Law and Human Behavior, Vol. 26, No. 4, August 2002 (© 2002)

Validity of the Mockwitness Paradigm: Testing the Assumptions[1]

Dawn E. McQuiston[2] and Roy S. Malpass[3,4]

Mockwitness identifications are used to provide a quantitative measure of lineup fairness. Some theoretical and practical assumptions of this paradigm have not been studied in terms of mockwitnesses' decision processes and procedural variation (e.g., instructions, lineup presentation method), and the current experiment was conducted to empirically evaluate these assumptions. Four hundred and eighty mockwitnesses were given physical information about a culprit, received 1 of 4 variations of lineup instructions, and were asked to identify the culprit from either a fair or unfair sequential lineup containing 1 of 2 targets. Lineup bias estimates varied as a result of lineup fairness and the target presented. Mockwitnesses generally reported that the target's physical description was their main source of identifying information. Our findings support the use of mockwitness identifications as a useful technique for sequential lineup evaluation, but only for mockwitnesses who selected only 1 lineup member. Recommendations for the use of this evaluation procedure are discussed.

KEY WORDS: mockwitness; eyewitness identification; sequential lineups.

INTRODUCTION

Procedural standards for the collection of eyewitness identification evidence, particularly the construction of lineups, have been the focus of many court decisions (Lipton, 1996; Sobel, 2001), a substantial literature in experimental psychology, and a number of documents developed by law enforcement agencies (Wells et al., 1998, 2000). Wells et al. (1998) review many instances of good practice recommendations and offer four of their own in the first "white paper" endorsed by the American Psychology – Law Society. More recently, the National Institute of Justice convened

[1]Portions of these data were presented at the 1999 Biennial Meeting of the Society for Applied Research in Memory and Cognition and the 1999 Meeting of the European–American Psychology-Law/American Psychology-Law Society. This research is based on the first author's Master's thesis.

[2]Department of Psychology, University of Texas at El Paso, El Paso, Texas.
[3]Department of Psychology and Criminal Justice Program, University of Texas at El Paso, El Paso, Texas.
[4]To whom correspondence should be addressed at Department of Psychology, University of Texas at El Paso, El Paso, Texas 79968; e-mail: rmalpass@utep.edu.

a Technical Working Group consisting of law enforcement officers, prosecuting and defense attorneys, and psychological scientists to write a guide for law enforcement regarding the collection of eyewitness identification evidence (Technical Working Group for Eyewitness Evidence, 1999). The proper construction of eyewitness lineups is an important focus in both efforts to compose guidelines. Wells et al. (1998) clearly stated the central proposition of lineup construction as their Rule #3:

> The suspect should not stand out in the lineup or photospread as being different from the distracters based on the eyewitness' previous description of the culprit or based on other factors that would draw extra attention to the suspect.

Mistaken identifications of innocent people are more likely when the suspect stands out in a lineup (i.e., a structurally biased lineup; Wells et al., 1998). The literature on eyewitness identification has proposed several quantitative indices to measure structural bias based on the use of *mockwitnesses* (see Malpass & Lindsay, 1999, for a review): individuals who have not witnessed a crime but are asked to select the individual they think perpetrated the crime from a lineup based on only his/her physical description given by the eyewitness. The rationale is that if a lineup is constructed properly, each member of it should have an equal probability of being selected by individuals who did not witness the crime (Doob & Kirshenbaum, 1973; Malpass, 1981; Wells, Leippe, & Ostrom, 1979). According to this paradigm, a lineup is considered to be biased to the extent that (1) the proportion of suspect identifications by mockwitnesses significantly departs from chance expectancy (*lineup bias*), and (2) the fillers in the lineup are not viable alternatives to the suspect (*lineup size*).[5] If the frequency with which mockwitnesses select the suspect based on only the description information given departs significantly from chance expectancy (where chance is $1/n$ and n is the number of lineup members), it becomes questionable whether an eyewitness's identification of the police suspect is due to a true match between the eyewitness's memory trace of the culprit and that lineup member, or is a result of a poorly constructed lineup in which the police suspect stands out in some way relative to the other members (Doob & Kirshenbaum, 1973).

Assumptions of the Mockwitness Paradigm: Empirically Sound?

