To appear as Chapter 9 in R. Lindsay, D. Ross, J. D. Read & M. P. Toglia (Eds.), Handbook of Eyewitness Psychology (Vol. 2): Memory for People. Lawrence Erlbaum & Associates.

Lineup Construction and Lineup Fairness

Roy S. Malpass

University of Texas at El Paso

Colin G. Tredoux

University of Cape Town

Dawn McQuiston-Surrett

Arizona State University

1. Introduction.

Police lineups come from English criminal law and procedure. According to Devlin (1976), lineups were instituted through a Middlesex magistrate's order in the mid 19th century. They were intended as a 'fair' replacement for the practices of courtroom identification, and showups, which were widely used in 19th century England, but widely recognized as potentially unfair to the defendant. Their origin indicates that the notion of 'fairness' is their *raison d'etre*. They are intended to secure an identification that can potentially incriminate someone, but also is fair to those who are subjected to it, particularly those who are innocent of the crime.

Study of the case law in many countries, as well as recent DNA-based exonerations in the U.S., indicates that lineups are not invariably fair – many innocent people are convicted after identification from a lineup by an eyewitness. The problem is significant because eyewitness evidence is cited as the most significant source of wrongful conviction (Scheck, Neufeld, & Dwyer, 2001). The DNA exoneration cases where the false conviction is established with near certainty show that eyewitness evidence has been largely responsible for false conviction in more than 70% of cases (www.innocenceproj.org). Wrongful identifications result from a number of failures of police procedure, however many of these are minimized when the lineups presented to witnesses are fair. If witnesses are induced to make a lineup identification, a fair lineup will expose the innocent suspect to an identification risk of 1/the number of persons in the lineup. A lineup biased against the suspect (because s/he stands out in some manner) adds additional risk. A lineup in which only two members (the defendant and 1 other) fit the witness' description of the offender increases the identification risk to 1/2. It follows that we must be able to construct fair lineups, and to diagnose the fairness of lineups.

The original guidelines of how to construct lineups, as they have come down to us in the English law, contained these key ideas i) to put a 'sufficient number of men…' in a line, ii) '…who bear a reasonable resemblance to each other in physical appearance…', iii) '…such that the identity of the suspect is not suggested to the witness…'. In practice, the 'sufficient number' came to mean eight in England, and six in the U.S. The degree of resemblance proved impossible to specify precisely, but case law in England evolved to specify that the resemblance should extend to height, weight, clothing and general appearance.

We can surmise from this that the original thinking behind the lineup was that it should consist of i) a number of 'plausible' fillers (i.e., non-suspect lineup members who adequately resemble the suspect), and ii) that the identity of the suspect should not be suggested by the manner of construction of a lineup, or the way in which it is conducted. From the beginning, then, these two principles persist:

- The suspect should not stand out from the other lineup members, and neither should any filler.
- Fillers should be equally good alternatives to the suspect.

2. Lineup Structure.

Decades of empirical research suggests that mistaken eyewitness identifications are more likely to occur when the suspect stands out in a lineup. Police lineups can even appear unfair to independent observers. For example, in a South African case where an eyewitness reported a robbery involving three Indian perpetrators, the police used a lineup containing three Indian suspects and three White fillers (Pelwani v. S.) 1963 (2) (PH) H237 (T). The court dismissed evidence of identification from the lineup, reasoning that the witness had pointed out the only three people on the lineup that he could have pointed out: that is, the lineup consisted in effect of only three members. A similar case in the U.S., in which the suspect was described as a Black man, used a 6 person lineup containing one Black suspect and 5 White fillers (Ellison & Buckhout, 1981). The egregious unfairness of these lineups is probably beyond dispute, but for many lineups the case is not as clear. Constructing lineups that are fair from the onset of the police investigation is important. Indeed, the National Institute of Justice Research Report entitled *Eyewitness Evidence: A Guide for Law Enforcement*[1] (The Guide) (*Technical Working Group, 1999)*, addresses this as the opening principle in the section on composing lineups: "Fair composition of a lineup enables the witness to provide a more accurate identification or nonidentification." (p. 29).

2.1. Constructing Fair Lineups.

Two separate but related constructs have been developed in psychological research on lineup fairness: lineup bias and lineup size (Malpass, 1981; Malpass & Devine, 1983; Wells, Leippe & Ostrom, 1979).

Lineup bias. An unbiased lineup is one in which persons without the visual experience possessed by a witness choose the suspect with a frequency approximating chance expectation, where chance expectation is defined as 1/number of lineup members. Lineups can be biased towards the suspect (more identifications of the suspect than expected) or away from the suspect (fewer identifications than expected). From the perspective of protecting the potentially innocent suspect from false identification, the suspect should not stand out in the lineup as being physically different from the fillers so as to draw extra attention to him, or to suggest his status as "suspect". The Guide expresses this point in the following way: "The investigator shall compose the lineup in such a manner that the suspect does not unduly stand out" (p 29). Pelwani v. S. (1963) provides an excellent example of a lineup that is unfair because it is biased towards the suspect – in that case, three of them! In a fair lineup, attempts should be made to include fillers in the lineup that have been matched on the suspect's general physical characteristics as stated in the verbal description of the suspect given by the witness, or, if an adequate description is not available, with reference to the appearance of the suspect or his/her lineup photo. When the suspect stands out in the lineup relative to the other lineup members, uncertain eyewitnesses may be cued to identify the suspect based simply on his distinctiveness rather than a true match between their memory of the culprit and that lineup member.

The concept of lineup bias is rather obvious. If the witness describes an offender with a scar on his right cheek and the suspect is the only person in the lineup with a scar on the right cheek, the lineup is clearly biased towards the suspect. The Guide provides this caution: "Create a consistent appearance between the suspect and fillers with respect to any unique or unusual feature (e.g., scars, tattoos) used to describe the perpetrator by artificially adding or concealing that feature." (p 30). Any attribute of the photographs or their presentation that causes the suspect to stand out is potentially biasing.

Lineup size is the degree to which the fillers in the lineup are viable alternatives to the suspect based on the witness's description of the culprit or based on their physical similarity to the suspect. The Guide puts it this way: "Select fillers who generally fit the witness' description of the perpetrator. When there is a limited or inadequate description of the perpetrator provided by the witness, or when the description of the perpetrator differs significantly from the appearance of the suspect, fillers should resemble the suspect in significant features." (p 29).

The effectiveness of the fillers in a lineup as alternatives to the suspect is of great importance. When a filler is not a plausible alternative to the suspect, the function of that individual in the lineup

[1] The section of the Guide (Technical Working Group, 1999) on composing a photo lineup is quoted in Appendix 1.

is reduced. The number of individuals in the lineup is an important matter. When the number of persons in a lineup is 6, a witness who cannot identify the offender but makes a lineup choice anyway will have a 1 in 6 chance of wrongfully choosing the innocent suspect when the offender is actually absent. The corresponding risk in a four person lineup is 25%. Few people would want to risk the loss of their freedom on a 3:1 bet!

When a lineup includes members who do not fulfill their role as acceptable alternates to the suspect, the lineup is effectively smaller than its actual size. Fillers who are not viable alternatives serve no purpose, and the risk of mistaken identification is increased because the expectation that an innocent suspect will be chosen merely by chance increases dramatically. Two measures of lineup size based on this principle and an analysis of the contribution of individual fillers to lineup size have been discussed in the literature: Effective Size (Malpass, 1981, Malpass & Devine, 1983) and E' (Tredoux, 1998, 1999). Both measures adjust the size of the lineup downward to reflect the failure of one or more fillers to perform as viable alternatives to the suspect.

