Bias in Police Lineups and its Reduction by an Alternative Construction Procedure

Michael S. Wogalter, Laura J. Van't Slot, and Michael J. Kalsher
Department of Psychology
Rensselaer Polytechnic Institute
Troy, NY 12180

ABSTRACT

The present research examined whether eyewitness identification lineups produced by law enforcement personnel are biased or suggestive. Experienced police officers were asked to construct two six-face photographic lineups, first using their usual (traditional) method, and second using an alternative method. The primary basis of the traditional method is that foils are selected based on their similarity to the target. The alternative method includes foils that are not only similar to the target but also similar to other foil faces in the lineup. Both types of lineups were shown to subjects who had not seen the faces before (mock witnesses) and were asked to guess the respective targets. The results showed that mock witnesses selected the targets significantly more often than expected by chance (1/6 probability) when embedded in the traditional lineups, thus demonstrating that these lineups were suggestive. Mock witnesses did not select alternative-method targets more often than expected by chance. These results indicate that foil selection procedures incorporating foil-to-foil similarity produce fairer lineups than those exclusively based on target similarity. Implications for forensic lineup construction procedures and for future research are discussed.

INTRODUCTION

Eyewitness identification is an important part of many criminal investigations. In some criminal cases, such as rape and assault, eyewitness identification may be the only evidence available (Malpass & Devine, 1984). One technique used in identification procedures involves live lineups (or the picture version, photo spreads) which contain several persons known to be innocent (foils) plus the suspected offender. The purpose of lineups is to allow the witness to identify the person he/she saw under conditions that avoid the false identification of an innocent suspect.

Protection of the innocent suspect is usually afforded by a fair lineup. A fair lineup presents the suspect in a manner that is not conspicuous relative to the other members of the lineup. Unfair or suggestive lineups provide cues that inform the witness of the identity of the police suspect. Wall (1965) ascribes the influence of suggestion as accounting for more miscarriages of justice than any other factor involved in eyewitness identification cases. If suggestion was not considered a problem then a showup, where the suspect is presented alone to witnesses, would be adequate. Theoretically, lineups are more fair than showups because the probability of choosing an innocent suspect is distributed across several faces of a lineup.

The need for fair lineups is particularly important in situations where witnesses feel obligated to make an identification for reasons that are extraneous to face recognition. For example, when confronted with authority figures, witnesses may feel pressured to comply with police demands (Doob & Kirshenbaum, 1973). Doob and Kirshenbaum (1973) suggest that the witness may take a role analogous to the "good" subjects in psychological experiments (Orne, 1962). "Good" subjects seek cues from the experimental situation to provide direction on how to behave properly, and in particular, how to behave in ways they think will help to support the experimenter's hypothesis. Similarly, witnesses may surmise that the police believe the suspect is guilty and will be disappointed or annoyed if an identification is not made. Thus, the witness may feel pressured to make a selection because a "correct" identification would help corroborate police suspicions. If the offender is absent (i.e., the police suspect is not the actual offender), identification might produce wrongful incrimination and possibly lead to false conviction. Fair lineup identification procedures reduce the likelihood that innocent suspects will be selected by witnesses who are inclined to make a choice.

Lineup fairness cannot be assessed using actual eyewitnesses because of the inability to separate how much of their performance was due to recollection of the offender and how much of it was due to lineup suggestiveness. However, suggestiveness can be assessed by measuring the base rate selections of non-eyewitnesses (mock witnesses). If the suspect is selected by mock witnesses more often than other members of the lineup, then the lineup's physical construction is biased or suggestive with respect to the suspect. More formally, a lineup is biased when mock witnesses select the suspect more often (or less often) than expected by chance. Chance is defined as 1/n, where n is the number of people in the lineup (Doob & Kirshenbaum, 1973). For example, in a fair six-person lineup the proportion of mock witnesses choosing the suspect should equal 0.167 (1/6). Fairness decreases as actual selection departs substantially from this value.

To avoid bias, the traditional construction procedure includes foils who appear similar to the suspect. Malpass and Devine (1983) reported that the criterion of similarity appeared in all of the lineup construction guidelines that they reviewed. Usually, these guidelines recommend that the foils be about the same age, height, build, race, have the same hair length, hair color, and be similar in general demeanor and position in life as the suspect. This guideline has not always been followed, however. Buckhout (1977) described a police lineup in which a black suspect was placed in a lineup with five white foils. Several cases of suggestive lineups have been cited by the United States Supreme Court. In one notorious case, United States v. Wade (1967) the suspect was known to be a young man but the lineup consisted of several men over 40 years of age and one teenager (the suspect). In another case, a male Oriental suspect was placed in a lineup in which he was the only
person of Asian descent. These cases illustrate that the lineup is regarded as unfair if the suspect is not similar to the other lineup members. In these examples, the suspects were so distinctive that witnesses were essentially presented with lineups containing only one real choice, not unlike showups.