Since its introduction in 1973, the mockwitness approach to estimating lineup bias has become generally accepted by researchers in the field of eyewitness memory, particularly as a means of evaluating the reliability of lineup identifications offered as evidence in court (Brigham & Brandt, 1992; Brigham, Ready, & Spier, 1990; Buckhout, Rabinowitz, Alfonso, Kenellis, & Anderson, 1988; Corey, Malpass, &

[5]There are two conceptually different measures of lineup fairness: lineup *bias* and lineup *size* (Malpass & Lindsay, 1999). The concept *lineup bias* refers to the degree to which the suspect is distinctive in appearance relative to the other lineup members. Statistical measures of lineup bias such as Functional Size take into consideration the number of mockwitnesses who evaluate the lineup and the proportion of those who identify the suspect (Wells, Leippe, & Ostrom, 1979). *Lineup size*, on the other hand, reflects the number of viable choice alternatives to the suspect in the lineup given the eyewitness's verbal description of the culprit. Lineups should be sufficiently large to distribute the probability of an innocent suspect being chosen by chance alone (16.67% in a six-member lineup). Statistical measures of lineup size such as Effective Size and Tredoux' E take into account the total number of mockwitnesses evaluating the lineup and the frequency with which each lineup member is chosen (Malpass, 1981; Tredoux, 1998).

McQuiston, 1999; Doob & Kirshenbaum, 1973; Malpass, 1981; Malpass & Devine, 1983; Wells, Leippe, & Ostrom, 1979). Although the logic underlying the mockwitness paradigm is sensible, its validity as a lineup evaluation procedure is based on a number of theoretical and practical assumptions that seem not yet to have been subject to empirical examination. We outline three questions that have received little attention in the literature.

Is the Mockwitness Method Robust to Variations in Lineup Procedure?

Mockwitness choices have traditionally been used to estimate structural bias in simultaneously presented eyewitness lineups in both experimental studies and field work. There is a literature, however, that supports and recommends the use of sequential lineups in the field—a procedure whereby lineup members are presented one at a time and the witness is asked whether each individual is the offender before the next is shown (Lindsay & Wells, 1985). Some studies show a sequential–superiority effect in that sequential presentation results in fewer false identifications and more correct rejections (Cutler & Penrod, 1988; Lindsay et al., 1991; Lindsay & Bellinger, 1999; Lindsay, Lea, & Fulford, 1991; Lindsay & Wells, 1985; Sporer, 1993), and reduces various forms of lineup bias (i.e., structural, clothing, instructional; Lindsay et al., 1991) in comparison with simultaneous presentation. Based on these results, Levi (1999) suggests an "honorable discharge" for the mockwitness approach to lineup evaluation because sequential lineups sufficiently control for lineup bias.

Results of a recent meta-analysis examining the sequential–superiority effect provide important additional data from this research base (Steblay, Dysart, Fulero, & Lindsay, 2001). For overall correct decisions (whether the target is present or absent), simultaneous lineups were found to be superior to sequential lineups. Considering target present and target absent data separately, simultaneous lineups were found to be superior to sequential lineups when the *target is present* in that eyewitnesses made significantly more correct identifications and fewer false rejections with simultaneous lineups. The data show that only when the *target is absent* were sequential lineups superior to simultaneous lineups for obtaining fewer false identifications of an innocent person and more correct rejections of the lineup. The meta-analysis also shows that publication status significantly moderated the sequential–superiority effect in that unpublished studies are less likely to support this effect. In addition, a large majority of published and unpublished studies reporting the sequential–superiority effect come from a small number of research laboratories, and have disproportionately used high-similarity-to-culprit suspects in target absent lineups, raising questions about ecological validity (Maclin, McQuiston, Kroeger, & Malpass, 2001; McQuiston, Malpass, & MacLin, 2002).

In view of these data on sequential lineups and the questions raised, we argue that it would be premature to retire the mockwitness approach and efforts should be made to devise a mockwitness procedure suitable to evaluate sequential lineups. There are no published data that address the extent to which the traditional (simultaneous) mockwitness approach can be used with sequential lineups. How well does the simultaneous mockwitness approach calibrate to a sequential mockwitness

procedure? How well can mockwitnesses detect structural bias when a lineup is presented sequentially?

Are Mockwitnesses Sensitive to Instructional Variation?