Lineups should not be composed of individuals who are too similar to one another. The extreme form of a high similarity lineup is one where the lineup members are clones of each other. It is self-evident that such a lineup would render witnesses unable to make discriminatory identifications. While it is argued that lineup members should resemble each other in terms of general physical appearance and characteristics based on a witness's description of the culprit, some degree of variation among them is desirable (Wells, 1993; Wells, Seelau, Rydell, & Luus, 1994). The Guide contains this instruction: "Consider that complete uniformity of features is not required. Avoid using fillers who so closely resemble the suspect that a person familiar with the suspect might find it difficult to distinguish the suspect from the fillers." p. 30). Natural variation among lineup members is desirable because it allows a witness with a clear memory of the culprit to distinguish between similar individuals, while a witness without a clear memory may be unable to do so. Likewise, if the suspect is the offender, witnesses will be better able to make an identification, whereas when the suspect is not the offender witnesses will not be aided in identifying him. Variation in the physical appearance of lineup members is not bad unless it causes the suspect to stand out. The variation among lineup members should be meaningful to witnesses in the case where the suspect is actually the offender but not meaningful when the suspect is not the offender.

2.2. Selection of Fillers.

A number of methods for selecting fillers in lineup construction have been proposed and studied.

Perceptual Similarity to Suspect. The traditional method of lineup construction is a procedure in which lineup fillers are selected based on their physical similarity to the suspect. There are at least two methods for achieving this. First, fillers can be judged perceptually (subjectively) by the investigator (police officer or researcher) as to their degree of similarity to the suspect. This is the approach adopted in most police practice, but it is also prevalent in laboratory and field studies of eyewitnesses. Second, filler-suspect similarity can be judged by persons independent of the lineup task, and some criterion can be set so as to utilize as fillers only those persons that exceed the criterion; for example, a specified proportion of judges may have to agree that the photo is appropriate for the lineup, or the photo must surpass a predetermined point on a scale representing similarity of the prospective filler to the suspect. This approach is taken in a number of studies, including Lindsay and Wells (1980) and Malpass and Devine, (1983). However, little is in fact known of the systematic relation between degree of suspect-filler similarity and the fairness of lineups apart from the finding that lineups in which the fillers have low similarity to the suspect have high measures of bias, and low measures of Effective Size (Malpass and Devine, 1983). A recent study by Tredoux (2002) reports a potential method for measuring the facial similarity of lineup members from a principal component analysis (PCA) of facial images. The physical measure of similarity derived from the PCA predicted both the result of lineup evaluation procedures based on "mock witnesses" (see below) and eyewitness performance from photospreads. However, whereas low similarity lineups were shown to result in lineups with high bias and low Effective Size, low similarity lineups led to greater accuracy, in terms of hits and correct rejections, than moderate or high similarity lineups when used with eyewitnesses.

These lineup construction methods are an important attempt to follow the guideline discussed above that the suspect not be physically distinctive from the fillers. Indeed, the similarity strategy is reportedly used most often by police (over 80% of the time) when constructing lineups and photospreads (Wogalter, Malpass, & McQuiston, 2004). However, a series of studies by Wogalter and colleagues suggest that this lineup construction method can actually result in bias or suggestiveness toward the suspect (Laughery, Jensen, & Wogalter, 1988; Marwitz & Wogalter, 1988; Wogalter & Jensen, 1986; Wogalter, Marwitz, & Leonard, 1992). Specifically, research indicates that choosing fillers with reference to physical similarity to the suspect can result in the suspect having unique properties, such as being the "prototype" or having the most familiar face of the group. Paradoxically, this suspect-matched lineup procedure can still result in the suspect standing out. Luus and Wells (1991) also argue that a suspect-matched lineup can result in the lineup members being too similar to one another.

Witness' Verbal Description of Culprit. An alternate lineup construction method is the description-matched strategy in which fillers are selected based on the witness's description of the culprit instead of the suspect's physical appearance. The rationale for the description-matched procedure is that it meets the criteria for constructing a fair lineup in which the suspect does not stand out relative to the fillers, and at the same time allows for all lineup members to vary on general physical characteristics so as to not make the lineup task impossible for an eyewitness who has a good memory of the culprit (see Wells, Rydell, Seelau, & Luus, 1994). Indeed, the research literature shows an advantage for description-matched lineups over suspect-matched lineups in promoting accurate identifications (e.g., Wells, Rydell, & Seelau, 1993). However for this procedure to be effective a useful description of the offender is required from the witness(s). "White male, between 20 and 30 years old, 5'6" to 6'0" tall, brown hair and brown eyes wearing faded jeans, white athletic shoes and a black T shirt" will not prove very helpful in constructing an acceptable lineup using the description matched procedure since it is quite vague. This underscores a common problem: while they may be accurate, verbal descriptions elicited from witnesses and recorded in police reports are often not very descriptive of the offender in ways that will help to differentiate him/her from others with similar general features.

An additional concern arises when there are multiple eyewitnesses to a crime, each of whom give a separate description of the culprit to police. It is common for the general descriptions given by separate eyewitnesses to be similar, but for differences to appear in the more specific (and more useful) descriptors. If the descriptions are adequate for use of the match to description strategy, one way to deal with multiple eyewitnesses is to base the selection of fillers on a composite description of the culprit that incorporates the descriptors given by all of the witnesses. This, of course, depends largely on the degree of overlap between descriptions. If there are important differences between separate witness descriptions it may be necessary to construct different lineups for each witness. Of course if the witness descriptions are inadequate then a match to suspect approach would be more appropriate, and in that case variation in witness descriptions is not of great importance.

Modal Verbal Description Of The Suspect. If a verbal description of the culprit is impoverished, or absent, there is an alternate means of choosing fillers based on descriptions. It is to obtain a number of descriptions of the suspect from persons similar to the witness(s) in age, sex, ethnicity (and any other demographic variables thought appropriate to local conditions). The modal descriptors can be found and used to guide filler selection.

Our preference is to base filler choice on a verbal description of the culprit, if a useful one is available, and secondarily on physical similarity to the suspect (to augment the verbal.)

3. Evaluating the Success of Lineup Construction.

There are at least three contexts in which lineup fairness should be evaluated:

(1) Law enforcement should evaluate and document the size and bias of their lineups prior to use;

(2) Scholars using lineups in research should evaluate and document lineup size and bias as a matter of quantifying this aspect of the stimulus materials used in their work, as a guide to replication; and

(3) Defense attorneys should routinely evaluate the lineups that form a basis for an eyewitness

identification in their cases. If the lineup can be shown to have been structurally or procedurally unfair it may be possible to win a motion to suppress the identification evidence. Failing that, evidence of unfairness may be introduced in court.

It is not enough to follow the procedures outlined here to ensure that a fair lineup has been produced. Its degree of fairness should be demonstrated, and adjustments made if necessary. Psychologists have pioneered a methodology for doing this by developing techniques for constructing fair lineups and creating quantitative fairness measures. Research on lineup fairness started in earnest with a paper by Doob and Kirshenbaum (1973). They served as consultants in a Canadian case, R v. Shatford, which turned on a lineup identification made by a single eyewitness. The eyewitness made an identification at the lineup despite the fact that the only description she was able to give of the perpetrator was that he was 'attractive'. Doob and Kirshenbaum suspected that the witness was basing her identification on a 'memory fragment' or perhaps even on her earlier insubstantial description to the police. The second possibility seemed to have some weight, since the suspect was particularly attractive. In order to test this possibility, Doob and Kirshenbaum had 20 naïve subjects rate members of the lineup for attractiveness. The suspect received an attractiveness rating substantially higher than any of the fillers, suggesting that the suspect could be selected by a witness who remembered only that he was attractive.