The logic of using foils similar to the suspect is that eyewitnesses, who possess specific information about the offender’s individual and unique appearance, should be able to distinguish between the offender (if present) and other lineup members who possess the same general characteristics. Without knowledge of the offender’s unique appearance, non-eyewitnesses should only be able to select the suspect with a probability equal to chance.

However, recent research casts a shadow on the similarity criterion (Laughery, Jensen, & Wogalter, 1988; Marwitz & Wogalter, 1988; Wogalter & Jensen, 1986; Wogalter, Marwitz, & Leonard, 1991). For example, Marwitz and Wogalter (1988) found that mock witnesses selected the target more often than chance would predict in lineups formed by students based on target similarity. Other research using facial stimuli taken from composite kits (e.g., the Identikit) has shown similar results (Laughery et al., 1988; Wogalter & Jensen, 1986). Thus, this research demonstrates that lineups based solely on foil-to-suspect similarity make the suspect stand out because it is the most similar face in the lineup.

The question remains whether the lineups constructed by law enforcement officials show the same similarity bias found in lineups constructed by college students (Marwitz & Wogalter, 1988; Wogalter et al., 1991). One purpose of this research was to determine whether seasoned police officers produce suggestive lineups using traditional techniques.

This research explored an additional issue. Would lineups produced in a different manner be less biased? In earlier research, Wogalter et al. (1991) evaluated three alternative construction methods. In each method, foil selection was based not only on target similarity but also on similarity with one or more of the other lineup faces. The results showed that alternative-lineup targets were not selected significantly more often than by chance, suggesting that bias was reduced. In addition, an overall analysis showed that the alternative lineups were significantly less suggestive than target-based lineups. However, Wogalter et al. (1991) used college students to construct all alternative-method lineups. It is unknown whether law enforcement officers would also produce less biased lineups using an alternative construction method. In the current study, police officers were told to construct a second lineup in which foils were selected based not only on their similarity to the target, but also on their similarity to other lineup foils. It was expected that lineups constructed in this manner would make the target less "prototypical" by distributing similarity among the lineup members.

**METHOD**

The experiment involved three phases: stimulus preparation, construction, and presentation. In the stimulus preparation phase, sets of targets and potential foil faces were assembled. In the construction phase, police officers assembled two lineups. In the presentation phase, the lineups were given to mock witnesses who attempted to guess the targets in the lineups.

**Subjects**

Sixteen police officers from cities and towns around the metropolitan Albany, NY area constructed the lineups. All were experienced in lineup construction, having constructed an average of 138 lineups ($s = 24.0$), of which 87% were photographic (as opposed to live) lineups, for an average of 13.6 years ($s = 7.4$). Later, 56 Rensselaer Polytechnic Institute (RPI) students participated as mock witnesses.

**Stimulus materials**

In the stimulus preparation phase, approximately 1000 photographs of male college seniors from the 1984 to 1986 RPI yearbooks were cut into individual black and white pictures. All were front-portrait views of persons donning similar clothing (dark coat and tie). Sixteen faces were randomly selected to serve as targets. The remaining faces were sorted into 16 sets of 25 based on general similarity to the targets and these sets served as the foil pools used for the lineups. The other faces of the original pool were eliminated. All experimental pictures were duplicated to control for stimuli used between construction methods.

**Procedure**

In the construction phase, police officers each were given a picture of a target, one foil pool, and a generic witness description. They were told to use their usual method to assemble a six-face lineup. The verbal description given before all lineups were constructed was:

The suspect was a white male. He was relatively well groomed and he wore a suit or sports coat. He appeared to be of college age, probably in his late teens or early twenties. No distinguishing marks or characteristics were reported by the witness. The suspect was apprehended near the scene of the crime. He had no alibi and could not account for the time that the alleged incident occurred.