Traditionally, the mockwitness procedure involves giving noneyewitnesses a description of a culprit, showing them a lineup, and asking which lineup member they think is the person who perpetrated the crime. Interestingly, the literature on mockwitness identification has virtually ignored whether the appropriate questions are being asked of mockwitnesses during the task. An exception is Wells and Bradfield (1999) who investigated whether varying the specific question asked of mockwitnesses would have an impact on lineup fairness assessments. Their results showed that lineups were rated as significantly more fair (in terms of functional size) when mockwitnesses were asked "Which person best fits the description?" rather than "Which person was the witness describing?" or "Which person is the accused?"

It is possible that other instructional variations may have an impact on lineup fairness assessments. For instance, eyewitness identification research has reliably shown that eyewitnesses tend to make more accurate identifications when they are given an unbiased lineup instruction (i.e., instructions that allow witnesses to choose no one from the lineup) than when given biased instructions (Malpass & Devine, 1981; Steblay, 1997). Also, manipulation of witnesses' *decision criterion* level—the willingness to make a particular decision given its expected outcome—via the lineup instructions given can also affect identification outcomes (Dunning & Stern, 1994; Gonzalez, Ellsworth, & Pembroke, 1993; Malpass & Devine, 1980; Malpass & Devine, 1984). Results of a study by Cutler and Penrod (1988) show that eyewitnesses make fewer errors when they are told by experimenters that there will be serious consequences as a result of their lineup decision (high criterion instructions), as opposed to those given more lax, neutral criterion instructions. Given these mockwitness and eyewitness data, it would be useful to investigate the impact of a range of mockwitness task instructions on lineup fairness assessments, beginning with those already identified as having an effect on eyewitness performance.

What are the Main Sources of Identifying Information for Mockwitnesses?

Typically, the only source of information given to mockwitnesses is the physical description of the culprit as provided by an eyewitness. Given this information, mockwitnesses are then asked to select who they think is the culprit from a lineup. There is an underlying assumption that mockwitnesses' identification decisions are based solely on this descriptive information as opposed to other sources of information, but it is unknown whether this is actually the case. Psychological research shows that decision processes and attributions are often based on physical appearance (Bull & Rumsey, 1988; Clifford & Bull, 1978), and there is some evidence that people show stereotypes for criminal appearance when presented with a series of faces (Bull & Green, 1980; Goldstein, Chance, & Gilbert, 1984; MacLin, Malpass,

& Herrera, 2000; Shoemaker, South, & Lowe, 1973; Yarmey, 1993). To our knowledge, subjective accounts of mockwitnesses' decision processes with regard to the task have not been collected in order to assess on what basis mockwitnesses actually make lineup choices.

PRESENT RESEARCH

Because the mockwitness approach is often used by attorneys and research scientists in the forensic arena, it is important to identify procedures that will facilitate mockwitness performance and ultimately render the most useful and accurate lineup measurement tool. However, there is very little published data examining the utility of the mockwitness approach under various procedural conditions, as well as any examination of the cognitive approach to the task. With that in mind, this research was designed to examine (1) whether mockwitnesses' estimates of lineup bias would vary as a function of changes in the standard mockwitness procedure, and (2) the source(s) of mockwitnesses' identifying information. To our knowledge, "standard" procedure includes that of simultaneous lineup presentation and a forced (biased) task instruction. Based on this, all mockwitnesses in this study were presented with sequential lineups that contained one of two targets (to control for stimulus effects; see Wells & Windschitl, 1999), and were exposed to variations in structural bias (fair vs. biased lineups) and variations in task instructions (fair vs. biased instructions; high vs. low criterion instructions) in order to investigate their impact on target identification accuracy. It was hypothesized that

(1) When presented with a structurally biased lineup, mockwitnesses' target choices would be significantly greater than chance expectation. When presented with a structurally fair lineup, we expected that choice rates of the targets would not differ from that of the other lineup members (because of the high degree of similarity among lineup members). We also expected significantly more nonchoosers under the fair lineup conditions than under unfair lineup conditions (again, because of target-filler similarity).
(2) Fewer errors would result when mockwitnesses were given unbiased lineup instructions and high criterion lineup instructions as compared to biased, low criterion instructions.
(3) Mockwitnesses would report utilizing the physical description information of the culprit as their main source of identifying information during the lineup task as opposed to other information.