In a second stage of the study, Doob and Kirshenbaum showed a photograph of the lineup to 23 'mock witnesses' (persons who had not been present at the original crime), along with the original eyewitness description of the suspect. They reasoned that if the lineup was 'fair', those who had not been present at the crime should not be able to identify the suspect, except by guessing. There were twelve people in the lineup, so the expected rate of guessing would be approximately 1/12. If significantly more mock witnesses than this selected the suspect, it would be evidence that the lineup was unfair, and that the structure of the lineup somehow suggested the identity of the suspect to the witnesses. Fourteen of the 22 witnesses selected the suspect, a result that would occur randomly with a probability less than 0.001. They concluded that the lineup was unfair, specifically that it was biased against the suspect.

This innovative research procedure pioneered more than thirty years ago by Doob and Kirshenbaum has become known as 'mock witness evaluation', and is the basis for almost all further work on post-hoc assessment of lineup fairness. The purpose of mock witness evaluation is to assess the structural fairness of the lineup – whether the fillers are adequate alternatives to the suspect and whether the suspect stands out from the fillers. Its purpose is not to predict what witnesses would do. The central assumption is this: if persons who have had no exposure whatever to the suspect prior to viewing the lineup can select him from the lineup with a probability greater than chance, then the lineup is biased towards identification of the suspect.

If a lineup is structurally biased, when an actual witness identifies the suspect, one has to ask whether s/he makes the identification based on memory of the person from the witnessed event or whether s/he is merely using the same minimal information used by mock witnesses to make the same choice.

3.1. Evaluating an Existing Lineup: Principles.

Evaluating an existing lineup is in principle very similar to lineup construction, using similar analytical concepts. The basic questions are:

- Does the suspect stand out from the other members of the lineup?
- Are the fillers adequate alternatives to the suspect?

The standard for answering these questions requires some clarification. Fairness evaluation is not based on replicating or simulating the perspective of actual witnesses. If we were to write an equation (or model) for predicting the response of a witness to an eyewitness identification lineup, it would include at least these factors / categories:

1. the amount / quality of the information (visual memory) the witness has about the offender,
2. these, and other factors, related to the likelihood of the witness making an identification (decision criterion), independent of the information they have:
    a. The witness' understanding of the lineup identification task.
    b. The witness' willingness to cooperate with the perceived wishes of the administering officer.

c. The degree to which the witness believes that the administering officer wants him/her to choose someone from the lineup.

3. The fairness of the lineup.

The first category of variables cannot be known by psychologists or by the criminal justice system independently of the identification process, and are not under their control. The second category can be influenced to some degree by instructions / admonitions, and other aspects of the lineup procedure. But these are matters of lineup administration rather than the structural fairness of the lineup itself. The structural fairness of the lineup, on the other hand, can be known in great detail and is completely under the control of the criminal justice system. The purpose of the mock witness procedure is to quantify structural lineup fairness, not to estimate the effects of the other two classes of influences on the eyewitness' lineup choice. In this way assessment of the role of lineup structure in the actual witness's response to the lineup is separated from the other factors which influence them.

The mock witness paradigm controls the first classes of variables by holding them constant and eliminating them as contributors to the result of the mock witness process. It provides either no information about the suspect or very particular (but non-visual) information. It instructs each mock witness to choose one of the members of the lineup. In this way the mock witness procedure is sensitive only to variations in the structure of a lineup (the relationships in physical and attributional appearance among the lineup members), or the structure of a lineup given the descriptive information provided to the mock witnesses, and not other factors that might influence the likelihood of making a lineup choice at all.

When a witness makes an identification there may be ambiguity about how to interpret it. The identification could come about as a result of the witness choosing the suspect from the lineup solely on the basis of his or her memory that this specific person is the person he or she saw committing the witnessed crime. Or, the identification could come from a more complex process: the witness may make a cluster of judgments that result in an identification. The witness may believe that s/he would not be called to examine a lineup if the police did not have a suspect, and further that the police are seldom wrong, so it is highly likely that the actual culprit is in the lineup. Given these beliefs, it would be easy for the witness to approach the task by looking for the one member of the lineup who is most likely to be the culprit. In such a case a suspect who stands out from the fillers, even when factually innocent, is at risk of identification by a wholly rational, cooperative and well-intentioned witness.

Witness identifications cannot be interpreted directly, in an absolute way. The basis for this is developed by Wells (1993) in the "lineups as experiments" analogy. In order to interpret an identification a control comparison is required that allows the finders of fact to know the result of identifications made by people who had no visual information at all about the suspect in the lineup. Experiments include control groups because the effects of experimental treatments require interpretation against some base or standard condition. This idea also characterizes lineup identifications: the witness identification cannot be interpreted without information about the effects of background factors, prominent among them being the effect of the lineup's structure – its fairness in the sense of both size and bias.

When a lineup is biased towards the suspect the witness may not need any memory at all of the criminal or the event in question in order to know which member of the lineup is the suspect. A good analogy for this is multiple choice tests in educational settings: if a multiple choice item is given to students in which the correct choice is known because all of the alternative choices are obviously false, the item is no test of knowledge of the subject.

3.2. Evaluating an Existing Lineup: Procedure.

Information Given To Mock Witnesses. There is a range of information that can be provided to mock witnesses as a basis for their lineup choice. They could be given no information at all, and simply be asked to indicate which member of the lineup is the police suspect. If mock witnesses can reliably choose the suspect based on no particular information apart from a simple inspection of the lineup, then identification by an actual witness adds no information about the guilt of the suspect. Alternatively, mock witnesses could be given the verbal description the witness gave to law enforcement, as in Doob and Kirshenbaum (1973). Another alternative is to provide mock witnesses

with a modal or composite description of the suspect. Each of these is described below.

Culprit description. Verbal descriptions collected by law enforcement appear not to be elicited with the purpose of obtaining individuating information. Many printed / online forms used for police reports contain categories that are not conducive to this purpose. As a result descriptions are often far less effective than they might be. Particularly, they are less an aid to lineup construction than they could be.

If the lineup fillers were chosen on the basis of the culprit description, then one matter of interest is whether the suspect stands out with respect to the culprit description. In that case, the information given to mock witnesses is the description given by eyewitnesses as obtained from police records. This is the classic starting point for mock witness evaluation of lineups. If there are multiple but divergent descriptions given by a single witness a composite description can be made. But if there are divergent descriptions given by multiple witnesses, then as many lineup evaluations must be made as there are divergent descriptions.

Suspect description. Suspects are "nominated" through many routes besides the verbal description of the culprit. The police will often pick a suspect based on knowledge of persons in the area, and through forms of information unrelated to facial appearance. For these and other reasons, the suspect may bear little similarity to the description(s) given by eyewitnesses. When this occurs, the question still is whether s/he stands out from the fillers. To evaluate this possibility mock witnesses may be given a "modal" description of the suspect. To obtain a modal description the investigator should obtain a photograph of the suspect taken close to the time of the offense. Then using that photograph, obtain a description of the suspect from a sample of judges who are of the same demographic categories as the witness, with respect to approximate age, sex, ethnicity and social class (primarily because descriptive vocabulary may vary along these lines). While the purpose is not to simulate the eyewitness, using descriptive language unfamiliar to the witness(es) may give a misleading conclusion as to whether or not the suspect stands out, as described.

