtract its observed frequency (O) from E_a
  (b) Sum these differences, and
  (c) Divide this sum by E_a.
  This estimates the degree to which the line-up foils, taken together, fall short of fulfilling their adjusted nominal chance expectation. The resulting number is the number of persons effectively absent from the line-up.
(3) Subtract the results of (2c) above from the adjusted nominal size of the line-up. This result is the line-up's Effective Size (*ES*). If this result is expressed as a percentage of nominal size (*ES*/nominal size) × 100, the percentage reduction of Effective Size from nominal size is apparent.

Effective Size takes account of information from each of the members of the line-up. It produces an index that is bounded at the lower end by zero and at the upper end by the nominal size of the line-up. Also, like Functional Size, it is insensitive to the addition of useless (null) foils.

Effective Size has two main disadvantages. First, neither Effective Size nor any known transform of it has a known sampling distribution. Second, Effective Size is more complex to calculate than the previously proposed measures. This can be a drawback when communicating the purpose and meaning of Effective Size to non-psychologists such as lawyers, judges, and jurors.

DEFENDANT BIAS

Malpass (1981) developed a measure of the bias of a line-up: 'Defendant Bias'. It is similar to the difference between observed and expected proportions of mock witness identifications, except that the Effective Size of the line-up is used to estimate chance expectation rather than the nominal size of the line-up. Because it is based on Effective Size, Defendant Bias has the same drawbacks in lack of intuitive interpretation by the laity.

Malpass and Devine (1983) studied the operating characteristics if these indices for four line-ups varying in composite similarity among the foils and the suspect. In the process of doing so, they offered two additional indices: Chi^2 and 'Acceptable Foils'.

CHI^2

Chi^2 is simply the application of this statistical technique to the distribution of mock witness choices across the suspect and the foils. A significant result merely indicates that the distribution of mock witness choice frequencies differs from the distribution that would be expected given completely random choices by mock witnesses. The Chi^2 measure has the disadvantage of being very sensitive. Significant departures from chance expectation are found almost universally.

ACCEPTABLE FOILS

The Acceptable Foils index begins by designating an arbitrary percentage of chance expectation for suspect and foil identification as the level below which a foil must not go without becoming unacceptable as a member of the line-up. If chance expectation were 1/6 (= 16.67%), then at a 75% criterion, a foil falling below 12.5% mock witness choices would be deemed unacceptable. The Acceptable Foils measure equals the number of foils considered acceptable using this process. Acceptable Foils has the advantage of being intuitively interpretable and accessible to the understanding of nearly all the players in the criminal justice system, including jurors.

HOW MUCH BIAS? HOW MUCH SIZE? WHAT PERCENTAGE OF CHANCE EXPECTATION?

A central problem with all the indices proposed has to do with the criteria of evaluation. How many mock witnesses should be used so that the test is neither too sensitive nor insensitive? What alpha level should be used for tests on proportions? What proportion of chance expectation should be required for an Acceptable Foil? Other critical questions also require answers. What are the sampling distributions for Functional Size, Effective Size, and Defendant Bias? Successful application of measures of line-up fairness in applied settings will require either clear answers to these questions or at least some understanding of the implications of deciding to measure fairness in the various ways available. These issues have been addressed to some degree previously by Tredoux (1998) and are expanded on further by him in this issue (Tredoux, 1999).

HOW SHOULD THE QUESTION BE PHRASED?

A methodological issue not previously addressed in the literature is the phrasing of the question used to obtain decisions from mock witnesses. Should the procedure be forced choice or should the mock witness be permitted to state that no line-up member appears to be a good fit to the description? Should mock witnesses be instructed to select the suspect, the person they think committed the crime, the person they believe the witness was describing, or perhaps the person who best fits the description? Will different questions generate different results? Wells and Bradfield (1999) provide a first attempt to explore this important question.

IS THE FAIRNESS EVALUATION PROCESS ROBUST TO LINE-UP PROCEDURAL CHANGES?

The entire fairness evaluation process is based on the assumption that mock witnesses provide a valid basis for estimating the probable eyewitness choice behavior of real witnesses. If mock witnesses prefer the suspect beyond chance expectation, then real witnesses will be subjected to approximately the same degree of biasing pressure from the structure of the line-up. One can question whether techniques for fairness evaluation developed with one form of line-up presentation will apply when an alternative form of presentation is employed. Corey *et al.* (1999) examines the usefulness of the mock witness technique for assessing the fairness of line-ups presented simultaneously versus sequentially. Valentine and Heaton (1999) compare the fairness of British line-ups (or identity parades) presented via videotape to those presented live.