METHOD

Participants

Mockwitnesses were 480 undergraduate students of University of Texas at El Paso enrolled in psychology, sociology, or criminal justice courses. Cell size was determined using a statistical power analysis based on Cohen's estimate for a medium

effect and a power estimate of .80 (Cohen, 1988). One hundred and twelve additional participants were used to construct photographic lineups. The ethnicity of participants reflected the ethnic make-up of UTEP and El Paso (approximately 72% Hispanic). Participants received partial course credit for their participation in this research. All participants were treated in accordance with the "Ethical Principles of Psychologists and Code of Conduct" (American Psychological Association, 2001).

Design

The study was a 2 (lineup structure: fair vs. unfair) × 2 (lineup instructions: biased vs. unbiased) × 2 (criterion instructions: high vs. low) × 2 (target one vs. target two) between-groups design. Participants were randomly assigned to conditions (30 per cell). The four variables examined are described below.

Lineup structure. Half of the participants viewed a fair lineup (high target-filler similarity) and the other half viewed a biased lineup (moderate target-filler similarity). Individuals comprising the biased lineups were moderately similar in appearance to one another rather than very dissimilar in order to avoid ceiling effects.

Lineup instructions. Half of the participants were given biased instructions ("the culprit is in the lineup"; forced-choice) and the remaining participants were given unbiased instructions ("the culprit may be absent from the lineup"; including a none-of-the-above option) prior to viewing the lineup.

Criterion instructions. Prior to viewing the lineup half of the participants were given high decision criterion instructions ("This lineup is from a real court case; it is very important for the defendant and the police whether you are able to pick him out so do not identify someone unless you are absolutely certain.") and the remaining participants were given low decision criterion instructions ("This is simply a research experiment and there will be no consequences as a result of the decision you make.").

Target. Half of the participants viewed a lineup containing Target 1 (T1) whereas the remaining participants viewed a lineup containing Target 2 (T2).[6] Both Targets were Hispanic males.

Photographic Lineups

Four six-member lineups were created, two *fair* with high target-filler similarity (T1 and T2) and two *unfair* with moderate target-filler similarity (T1 and T2). Black and white, frontal pose, head and shoulder photographs of Hispanic males obtained from a computerized mugshot database at the El Paso (Texas) Police Department were used to construct the lineups. Two targets were initially selected from

[6]The target presence versus absence paradigm—an important factor generally included in eyewitness identification research designs—was not a factor in this experiment. Because the purpose of mockwitness evaluation is only to detect structural bias in a lineup, whether the suspect is actually the culprit or not is irrelevant.

the database that did not have any distinguishing facial characteristics (e.g. scars, tattoos). Independent raters ($N = 15$) provided detailed descriptions of the targets and the modal descriptors comprised their physical descriptions. A bank of photographs was then obtained from the database to serve as potential fillers in the lineups, selected based on both their degree of similarity to the targets' physical appearance and the degree to which they matched the targets' physical descriptions (Wells, Rydell, & Seelau, 1993). Photographs were chosen that matched each target's description completely and were highly similar in appearance to each target (potential fillers for fair lineups), and that matched approximately half of each target's description and were moderately similar in appearance to each target (potential fillers for unfair lineups). Independent judges ($N = 60$) rated the photographs on their degree of similarity to each target on a scale ranging from 1 (*not at all similar*) to 10 (*extremely similar*), and then rank ordered the photographs based on their similarity to each target. Each photograph's mean rating score and mean ranking score were converted into z scores and then averaged in order to generate one similarity score for each photograph. Based on this, the targets plus the five photographs with the highest similarity scores from the bank of similar photographs comprised the *fair* six-member lineups. For the *unfair* lineups, five photographs were randomly selected from those photographs comprising the middle tercile of similarity scores among the moderately similar bank of photographs. Those five photographs plus the targets comprised the six-member unfair lineups. The photographs were converted into 35 mm transparencies to be displayed as slides.