There is at least one special instance in which a modal description can be useful. If the witness has seen or helped to construct a composite image of the culprit during the investigation process his/her memory for the appearance of the culprit is likely to have been modified (Topp, McQuiston & Malpass 2003). In such a case it would be appropriate to use modal description information based on the composite image as the description given to mock witnesses.

No description. Mock witnesses may generally believe that their task is to choose based on the descriptive information provided them (McQuiston & Malpass 2002). However, we know that there is more to lineup identifications than that. Research shows that if a lineup member's appearance fits a criminal appearance stereotype his/her likelihood of being identified as the culprit is increased (McQuiston & Malpass, 2002). For this reason it is possible that the suspect may stand out from the fillers on a basis that is not related to culprit or suspect descriptions, but is based on other grounds that may not be clearly discernable. As argued above, the most basic evaluation of lineup fairness is one in which the lineup is displayed with no other information given. If mock witnesses choose the suspect with greater than chance frequency the lineup is unfair in either size, bias, or both.

Composite. Composite images made during the course of an investigation are known to be highly variable in their similarity to the offender (see Davies & Valentine, this vol.; Shepherd & Ellis, 1996). It may be of interest to show mock witnesses a copy of the composite image when the witness has constructed or has been shown a composite, and when the original witness description does not resemble the suspect. The rationale, of course, is to inquire into whether the witness' memory has been contaminated: Any identification of the suspect following viewing of a composite raises the question whether the identification is based on the witnesses memory for the face of the offender at the time of the offense, or based on their memory of the composite. Topp, McQuiston & Malpass (2003) show that for witnesses who participated in constructing a composite and then subsequently viewed it at least once there was a contamination of their memory for the target face in the direction of the composite.

Our recommended approach is to routinely do multiple mock witness evaluations based on no description at all, one of the description types, and on a composite if one was produced. This provides a rough quantitative dimension to the question of how much information is necessary for

identification of the suspect: mock witnesses with no personal individuating information, mock witnesses with a verbal description of the offender, or an eyewitness with the unique individuating information that comes from having seen the person before, presumably committing the offense in question. Once more, the logic is that if one can identify the suspect based on the knowledge provided to mock witnesses in the first two of these, one does not need to have had the knowledge possessed by an eyewitness, in the third, and the meaning of an eyewitness identification is ambiguous.

4. Quantitative Measures of Lineup Fairness.

To preserve the distinction between the size and bias aspects of lineups we will discuss these separately. Before proceeding with that discussion, however, the desired attributes of quantitative indicators of lineup fairness require a brief discussion.

4.1. Desiderata for Lineup Fairness Measures.

What are the *desiderata* for measures of lineup size and bias? First, the measures must reflect the theoretical meaning of the concepts to be measured. Therefore, bias should reflect the degree to which the suspect in a lineup will be chosen at a rate above or below chance expectation (where chance expectation is based on the number of people in the lineup). Lineup size should reflect the adequacy of the individual fillers as alternatives to the suspect.

- The measures should be bounded by the range of the phenomenon. Size measures, for example, must never yield "sizes" that are greater than the number of persons in the lineup.
- The measures should be a calibrated monotonic function of changes in the underlying latent variable. This means that the numerical index representing size should increase or decrease proportionately, with each increase or decrease in the adequacy of a filler, and the index representing bias should change proportionately with each change in preference for the suspect. As bias or size are manipulated by experimenters, the measures should respond in a calibrated monotonic manner.
- The measures should have a sampling distribution that allows for inferential statistical analysis.
- The measures should be accessible to law enforcement officers and the public so that clear and understandable explanations of their meaning and interpretation can be given.

4.2. Does the Suspect Stand Out? Measures of Lineup Bias.

The mock witness procedure, as originally conceptualized by Doob and Kirshenbaum, uses a measure of lineup bias. Bias is the extent to which the proportion of mock witnesses choosing the suspect is greater or less than expected by chance. When the proportion of mock witnesses choosing the suspect equals that expected by chance (i.e. 1/k, where k is the number of lineup members), the lineup is unbiased. When it deviates from the expected value, it is biased. The measure has interpretable limits at both the upper and lower ends. As the proportion approaches unity, mock witnesses are choosing the suspect to the exclusion of the fillers, and when the proportion approaches zero, witnesses are failing to choose the suspect at all. However, one can expect the proportion to show random sampling variation, and an important question thus concerns how to interpret the observed proportion. For example, if 7 of 20 mock witnesses (35%) choose the suspect from a five-person lineup, we need to know whether this could reasonably be explained as chance variation from the expected value of 4 of 20 (20%). Doob and Kirshenbaum used a z-test to make this decision, but Tredoux (1998) recommends the direct calculation of Binomial probabilities instead. The latter method does not make the assumption of an approximating distribution. Both Wells, Leippe and Ostrom (1979) and Tredoux (1998) recommend reporting the proportion as a confidence interval rather than as a point estimate, and Tredoux (1998) refers readers to a formula that is more accurate than that regularly used for computing confidence intervals around proportions.

Usually, eyewitness researchers are interested in situations where the lineup is biased against the suspect (i.e., where the suspect is chosen by a higher proportion of mock witnesses than is expected by chance), but it is also possible that the lineup could be biased in favor of the suspect. Imagine a lineup where 1 of 30 mock witnesses (3%) chooses the suspect from a five-person lineup. This is significantly less than we would expect by chance (binomial $p < 0.01$). On the one hand, this might appear to be of no consequence to the police

or to eyewitness researchers, as the suspect's liberty has not been jeopardized unfairly. However, this approach only recognizes the problem of false identifications, and fails to recognize that a second kind of error can be committed when using a lineup, namely the failure to identify a guilty perpetrator. Lineups in which suspects are chosen by mock witnesses at levels significantly below chance responding are poorly constructed and run the risk of committing the second kind of error. Such lineups may be rare, but it is useful to extend the reasoning behind the measure of lineup bias so that it can be used as a warning indicator for both types of fundamental error.

4.3. 'Nominal', 'Functional' and 'Effective' Size.

The nominal size of a lineup is simply the number of people appearing: suspect plus fillers. The nominal size sets a limit on the *a priori* risk to which an innocent person is exposed. For the most frequent nominal size used in the United States (6), the risk of a false identification from a perfectly fair lineup by a poor but willing eyewitness is 16.67%. A major step towards decreasing the risk of false identification to innocent suspects would be to increase the nominal size of the lineup. An increase to 12 reduces the risk to 8.3%. The degree of the false identification risk ought to be a matter for policy discussions. Assuming that the lineups are fair, increasing nominal size is a simple and effective method for reducing the probability of false alarms. The technology to support this is available where photospread or video lineups are considered acceptable alternatives to corporeal lineups.

A lot depends on whether lineups which contain 5 (or 11, or 19) fillers actually contain that many realistic alternatives to the suspect: are fair in the "size" sense. A lineup containing a White suspect, two White fillers, an African American, a large dog and a refrigerator has a maximum size of three (similar in effect to Pelwani v. S, 1963), and if the two other White fillers are very different from the suspect, perhaps even lower. The problem here is to develop quantitative measures that reflect the number of lineup members who act as viable alternatives to the suspect, thus detecting whether the lineup size is effectively smaller than it appears.