After completing the first lineup, each officer was informed about the problems with lineups constructed solely around the suspect. Specifically, they were told that research has shown that this construction procedure makes the suspect the most similar face in the lineup, and as a consequence, there is increased likelihood that people can guess the police suspect without any prior knowledge of the perpetrator’s appearance or criminal incident. With this background, the officers were told to construct lineups in which all members were equally similar to each other. No explicit direction was given on how they should select the foils except they were told that at any point they could exchange or replace any of their earlier choices. They were encouraged to try out different collections of faces using any approach that they deemed appropriate to meet the goal of equal similarity. The only constraint was that the target had to remain in their final lineup.

Police officers were given a questionnaire that assessed the methods they used to construct lineups in the course of their work. Included in the questionnaire were items that requested information on how they routinely construct lineups, where and how they learned lineup construction procedure, the facial features they paid the most attention to, and whether and why their lineups had been challenged in court. Most questionnaire items were given after the first
lineup and before the instructions for the second-lineup. The purpose of this procedure was: (1) to avoid any possible influence of the questionnaire on the construction of the traditional lineup, and (2) to avoid any possible influence of the alternative-lineup instructions on the questionnaire responses. The questionnaire responses will not be described in detail in this report, except for the two following items. The officers were asked after each lineup was constructed to "Rate your professional opinion of the quality of the lineup" using a 9-point Likert-type scale with the even numbered anchors having the following verbal labels: (0) extremely poor, (2) poor, (4) moderate, (6) good, and (8) excellent. They were also asked, "Considering all of the foil faces together, how similar are they to the suspect/target face?" and they responded on an 8-point Likert-type scale with the anchors: (0) not at all similar, (1) remotely similar, (2) somewhat similar, (3) moderately similar, (4) similar, (5) very similar, (6) extremely similar, and (7) identical.

After the lineup construction phase was completed, pictures from each lineup were affixed to the inside of manila folders. The folders were assigned to one of two sets, each containing 16 lineups. Half of the lineups in each set comprised traditional lineups and half comprised alternative lineups. Sets were matched with respect to targets in the traditional and alternative lineups, such that a target that appeared in a traditional lineup in one set appeared in an alternative lineup in the other set.

In the presentation phase, mock witnesses received a randomly ordered stack of lineups from one of the two folder sets. They were told that the lineups were assembled by police officers and that a "police suspect" was present in each lineup. Mock witnesses were told to study each lineup carefully and to choose the face that they thought was the suspect. They were told to ignore facial expression and perceived "guilty" appearances. Mock witnesses marked their choice and the viewing order of each lineup on a response sheet. They were not given feedback as to the correctness of their choices.

RESULTS

Correct responses (selections matching the target) were given a score of 1, and incorrect responses were given a score of 0. Target selection was compared to what would be expected by random/chance selection. If the subjects were merely selecting faces at random, the rate of target selection would be one out of six or a mean of .167. Selection rates above this level would indicate the lineups are biased in the direction of the suspect/target. Targets in the traditional lineups ($M = 2.50$, $s = .158$) were chosen significantly more often than expected by chance, $t(55) = 3.94$, $p < .001$. However, targets in the alternative lineups ($M = .179$, $s = .119$) were not chosen significantly more often than expected by chance, $t(55) = 0.75$, $p > .05$. Direct comparison between methods showed that selection of targets in traditional lineups was more likely than for the alternative lineups, $t(55) = 2.66$, $p < .01$.

A more conservative test was also used. For every lineup, the mean selection rate was compiled by collapsing across mock witnesses resulting in 16 pairs of scores that were matched to police officer. Analysis showed that targets in the traditional lineups were chosen more often than expected by chance, $t(16) = 2.56$, $p < .05$. In addition, targets in the alternative lineups were not chosen significantly more often than expected by chance, $t(15) = 0.43$, $p > .05$. However, unlike the more powerful analysis described earlier, the difference between the two lineup methods was marginal, but not significant at the conventional probability level, $t(15) = 1.98$, $p < .07$.