Manipulation Check

Mockwitnesses ($N = 28$) evaluated all lineups as a check on their fairness and unfairness. They were told that a man had recently committed an armed robbery and were given his physical description. They were then asked to choose which photograph was that of the culprit from a lineup displayed sequentially. Protocols contained 20 numbered blanks in which to write "yes" or "no" as to whether each photograph was that of the culprit, although only six photographs were shown.[7] Participants were instructed they could answer "yes" to only one photograph and each slide was shown for approximately 20 s. A z test for the significance between the proportion of target choices and chance expectation (16%) was used to statistically evaluate lineup bias whereas lineup size was determined using the Tredoux' E index[8] (Tredoux, 1998). For the fair lineups, Target 1 was chosen by 14% of mockwitnesses ($z = 0.29, ns$) and lineup size was 4.45 whereas Target 2 was chosen by 16% of mockwitnesses ($z = 0.32, ns$) and lineup size was 4.30. For the unfair lineups, 64% of mockwitnesses chose Target 1 ($z = 3.70$, $p < .01$) and lineup size was 2.18 whereas 60% chose Target 2 ($z = 3.18$, $p < .01$) and lineup size was 2.33.

[7]It is common practice in eyewitness identification studies for participants presented with sequential lineups to be misled about the number of lineup members to be shown. This is done as an attempt to avoid a shift in choice criterion as participants approach the final lineup members (Lindsay, Lea, & Fulford, 1991; Lindsay & Wells, 1985; Melara, DeWitt-Rickards, & O'Brien, 1989).

[8]Lineup size is calculated using Tredoux' E, a statistic favored over Effective Size (Malpass, 1981) because a sampling distribution can be computed, allowing for statistical analysis (Tredoux, 1998; see Corey, Malpass, & McQuiston, 1999).

Mockwitness Procedure

The mockwitness procedure was conducted using participant groups ranging between one and six. Following informed consent procedures, participants were given experimental instructions including the physical description of a man who had recently committed an armed robbery, and a set of instructions depending on their experimental condition. Instruction sheets contained 20 numbered blanks for a "yes" or "no" decision as to whether each photograph was that of the culprit, although six lineup members were actually displayed. Participants were not explicitly told how many photographs would be shown. They were not instructed to choose only one lineup member, as we were interested in examining the pattern of choices that would emerge when mockwitnesses are not constrained to make one choice. They then viewed one of the four lineups sequentially with each lineup member displayed for approximately 20 s. To prevent order effects, lineup photographs were counterbalanced so that the targets appeared in positions one through six equally often. Following the identification procedure, participants completed a four-item questionnaire that addressed the decision processes used during the identification task. Specifically, they were asked to indicate why they chose a particular lineup member (if any), what strategy they used when making a lineup decision, whether they chose more than one lineup member and why, and what types of information during the task had an influence on their lineup decision. They were then debriefed, thanked, and dismissed.

RESULTS

Four hundred and three participants made a lineup choice (84%). Of these, 167 participants (41%) identified one lineup member (87 with T1, 80 with T2). Two hundred and thirty-six participants (59%) identified two or more lineup members (119 with T1, 117 with T2). The breakdown of multiple lineup identifications includes the following: 121 participants chose two lineup members, 76 chose three lineup members, 30 chose four lineup members, 5 chose five lineup members, and 4 chose all six lineup members. Seventy-seven participants chose no one.

Target Identification Accuracy

Because participants were not constrained to make only one lineup choice, the large number of multiple identifications made is not surprising. What is not entirely clear, however, is the most appropriate way to deal with these data. It was also unclear what, if any, differences existed between participants who chose one lineup member versus those who chose multiple lineup members in terms of the dependent variable (target identification accuracy); therefore, it was important to examine these sets of participants separately. For the purposes of this analysis, participants who made multiple lineup identifications were considered accurate if the target was one of their selections in the combination, otherwise they were considered inaccurate. Two separate $2 \times 2 \times 2 \times 2 \times 2$ hierarchical loglinear analyses were conducted in order to evaluate the impact of the four independent variables—lineup structure, lineup

instructions, criterion instructions, and target—on (1) target identification accuracy for participants who made a *single* choice ($N = 167$) and (2) target identification accuracy for participants who made *multiple* choices ($N = 236$).[9]

Single Choice Analysis

Considering only those participants who made a single lineup choice, a significant association was found between target identification accuracy and lineup structure, $\chi^2(1, N = 167) = 42.96$, $p = .000$; $\Phi = .49$. Eleven percent of participants who viewed a fair lineup correctly identified the target from the lineup in comparison with 58% of those who viewed an unfair lineup.