Most legal jurisdictions prescribe the number of lineup members (e.g. six in most of the U.S., eight, ten, or twelve in various parts of Canada, nine or ten in England), and fillers who are selected for the lineup are required to be reasonable matches to the suspect on general attributes such as height, weight, hair color, and facial appearance. The notion of a 'plausible filler' is inscribed in legal understandings of what a fair lineup should be, and it is important therefore to attempt to measure the number of plausible members in a lineup. We can distinguish between the nominal size of a lineup (i.e. how many people are in it), and its 'Effective Size' (i.e. the number of plausible lineup members) (Malpass, 1981).

Wells, Leippe and Ostrom (1979) recognized this, and proposed a proxy measure known as 'Functional Size', which is the reciprocal of the proportion of mock witnesses choosing the suspect. Thus, if 10 of 20 people choose the suspect from a 5 person lineup, the reciprocal is 20/10 = 2, and the lineup has a Functional Size of only 2. In the hypothetical lineups d and e of Table 1, Functional Size is 2 and 6 respectively, and this corresponds quite well to the apparent difference in number of plausible fillers between the lineups (i.e. from visual inspection of the array frequencies).

Insert table 1 about here

The concept underlying 'Functional Size' (i.e. to distinguish the number of nominal and plausible lineup members) is a good one, but there are a number of reasons to consider alternative measures. First, Functional Size is not a sufficient estimator: it takes account only of the number of mock witnesses choosing the suspect and does not consider the fillers except in aggregate. It is possible to arrive at values of Functional Size which suggest a lineup with an acceptable number of plausible fillers, where this is clearly not the case. Such an example is shown in lineup f of Table 1.

Second, it is possible to arrive at values of Functional Size that cannot be interpreted in any reasonable sense as the number of plausible lineup members. In the case of lineup g in Table 1, for instance, the nominal size is 6, and the Functional Size is 100. This estimate is not meaningful in the sense of indicating the number of plausible lineup members.

Third, it is possible for the Functional Size of a lineup to be identical to its nominal size and for the distribution of identifications to exhibit a clearly different picture about the number of plausible fillers. Lineup f of Table 1 is one example. The

Functional Size of the lineup is 6, suggesting 6 plausible lineup members, when there are only two. It is easy to imagine many other patterns of distribution of mock witness choices across the fillers that show a similar result.

At the heart of the matter is the fact that Wells et al.'s (1979) measure of Functional Size is not a size measure, but rather a bias measure. It is the reciprocal of the proportion of mock witnesses who choose the suspect, and although it may be possible to give it a quasi-size interpretation, its computational basis restricts it to a statement about how frequently the suspect is chosen.

Effective Size. Malpass (1981) suggested 'Effective Size' as a measure of the number of plausible lineup members. Effective Size has a maximum of k, the number of lineup members, and a minimum of 1 (assuming that mock witnesses must choose one member of the lineup). Starting from the maximum value, a subtraction is made for each lineup member who is chosen at a rate that differs from chance expectation.

The assumption underlying the notion of Effective Size is appealing: one or more of the fillers in a lineup may present an inadequate test of a witness who has little more than general knowledge of the appearance of the offender, and we shouldn't take the ability of a witness to reject such fillers very seriously. The calculation of Effective Size acts on this assumption by reducing the nominal size of the lineup according to departures of proportionate identification of individual fillers from that expected by an equiprobability model (every lineup member drawing the same number of mock witness choices). For many distributions of identifications the measure gives an indication of the number of fillers that could reasonably be considered present, at least from visual inspection of array frequencies. Lineups h, i and j in Table 2 are clear examples.

Insert table 2 about here

However, there are a number of weaknesses with the Effective Size measure (see Tredoux, 1998). Most importantly, there is no known sampling distribution for Effective Size, which weakens the kinds of conclusions researchers or practitioners can draw about particular lineups.

Tredoux (1998) suggested an alternative computational formula for Effective Size which retains most of the desirable properties, with the important added benefit of a known sampling distribution. Specifically, the measure has a maximum of k, the number of lineup fillers, and a minimum of 1 (assuming that mock witnesses are required to choose a lineup member). If some lineup members attract more choices than others, this will result in a reduction of the value of E' from k towards 1. The formula is as follows:

$$'E' = \frac{1}{1 - \underline{I}}$$

where I is defined as $1 - \sum_{i=1}^{k} \left(\frac{O_i}{N} \right)^2$

where oi = the observed number of mock witnesses who choose lineup member i; N = the number of mock witnesses; and k = the number of lineup members).

Methods for using E' inferentially can be found in Tredoux (1998). These are relatively uncomplicated, and can be incorporated into a spreadsheet for easy computation (download http://www.eyewitness.utep.edu/images/size-calc.xls for an example). Some authors have reported a high correlation between E' and Effective Size (Corey, Malpass, & McQuiston, 1999; Tredoux, 2002), but the relation has not been systematically investigated. In order to do so, we computed some simulation data over different numbers of mock witnesses and different sizes of lineup. Lineup frequencies were randomly generated to produce a range of values for E between 1 and nominal lineup size. Five thousand lineups were generated for each combination of nominal size and number of mock witnesses. The correlation between E' and Effective Size is shown in Table 3 for each of these combinations. It is clear from Table 3 that E' and Effective Size are very closely related, even for relatively small samples of mock witnesses.

Insert table 3 about here

4.4. Are Lineup Size And Lineup Bias Independent?

There are important reasons why we should attempt to measure Effective Size as well as lineup bias when we evaluate lineups. A lineup may appear biased if the suspect is chosen at a rate

higher than expected by chance, but this might not be a fair conclusion if the Effective Size of the lineup is lower than the nominal size. Consider the lineup shown in Table 4. The suspect is chosen by 31% of the mock witnesses, whereas chance expectation is 16.67%. This is a statistically significant difference, but is the lineup biased against the suspect? If we take into account that two other fillers are chosen at exactly the same rate as the suspect, we will probably conclude that the lineup is not biased against the suspect, as there are other fillers that receive as many mock witness choices. What is much more of a problem is that the lineup has a low Effective Size. There are only three plausible identification alternatives in the lineup: the suspect and two fillers. The real danger to the suspect is a witness who chooses randomly among the plausible lineup members: the risk of mistaken identification is not the intended 1/6, but more like 1/3.

Insert table 4 about here

Effective Size clearly provides information about lineups over and above that given by lineup bias, but the measures are not independent. A lineup with high bias can also be a lineup with low Effective Size. However, the converse is not true. The relation between lineup bias (when defined as proportion of mock witnesses choosing the suspect) and Effective Size (when defined as by Tredoux, 1998) is shown in Figure 1. Note that in the right hand side of the figure the measures are strongly dependent, but in the left hand side they are much less dependent.

Insert figure 1 about here

4.5. Other Measures Of Lineup Fairness.

Defendant bias. Malpass (1981) proposed an alternate measure of lineup bias, which takes into account the Effective Size of the lineup. Ordinarily, we would consider 1/k to be the proportion of suspect identifications expected by chance guessing, but Malpass suggests that this might better be calculated as 1/[Effective Size]. The idea here is that the likelihood of being selected randomly by a witness is less a function of the nominal size of the lineup than a function of the number of plausible fillers present in the lineup. Then, instead of evaluating the proportion of suspect identifications against 1/k, we evaluate it against 1/[Effective Size]. This suggestion seems very sensible to us: in most police or research lineups we have seen, Effective Size is rarely equal in value to nominal size, and testing lineup bias for significance by comparing the proportion of suspect identifications to 1/k would be fairly meaningless – it will nearly always be significant. In line with our discussion earlier on lineup size, it makes more sense to us to first establish the Effective Size of the lineup, and then to determine whether the suspect is chosen from the remaining plausible fillers at levels greater than any of the other fillers.