DO THE MEASURES SHOW CONVERGENT VALIDITY?

The validity of line-up fairness measures is certain to be influenced by the degree to which the various measures lead to the same versus different conclusions regarding the fairness of line-ups. In addition to the measures defined by psychologists, the fairness of line-ups is assessed intuitively by police and the courts (Wogalter *et al.*, 1993). As a result, agreement between criminal justice personnel and researchers on the fairness of line-ups is also an issue. This topic has been addressed previously (e.g. Brigham and Brandt, 1992) and is expanded on further in this issue by Brigham *et al.* (1999).

DO LINE-UP FAIRNESS MEASURES GENERALIZE ACROSS RACES?

The cross-race issue has been extensively researched with regard to identification. Generally, people are better at identifying members of their own race than members of other races (Chance and Golstein, 1996; Ng and Lindsay, 1994). Brigham and Ready (1985) demonstrated that the race of a mock witness can influence the result of a test of line-up fairness. Lindsay *et al.* (1999a) extend our knowledge in this area by considering the race of the witness providing the description, race of the mock witness, and race of the perpetrator as factors in measuring line-up fairness. Their studies examine these effects for Asian, Black, and White witnesses, mock witnesses, and criminals.

DO THE MEASURES SHOW CRITERION VALIDITY?

At least as important as the methodological issues outlined above is the issue of criterion validity. All the discussion to this point has accepted measures derived from the mock witness task as useful to assess line-up fairness; but there are few data addressing the validity of mock witness choices as predictors of identification decisions by eyewitnesses. A line-up procedure that leads to the identification of the suspect regardless of whether the suspect is guilty or innocent is clearly unfair. On the other hand, there is no known identification procedure that can guarantee correct identification of guilty suspects nor completely eliminate identifications of innocent suspects.

The original motivation that led to the mock witness task was the assessment of a line-up that had already been used. In court, mock-witness data will be presented by the defence, the side wanting to suggest that a false identification has occurred. The relationship between measures of line-up fairness and correct identification is thus unlikely to be of great practical significance.

On the other hand, data are needed to demonstrate the relationship of fairness measures to the probability of an innocent line-up member being identified. Furthermore, different identification procedures lead to varying rates of false identification (Lindsay and Bellinger, 1999; Lindsay and Wells, 1985; Lindsay *et al.*, 1997). This leads to the question of whether any measure of fairness provides a reasonable criterion for all identification procedures. Lindsay *et al.* (1999b) provide data addressing this issue.

ARE LINE-UP FAIRNESS MEASURES STILL RELEVANT?

The mock witness task and issues of measuring line-up fairness were first raised when all line-ups were conducted simultaneously and we knew little or nothing about the impact of alternative identification procedures. Research has shown that the impact of poor foils can be reduced dramatically by presenting line-up members sequentially (Lindsay *et al.*, 1991). Recently, more radical alternatives to traditional line-ups have been suggested as a means of reducing wrongful convictions as a result of identification errors (Levi, 1998). Levi (1999) argues that such results render moot the issue of line-up fairness as measured by the mock witness task. Both Levi (1999) and Lindsay *et al.* (1999b) conclude that our efforts should be directed toward encouraging the use of superior identification procedures rather than measuring the faults in traditional but inferior ones.

CONCLUSION

Fairness in the construction and administration of line-ups is of undeniable importance in criminal cases. Implementing identification procedures so as to maximize their validity and to minimize obtaining false evidence against innocent persons are important aspects of justice processes. It is clear from recent studies of exoneration through analysis of DNA evidence (e.g. Connors *et al.*, 1996) that eyewitness evidence can result in the conviction and incarceration of innocent persons. For these reasons it is important to have available processes and techniques for the evaluation of the fairness of a line-up prior to its use (by investigators) and after its use (by defence counsel). The technology of such evaluation is relatively young, and current research is changing our preferences for techniques and the way they are applied. The present set of papers explores a number of related issues and prepares the ground for a more advanced and refined round of research on these important matters.