Mock witnesses were not explicitly informed as to how they should go about selecting the targets from the lineups. It is possible that the bias effects seen in the traditional lineups was due to cues picked up as they went through the lineup sequence. That is, after viewing several lineups they might have "learned" that foil-to-target similarity points to the targets. Also, actual eyewitnesses usually do not see more than one or two lineups. Analyses were performed to determine whether the similarity bias effect tends to occur for later viewed lineups compared to earlier lineups. While lineup order was randomized for each mock witness, it was tracked and used to sort lineups into each individual’s viewing order. Interestingly, the first lineup viewed was the most biased of the 16 ($M = .41$ and .24, for the traditional and alternative lineups, respectively). Comparison to chance showed that participants receiving a traditional lineup first were significantly more likely to select the target than expected by chance, $t(26) = 2.50$, $p < .05$. However, participants receiving an alternative lineup first did not select the target significantly greater than chance, $t(26) = 0.92$, $p > .05$. However, no significant difference between methods was found for the first viewed lineup, $t(54) = 1.33$, $p > .05$. Further analysis showed no effect of lineup order in a one-way repeated-measures analysis of variance, $F(1, 825) = 1.10$, $MSe = 1.69$, $p > .05$. In addition, no relationship was found between lineup order and performance, $r(14) = .14$, $p > .05$. There was also no significant difference between the first eight vs. the second eight lineups, $r(55) = 0.34$, $p > .05$, and between the first four lineups vs. the last four lineups, $r(55) = 0.78$, $p > .05$.

After constructing each lineup, officers were asked to rate their professional opinion of the lineup's quality. The rating for the traditional lineup was significantly higher ($M = 6.50$, $s = 89$) than the rating for the alternative lineup ($M = 5.62$, $s = 1.20$), $t(15) = 2.91$, $p < .05$. The officers were also asked to rate the overall similarity of the foils to the suspect/target. The similarity of the foils in the traditional lineup was somewhat higher ($M = 4.38$, $s = .72$) than in the alternative lineup ($M = 4.19$, $s = .98$), but the difference was not significant, $t(15) = 0.90$, $p > .05$.

DISCUSSION

Previous research suggests that lineups based exclusively on target similarity are biased. However, these earlier studies used college students to assemble the lineups (Marwitz & Wogalter, 1988; Wogalter et al., 1991) or used artificial faces (Wogalter & Jensen, 1986; Laughery et al., 1988). The present research used face photographs and experienced police officers to construct lineups, thus producing the most realistic assessment of the similarity bias effect attempted to date. The current research supported earlier work. Police officers' (traditional) lineups were biased in the direction of increased target selection. Apparently, these lineups provide cues that enable mock witnesses to select the targets to a greater extent than would be expected in truly fair lineups.

Bias was also investigated using an alternative construc-
tion procedure in which foil selection was based in part on the facial characteristics of the foils. The results showed that target selection in the alternative lineups was not different from chance ( unlike the traditional lineups). An analysis comparing lineup methods confirmed the difference between the target-based and alternative-method lineups. These results suggest that distributing similarity among all lineup members produces fairer lineups than the traditional method.

In a more conservative analysis, the direct comparison between lineup methods failed to reach significance. However, this analysis involved only 16 pairs of scores and thus had lower statistical power compared to the analysis based on the 56 mock witness scores. Had there been more officers in the study, power would have been greater.

Interestingly, the officers' quality ratings showed a significant difference in favor of the traditional lineup compared to the alternative lineup. On the one hand, the officers rated their traditional lineup's quality higher than the alternative lineups. On the other hand, the assessment by the mock witnesses showed that the alternative method was less biased. However, the questionnaire result should be considered tentative, because the information available at the time of the two ratings was different. For example, when the police officers rated the alternative lineup they already had constructed the traditional lineup and had been given admonishments about lineup bias. This information might have raised doubts about the quality of their lineups, and thus lowered the ratings on the second lineup.

Five other methodological aspects of concerns should be mentioned. First, lineup construction order was not counter-balanced. In most repeated or within-subject designs, conditions are counterbalanced so that effects can be attributed to differences in conditions rather than possible order effects. Counterbalancing was not used because the primary interest of the current experiment concerned whether using their routine method, police officers construct biased lineups using their routine method. We were careful not to provide any information before this task so as not influence the police officers in any way. Following the completion of the first lineup, responses to the questionnaire confirmed that 81% of the officers exclusively based their foil selections on target similarity. It would have made little sense to have the police officers first construct the alternative lineups and then later construct the traditional lineup. With this order, the officers might have been influenced by the alternative-method lineup instructions when constructing the traditional lineup.

Second, it might seem possible that the lowered bias for the alternative-method lineups was due to practice gained by constructing the first lineup. However, as described earlier, all officers in the current study had extensive experience constructing traditional photographic lineups. Thus, it is unlikely that the bias reduction found for the second lineup was due to a practice effect as the police officers had considerable experience in the task before taking part in the study.