Percentage below expectation. Malpass and Devine (1983) suggest a method for evaluating the suitability of individual lineup members, based on the extent to which the member is chosen below chance expectation in a mock witness task. Only those fillers that are chosen by mock witnesses above a specified proportion of chance expectation are considered to be acceptable fillers.

An alternative approach to measuring lineup fairness would then be to set a minimum number of plausible lineup members, and to determine whether the lineup meets this minimum size requirement. Tredoux (1998) argued that this approach is too much at the mercy of random sampling variation, and suggests that a better method may be to construct confidence intervals around the observed proportion of identifications that each filler receives, and to apply the minimum criterion test to the endpoint(s) of the intervals. Assume, for example, that we set a minimum criterion of 7% of mock witness choices per lineup filler, i.e. no filler should attract fewer than 7% of the mock witness choices. Now, if filler A receives 10% of the choices, but a 95% confidence interval around that point estimate of 10% is [6% - 14%], then the 7% criterion falls within the acceptable range, and we would therefore accept the filler for inclusion in the lineup.This has the benefit of attaching some level of probability to decisions made about the plausibility of fillers.

5. Validity of Mock Witness Evaluations.

Measures of lineup fairness are derived from mock witness evaluations of lineups. It is implicitly assumed that these measures tell us something about how eyewitnesses will perform on the same lineups. If a mock witness evaluation of a particular lineup produces a high bias estimate, this is taken to suggest that an eyewitness is more likely to choose

the suspect from that lineup, even if he is innocent, than from a low-bias lineup. Is this assumption warranted? Are mock witness measures of lineup fairness valid measures? Do lineups that are unfair in either sense, size or bias, lead to the expected increases in false identification?

There are good reasons for concern, as mock witnesses and eyewitnesses differ in several important ways. Mock witnesses by definition have no direct knowledge regarding the perpetrator's (or suspect's) appearance. Eyewitnesses, on the other hand, have access to memory of the perpetrator's appearance as well as to the verbal description they gave. If an eyewitness to a particular event has given an inaccurate description of the perpetrator, or if law enforcement procedure has elicited only an impoverished description, s/he may at the same time have an intact and more differentiated memory for the prepetrator's appearance, and s/he may identify an innocent suspect who resembles the perpetrator. Because the witness' visual memory is different from the verbal description in important ways, the development of a fair lineup based on the description may not represent a fair lineup based on the witness' visual memory. This can happen even if the mock witness evaluation (based pon a verbal description) suggests that the lineup is unbiased, and has a reasonable number of plausible fillers. Some studies suggest that eyewitnesses are generally not good at providing accurate verbal descriptions (Meissner, Sporer & Schooler, this volume). If lineups are constructed so that fillers match the description of the perpetrator rather than the suspect's physical appearance we can expect this problem to be significant: a match to description strategy should ensure a favorable mock witness evaluation, but if it is based on a poor description it may not adequately test the fairness of the lineup. The lineup evaluation process needs better descriptions from the investigation process, beginning with first responder reports. And those engaging in lineup evaluation (during their initial construction as well as later evaluations) should assess the quality of available descriptions. In particular, great care should be used in deciding whether or not to use a match-to-description strategy in lineup construction. There is little research directly on the point of the quality of verbal descriptions of faces, or the description attributes that make them useful.

The assumption that mock witness evaluations give us information about the fairness of police lineups should clearly be validated. While there the number of validation studies is small they give general support to the idea that mock witness evaluation gives us information about the fairness of police lineups.

Brigham and Brandt (1992) report a study in which overall judgments of fairness made by samples of law officers and college students were compared to estimates of Functional and Effective Size, as derived from mock witness evaluations. Twenty-three photo lineups were created so that 8 could reasonably be classified as fair, an additional 8 as moderately fair and 7 as least fair. The law officers and students first assessed the overall fairness of each lineup, and then evaluated each of the 5 fillers on a 6-point 'acceptability' scale. Across the 23 lineups, two of the mock witness measures, proportion of mock witness suspect-identifications and Effective Size, were consistently related to fairness measures derived from the student and law officer samples.

Lindsay, Smith and Pryke (1999) argued that the most important validation for lineup measures from a forensic point of view is criterion related: the measures should be able to predict the occurrence of false positive identifications (i.e. identifications of innocent suspects). Presumably, false positives are brought about by the use of biased/unfair lineups, and bias/fairness estimates – if they are valid - should be correlated with their occurrence. Lindsay et al. exposed participants to a staged crime and then conducted 18 lineups, some of which contained the perpetrator, and others which did not. In order to generate a larger sample size than would ordinarily have been obtained for the mock witness evaluation task, Lindsay et al. treated all of the innocent lineup fillers in the perpetrator absent lineup as suspects, one at a time. They used identification rates for each filler to determine both lineup bias and Effective Size. They then correlated measures of bias and size with the rate of false positive identification made by eyewitnesses for each of the 18 lineups. Unfortunately, by treating each filler as a suspect replacement their calculations of lineup size are redundant (see Tredoux, 1999, for a full version of this criticism). One lineup is turned into six lineups by treating each lineup member as a suspect, and then the same set of mock witness choices is used to calculate

effective size each time. The effective size is thus the same for each 'lineup' created in this artificial way, and the resulting predictor is invalid. Lindsay et al. report a significant correlation of 0.64 ($p < 0.05$) between lineup bias (proportion of mock witnesses choosing the suspect) and the rate of false positive identifications by eyewitnesses, and a non-significant correlation of 0.1 between lineup size (Effective Size) and the rate of false positive identifications. A later study by the same authors (Smith, Lindsay and Pryke, 1999), using much the same methodology, found a significant correlation between lineup bias and 'eyewitness accuracy', albeit of considerably smaller magnitude ($r = 0.2$). They did not use a measure of lineup size as a predictor, though, and their method of calculating lineup fairness used non-independent data in much the same manner as the earlier study.

Tredoux, Parker and Nunez (2004), noting the statistical dependency in Lindsay et al.'s (1999) study, and the similar dependency in Smith et al's (2000) study, conducted a staged crime experiment, testing the ability of lineup measures to predict eyewitness identifications of innocent suspects. Three lineups of varying degrees of fairness were created for each of two perpetrators, and evaluated by mock witnesses, as well as being used to test eyewitness identification accuracy. Results showed a monotonically increasing (and statistically significant) relationship between lineup bias (proportion of mock witnesses choosing the suspect) and rate of false positive identifications, as well as a monotonically increasing (and statistically significant) relationship between lineup size (E') and rate of false positive identifications.

5.1. Mock Witness Evaluation of Sequential Lineups.

The mock witness approach to lineup evaluation has generally been applied to the evaluation of simultaneously-presented lineups, which is the traditional form of lineup presentation. However, sequential presentation is also used, and under some circumstances may be preferable. McQuiston and Malpass (2002) investigated whether the mock witness procedure could be applied to the sequential presentation of lineup photographs, and how accurately mock witnesses could detect lineup bias when the lineup was presented sequentially. They also examined the degree to which instructional manipulations affect mock witness' choices, as there is some evidence that the instruction given to mock witnesses affects their choice rates (Wells & Bradfield, 1999). Last, they examined the source(s) of information mock witnesses use when making a lineup choice.