ACKNOWLEDGEMENTS

Preparation of this manuscript was supported by a grant from the Social Sciences and Humanities Research Council of Canada and the National Science Foundation of the United States of America.

REFERENCES

Brigham, J. C. and Brandt, C. C. (1992). Measuring lineup fairness: Mock witness responses vs. direct evaluations of lineups. *Law and Human Behavior*, **19**, 475–489.

Brigham, J. C., Meissner, C. A. and Wasserman, A. W. (1999). Applied issues in the construction and expert assessment of photo lineups. *Applied Cognitive Psychology*, **13**, S73–S92.

Brigham, J. C. and Ready, D. J. (1985). Own-race bias in lineup construction. *Law and Human Behavior*, **9**, 415–424.

Chance, J. E. and Goldstein, A. G. (1996). The other-race effect and eyewitness identification. In S. L. Sporer, R. S. Malpass and G. Koehnken (Eds.), *Psychological issues in eyewitness identification*. Mahwah, NJ, USA: Lawrence Erlbaum Associates, Inc.

Connors, E., Lundregan, T., Miller, N. and McEwan, T. (1996). *Convicted by juries, exonerated by science; Case studies in the use of DNA evidence to establish innocence after trial*. Alexandria, VA: National Institute of Justice.

Corey, D., Malpass, R. S. and McQuiston, D. E. (1999). Parallelism in eyewitness and mock witness identifications. *Applied Cognitive Psychology*, **13**, S41–S58.

Doob, A. N. and Kirshenbaum, H. M. (1973). Bias in police lineups–partial remembering. *Journal of Police Science and Administration*, **1**, 287–293.

Levi, A. M. (1998). Protecting innocent defendants, nailing the guilty: A modified sequentual line-up. *Applied Cognitive Psychology*, **12**, 265–275.

Levi, A. (1999). An honourable discharge for lineup fairness measurement. *Applied Cognitive Psychology*, **13**, S121–S124.

Lindsay, R. C. L. and Bellinger, K. (1999). Alternatives to the sequential lineup: The importance of controlling the pictures. *Journal of Applied Psychology*, **84**, 315–321.

Lindsay, R. C. L., Lea, J., Nosworthy, G. J., Fulford, J., Hector, J., LeVan V and Seabrook, C. (1991). Biased lineups: Sequential presentation reduces the problem. *Journal of Applied Psychology*, **76**, 556–564.

Lindsay, R. C. L., Pozzulo, J. D., Craig, W., Lee, K. and Corber, S. (1997). Simultaneous lineups, sequential lineups, and showups: Eyewitness identification decisions of adults and children. *Law and Human Behavior*, **21**, 404–411.

Lindsay, R. C. L., Smith, S. M. and Pryke, S. (1999b). Measures of lineup fairness: Do they postdict identification accuracy? *Applied Cognitive Psychology*, **13**, S93–S107.

Lindsay, E. C. L., Ross, D. F., Smith, S. M. and Flanigan, S. (1999a). Does race influence measures of lineup fairness? *Applied Cognitive Psychology*, **13**, S109–S119.

Lindsay, R. C. L. and Wells, G. L. (1985). Improving eyewitness identifications from lineups: Simultaneous versus sequential lineup presentation. *Journal of Applied Psychology*, **70**(3), 556–564.

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.

Ng, W. and Lindsay, R. C. L. (1994). Cross-race facial recognition: Failure of the contact hypothesis. *Journal of Cross-Cultural Psychology*, **25**(2), 217–232.

Tredoux, C. (1998). Statistical inference on measures of lineup fairness. *Law and Human Behavior*, **22**, 217–237.

Tredoux, C. (1999). Statistical considerations when determining measures of lineup size and lineup bias. *Applied Cognitive Psychology*, **13**, S9–S26.

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

Wells, G. L. (1993). What do we know about eyewitness identification?. *American Psychologist*, **48**(5), 553–571.

Wells, G. L. and 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., Leippe, M. R. and Ostrom, T. M. (1979). Guidelines for empirically assessing the fairness of a lineup. *Law and Human Behavior*, **3**, 285–293.

Wogalter, M. S., Malpass, R. S. and Berger, M. A. (1993). How police officers construct lineups: A national survey. *Proceedings of the Human Factors and Ergonomics Society* (pp. 640–644). Santa Monica, CA: Human Factors and Ergonomics Society.