Third, mock witnesses were forced to make a choice for every lineup. Real witnesses are normally given the option of not choosing in fair lineup procedures. By forcing mock witnesses to make a choice, the pattern of selections should reflect those of liberal, compliant, or cooperative witnesses ( Doob & Kirshenbaum, 1973). As a consequence, the applicability of the current research may be limited to individuals who would lean toward making a choice or situations that encourage choosing.

Fourth, mock witnesses viewed 16 lineups, rather than a single lineup as many witnesses would. Because order of lineups viewed was tracked, it allowed analysis of whether the strategy of mock witnesses changed after viewing several lineups. However, the analyses indicated that there were no significant effects of order, other than the finding that mock witnesses who viewed a traditional lineup first, selected the target significantly more often than expected by chance.

Fifth, there is the possible concern that some foils might have been inadequate or distinctive. If so, mock witnesses could immediately eliminate them from consideration, and thus, functionally decrease lineup size. However, foil quality was probably not a problem for the following three reasons: (1) The faces in lineups were very similar in appearance as they were first included into foil pools based on similarity, and then in the construction phase, additional similarity decisions were made by the police officers. (2) Similar results were found by Laughery et al. (1988) and Wogalter and Jensen (1987) using foils that differed by only a single feature from the target (i.e., extremely similar foils). Moreover, recent research by Nosworthy and Lindsay (1990) showed that lineup size (having at least three "good" foils) had little effect on identification decisions. (3) Informal examination of the raw data showed that very few faces received no selections.

Recently, Wells and Luus (1990; Luus & Wells, in press) suggest that lineups should not be constructed around the appearance of the suspect, rather they should be based on the prelineup verbal description provided by the eyewitness. This may be a way to decrease the similarity bias described in this article but exclusive use of verbal description to construct lineups presents its own problems. People are not fluent in describing faces and the resulting descriptions are poor (Ellis, Shepherd, & Davies, 1980; Laughery, Duval, & Wogalter, 1986; Shepherd, Davies, & Ellis, 1978). Part of the problem arises from the witness's difficulty translating the memorial image of the suspect's face to language. Another source of error associated with verbal description is that police officers have to interpret the witness' description (i.e., translate it back to visual image to search for a suspect and later to form a lineup). Furthermore, the witness is often not thinking clearly shortly after a crime incident and may inadvertently omit crucial information from their verbal description (Navon, 1990). Using prelineup verbal description as the exclusive basis for lineup construction would allow the inclusion of dissimilar appearing distractors. Consider the quite general, but not at all unusual, verbal description given to the officers in the current study. It not only described virtually all faces in our sample, but also describes millions of other people. The exclusive use of verbal description as the basis of lineups opens the door to the use of foils which may have only a remote resemblance to the suspect, but who might still "fit" the verbal description. Moreover, support in court is questionable because the defense side could surely complain that the range of foils allowed by most descriptions would allow the suspect to stand out unfairly. It is our belief that a lineup based entirely on verbal descriptions is not adequate alone, but as future research may show, it may be part of an approach that, along with suspect and foil similarity considerations, yields fair lineups.
Some suggestiveness is probably inevitable in real-world identification situations. The goal is to minimize it. With the advent of computerized mug files coded on facial characteristics (Shepherd, 1986), police can construct lineups with high target-to-foil similarity. How similar should they make the lineup? The present results suggest that the most appropriate level of similarity is lower than maximal. Police investigators should be aware of this problem and take this into consideration when constructing lineups. The only way to know whether a lineup is suggestive is to test it using mock witnesses. A number of iterations of replacing and substituting foils might be necessary before an acceptable, fair lineup is found.

Finally, the present research calls attention to an interesting problem. An often-stated guideline for fair lineups is that foils be selected on the basis of similarity to the target (Malpass & Devine, 1983). However, as we have seen, the similarity rule taken to its limit produces suggestive lineups. As suggested above, the similarity-fairness function probably has an inverted-U shape. Very low similarity and very high similarity lineups are less fair than lineups somewhere between the two extremes. The question is how much and what kinds of similarity are needed to attain this, and what methods are most appropriate.

Research is beginning to investigate lineup construction methods and techniques that lead to fair lineups. Cost (in terms of money, time, and effort needed for implementation) should be addressed since it will no doubt determine whether the recommendations gained from research are accepted (e.g., Brigham, Ready, & Spier, 1990). These expenses must be carefully weighed against the possible costs to an innocent suspect who might be wrongly convicted with eyewitness identification evidence. The present research suggests one alternative procedure that may be useful, but future investigations will help to define the specific methods and procedures that foster the fairest lineups.

REFERENCES