A sample of mock witnesses read information about a robbery, along with a description of the culprit. Two instructional manipulations were included: (1) participants were either told that the culprit may or may not be in the lineup (unbiased instructions) or told that the culprit was in the lineup (biased instructions); and (2) participants were either told that the lineup was from a real court case and that they should be absolutely certain in the lineup choice they make (high decision criterion instructions) or told that they were only participating in a research experiment and that the lineup choice they made had trivial consequences (low decision criterion instructions). They then viewed a lineup sequentially that was constructed a priori to be either fair or biased. Participants were led to believe they would see 20 photographs when in fact the sequence stopped after the sixth photograph, as "backloading" the sequence of photos is standard sequential lineup administration procedure (Lindsay et al., 1991). They were asked to indicate for each photograph shown whether they thought that was or was not the culprit based on the information given.

Thirty-five percent of the participants chose one lineup member as being the culprit, 49% chose two or more lineup members and 16% chose no one. For those who made a single lineup choice or multiple lineup selections (for the multiple choice analysis, a participant's lineup choice was considered to be accurate if the target was one of the selections in the combination), the target was correctly chosen from the biased lineup significantly more often than from the fair lineup, p's=.000. The variation in instructions had no meaningful effect on lineup choices. Last, the majority of participants indicated that the description of the culprit was the driving force behind the lineup choice(s) they made. Based on these results, McQuiston and Malpass (2002) concluded that these data support the use of the mock witness approach to evaluate the fairness of sequentially-presented lineups. They also offered several recommendations for the use of the sequential mock witness procedure, including: (1) mock witnesses should be instructed to make only one lineup choice, (2) they should be given a forced-choice instruction, and (3) they should be

instructed to choose the lineup member that best fits the description provided.

6. Unexplored Aspects Of The Mock Witness Procedure.

When the mock witness procedure is used to evaluate lineups in a research study, the descriptions provided to witnesses are generally contrived by experimenters or research participants and are sufficiently complete so that there is little or no guesswork on the part of the witness. But when this procedure is used as a post hoc measurement, issues like the completeness and accuracy of descriptions becomes an important aspect of the procedure. As research indicates, mock witnesses rely almost solely upon the information provided to them (culprit / suspect description) as their main source of identifying information in the mock witness task (McQuiston & Malpass, 2002). Because the procedure seems to be inherently dependent upon the completeness and/or quality of the culprit's description, it becomes problematic when this information is either lacking or unclear. But we do not know what source(s) of information mock witnesses turn to in the absence of good descriptive information when attempting to make a lineup choice.

The rationale behind the use of mock witness responses as the basis for advocating for or against the fairness of a given lineup has been criticized by those in the judicial system (i.e., attorneys, law enforcement officers), based on either a lack of understanding of the mock witness procedure and its rationale, and/or an unwillingness to accept this information as valid (Brigham & Pfeifer, 1994). As the influence of research on eyewitness identification continues to grow we can also expect the use of mock witness indices as measures of lineup fairness to continue to be advocated in court. How do fact finders interpret this information? The degree to which jurors and judges perceive, understand, use, or reject this information and incorporate it into their evaluation of the evidence in a case is unknown, but certainly an important avenue for future research.

7. Making The Measures Accessible.

There are many audiences for the concepts discussed in this chapter: researchers, law enforcement officers, judges, attorneys acting for the prosecution or the defense, and jurors, who are charged with making decisions based in part on their understanding of lineup fairness issues and their quantification. It would be useful to make these ideas and their quantification available and conceptually accessible to these constituencies.

Conceptual accessibility. The problem is to represent the quantifications of size and bias to the various constituencies in ways that their meaning is easily grasped. This is less a problem for the professional participants in the criminal justice system, and much more a problem for the novices: jurors. In our experience simple graphic presentations of choice rates for lineup members against a background of chance expectation is easily understood.

Access to the tools. The techniques and procedures needed to implement mock witness evaluation of a lineup can be assembled from information in and cited in this chapter. For practitioners in the criminal justice system, however constructing a complete procedure in this way would be time consuming and contain many uncertainties and disincentives. Detailed instructions do not exist in the law enforcement literature. Malpass (2004) provides detailed instructions and access to computer based calculations of the relevant statistics. As training in lineup construction and evaluation techniques become more common for law enforcement personnel access to these techniques should become more widely available.

8. References.

Brigham, J. C. & Brandt, C. C. (1992). Measuring lineup fairness: Mock witness responses versus direct evaluations of lineups. *Law and Human Behavior, 16 (5),* 475-489.

Brigham, J. C. & Pfeifer, J. E. (1994). Evaluating the fairness of lineups. *Adult Eyewitness Testimony: Current Trends & Developments* (pp. 201-222). Melbourne: Cambridge University Press.

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

Davies, G. M. & Valentine, T. (this volume). Facial composites: Forensic utility and psychological research. In R. Lindsay, D. F. Ross, J. D. Read, & M. P. Toglia (Eds). *Handbook of Eyewitness Psychology: Memory for People*. Mahwa, New Jersey: Lawrence Erlbaum and Associates.

Devlin, Hon Lord Patrick (1976). *Report to the Secretary of State for the Home Department of*

the Departmental Committee on Evidence of Identification in Criminal Cases. HMSO.

Doob, A. N. & Kirshenbaum, H. M. (1973). Bias in police lineups — partial remembering. *Journal of Police Science and Administration, 18*, 287-293.

Ellison, K. W. & Buckhout, R. (1981). *Psychology and Criminal Justice*. New York: Harper & Row.

Laughery, K. R., Jensen, D. G., & Wogalter, M. S. (1988). Response bias with prototypic faces. In M. M. Gruneberg, P. E. Morris, & R. N. Sykes (Eds.), *Practical Aspects of Memory: Current Research and Issues* (pp. 157-162). Chichester: Wiley.

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

Lindsay, R. C., Lea, J. A., & Fulford, J. A. (1991). Sequential lineup presentation: Technique matters. *Journal of Applied Psychology, 76(5),* 741-745.

Luus, C. A. E. & Wells, G. L. (1991). Eyewitness identification and the selection of distractors for lineups. *Law and Human Behavior, 15*, 43-57

Malpass, R.S. (2004). A Lineup Evaluation "Do-It-Yourself Kit" for Attorneys and Law Enforcement. http://eyewitness.utep.edu/diy.html

Malpass, R. S. & Devine, P. G. (1983). Measuring the fairness of eyewitness identification in lineups. *Evaluating Eyewitness Evidence* (pp. 81-102). London: John Wiley & Sons Ltd.

Malpass, R. S. (1981). Effective Size and defendant bias in eyewitness identification lineups. *Law and Human Behavior, 5(4),* 299-309.

Marwitz, D. B. & Wogalter, M. S. (1998). *Bias in photo-spreads of faces: A comparison of two lineup construction methods.* Paper presented at the Proceeding of the Human Factors Society, Santa Monica.

McQuiston, D. E. & Malpass, R. S. (2002). Validity of the mock witness paradigm: Testing the assumptions. *Law and Human Behavior, 26(4),* 439-453.

Meissner, C. A., Sporer, S. L., & Schooler, J. (in press). Person descriptions as eyewitness evidence. In Lindsay, R. C. L., Ross, D. F., Read, J. D., & Toglia, M. P. (Eds.), *Handbook of Eyewitness Psychology, Vol. 2 (Chapter 9)*. Mahwah: Lawrence Erlbaum.

Pelwani v. S. 1963 (2) (PH) H237 (T)

Scheck, B., Neufeld, P., & Dwyer, J. (2001). *Actual Innocence: When Justice Goes Wrong And How To Make It Right.* New York: Doubleday.

Shepherd, J. W. & Ellis, H. (1996). Face recall – methods and problems. In S. L. Sporer, R. S. Malpass & G. Köhnken (Eds.), *Psychological Issues in Eyewitness Identification* (pp. 87-115). Mahwah: Lawrence Erlbaum Associates, Inc.

Smith, S. M., Lindsay, R. C. L., & Pryke, S. (2000). Postdictors of eyewitness errors: Can false identifications be diagnosed? *Journal of Applied Psychology, 85(4)*, 542-550.

Steblay, N., Dysart, J., Fulero, S., & Lindsay, R.C.L. (2001). Eyewitness accuracy rates in sequential and simultaneous lineup presentations: A meta-analytic comparison. *Law and Human Behavior, 25(5),* 459-474.

Technical Working Group on Eyewitness Evidence (1999). Eyewitness Evidence: A Guide for Law Enforcement. Washington, D.C.: National Institute of Justice (i-x, 1-44). http://www.ojp.usdoj.gov/nij/pubs-sum/178240.htm

Topp, L.D., McQuiston, D.E. & Malpass, R.S. (2003, July). Exploring Composite Production and Its Subsequent Effects on Eyewitness Memory. Poster presented at the Meetings of the Society for Applied Research in Memory and Cognition, Aberdeen, Scotland.

Tredoux, C. G. (1998). Statistical inference on measures of lineup fairness. *Law and Human Behavior, 22(2)*, 217-237.

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

Tredoux, C. G. (2002). A direct measure of facial similarity and its relation to human similarity perceptions. *Journal of Experimental Psychology: Applied, 8(3)*, 180-193.

Tredoux, C.G., Parker, J.F. & Nunez, D. (2004). *Optimality of encoding and process/outcome parallelism in eyewitness and mock witness identifications*. Unpublished manuscript, University of Cape Town.

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., Leippe, M. R., & Ostrom, T. M.. (1979). Guidelines for empirically assessing the fairness of a lineup. *Law and Human Behavior, 3 (4),* 285-293.

Wells, G. L., Rydell, S. M., & Seelau, E. P. (1993). On the selection of distracters for eyewitness lineups. *Journal of Applied Psychology, 78*, 835-844.

Wells, G. L., Seelau, E. P., Rydell, S. M., & Luus, C. A. E. (1994). Recommendations for properly conducted lineup identification tasks. In D. F. Ross, J. D. Read, & M. Toglia (Eds.), *Adult Eyewitness Testimony: Current Trends and Developments* (pp. 223-244). Cambridge University Press: NY.

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

Wogalter, M. S., Malpass, R. S., & McQuiston, D. E. (2004). A national survey of U.S. police on preparation and conduct of identification lineups. *Psychology, Crime, & Law, 10(1),* 69-82.

Wogalter, M. S., Marwitz, D. B., & Leonard, D. C. (1992). Suggestiveness in photospread line-ups: Similarity induces distinctiveness. *Applied Cognitive Psychology, 6(5)*, 443-453.

Wogalter, M.S., & Jensen, D. G. (1986). Response bias in lineups with prototypic faces. In *Proceedings of the Human Factors Society 30th Annual Meeting* (pp. 725-728), Santa Monica, CA: Human Factors Society.

Table 1. Functional Size in a number of hypothetical lineups

Lineup	Member							
	1	**2**	**3**	**4***	**5**	**6**	**Not Present**	**Functional Size**
d	5	3	6	30	9	3	4	2
e	8	9	8	10	9	9	7	6
f	1	3	20	10	21	4	1	6
g	20	50	4	1	20	5	0	100

* = suspect.

Table 2. Effective Size in a number of hypothetical lineups.

Lineup	Member						
	1	**2**	**3**	**4***	**5**	**6**	**E_a**
h	0	25	5	25	3	2	2.83
i	10	10	9	10	11	10	5.90
J	1	0	12	12	0	11	3.17
K	7	7	7	24	8	7	4.60
L	12	6	9	13	14	6	5.10
m	6	19	3	20	8	10	4.45

E_a = Effective Size calculated with adjustment for null fillers.

* = suspect

Table 3. Correlations between E' and Effective Size in a series of simulated lineups

	N = 20	N= 30	N = 50	N= 100	N = 1000
k = 6	0.97	0.987	0.994	0.997	0.998
k =9	0.986	0.996	0.998	0.999	0.999
k = 12	0.982	0.991	0.996	0.999	0.999

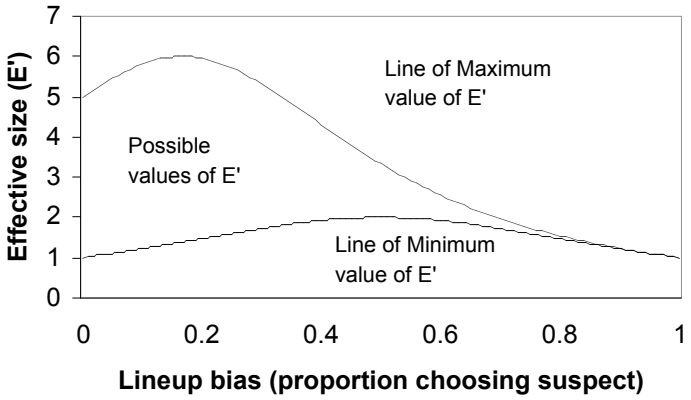
All correlations are significant at $p < 0.001$

Table 4. Effective Size in a hypothetical lineup

Lineup Member							
1	**2**	**3**	**4***	**5**	**6**	**E_a**	**E**
1	25	2	25	2	25	3.38	3.4

E_a = Effective Size (Malpass, 1981) E' = Effective Size (Tredoux, 1998)
* = suspect

Figure 1. The relationship between lineup bias and lineup size



Values calculated for a six person lineup

Appendix 1. “Composing Lineups”, excerpted from Eyewitness Evidence: A Guide for Law Enforcement (Technical Working Group, 1999)

A. Composing Lineups

Principle: Fair composition of a lineup enables the witness to provide a more accurate identification or nonidentification.

Policy: The investigator shall compose the lineup in such a manner that the suspect does not unduly stand out.

Procedure: ***Photo Lineup:*** In composing a photo lineup, the investigator should:

1. Include only one suspect in each identification procedure.
2. Select fillers who generally fit the witness’ description of the perpetrator. When there is a limited/inadequate description of the perpetrator provided by the witness, or when the description of the perpetrator differs significantly from the appearance of the suspect, fillers should resemble the suspect in significant features.
3. If multiple photos of the suspect are reasonably available to the investigator, select a photo that resembles the suspect description or appearance at the time of the incident.
4. Include a *minimum* of five fillers (nonsuspects) per identification procedure.
5. Consider that complete uniformity of features is not required. Avoid using fillers who so closely resemble the suspect that a person familiar with the suspect might find it difficult to distinguish the suspect from the fillers.
6. Create a consistent appearance between the suspect and fillers with respect to any unique or unusual feature (e.g., scars, tattoos) used to describe the perpetrator by artificially adding or concealing that feature.
7. Consider placing suspects in different positions in each lineup, both across cases and with multiple witnesses in the same case. Position the suspect randomly in the lineup.
8. When showing a new suspect, avoid reusing fillers in lineups shown to the same witness.
9. Ensure that no writings or information concerning previous arrest(s) will be visible to the witness.
10. View the spread, once completed, to ensure that the suspect does not unduly stand out.
11. Preserve the presentation order of the photo lineup. In addition, the photos themselves should be preserved in their original condition.