Multiple Choice Analysis

When considering participants who made multiple lineup identifications, significant associations were found between target identification accuracy and lineup structure, $\chi^2(1, N = 236) = 26.69$, $p = .000$; $\Phi = .32$, and between target identification accuracy and target, $\chi^2(1, N = 236) = 5.66$, $p = .017$; $\Phi = .11$. Participants who viewed a fair lineup were less likely to correctly identify the target (45%) than those who viewed an unfair lineup (76%). Also, 65% of participants who viewed Target 1 correctly identified him, as compared to 54% who viewed Target 2. These main effects, however, are qualified by a significant interaction effect between Target Identification Accuracy × Lineup Structure × Lineup Instructions × Target, $\chi^2(1, N = 236) = 4.33$, $p = .037$. Follow-up analyses showed that, when given unbiased instructions, mockwitnesses made significantly more correct identifications when the lineup was unfair than fair for both Target 1 (92% vs. 54%, $z = 3.19$, $p = .001$) and Target 2 (72% vs. 44%, $z = 2.11$, $p = .035$). Under the biased instructions condition, a similar pattern of correct identifications for the unfair versus fair lineup manipulation was found, but only for Target 2 (72% vs. 26%, respectively, $z = 3.58$, $p = .000$). Also, participants were more likely to correctly identify Target 1 (53%) than Target 2 (26%; $z = 2.19$, $p = .028$) when presented with a fair lineup and biased instructions. A similar pattern of results was found under the unfair lineup, unbiased instructions conditions for Target 1 versus Target 2 (92% vs. 72%, respectively), but this comparison only approached statistical significance ($p = .058$).

Choosers Versus Nonchoosers

Seventy-seven participants (16% of the total sample) made no identification from the lineup they were shown. A 2 × 2 × 2 × 2 × 2 hierarchical loglinear analysis was conducted to assess the association between the independent variables and the

[9] A similar analysis was conducted that included both of these groups of participants ($N = 403$). The results were similar to those found with only multiple choice participants: target identifications were more likely when participants viewed an unfair lineup (67%) than a fair lineup, 33%; $\chi^2(1, N = 403) = 56.70$, $p = .000$; $\Phi = .36$, and participants correctly identified Target 1 more often than Target 2, 58% versus 42%; $\chi^2(1, N = 403) = 8.56$, $p = .003$; $\Phi = .13$. These effects are qualified by a marginally significant interaction effect between Target Identification Accuracy × Lineup Instructions × Criterion Instructions, $\chi^2(1, N = 403) = 3.85$, $p = .049$; however, follow-up analyses did not reach statistical significance (all p's $> .05$).

dependent variable, choosing rate (choosers vs. nonchoosers). No significant interactions or main effects were found.

Target Position

We were interested in examining whether participants showed a preference for any of the six positions in the sequence of faces. Of those who made a lineup choice ($N = 403$), the frequencies with which positions one through six were chosen were 200, 131, 108, 127, 113, 127. The frequencies within these categories are not mutually exclusive; participants who selected more than one lineup member are represented in multiple categories. Still, there is an obvious preference for the first face shown. Broken down by choice type, the proportions with which the photograph in position one was chosen by single choice participants (29%) and multiple-choice participants (64%) both significantly departed from chance expectation, $\chi^2(5, N = 167) = 28.51$, $p = .000$ and $\chi^2(5, N = 236) = 18.46$, $p = .000$, respectively.

Basis for Mockwitness Identifications

Following the identification task, participants were asked to respond to four items designed to address the decision processes of mockwitnesses when faced with the task. We were interested in attempting to identify the types of factors that influence the decisions mockwitnesses make. The first three questions were open-ended and the fourth question was close-ended. For the open-ended questions, all of the information participants provided was coded into multiple response categories. These categories were derived post hoc from the particular responses given for each question. This was done in order to reduce the information given into a more manageable and meaningful response set so that proportions of responses could be calculated. Response categories were nonmutually exclusive because many participants gave information that could be classified into multiple categories.

When asked to (1) *Describe in as much detail as possible why you made the choice(s) you made concerning the slides you were shown* and to (2) *Describe the strategy you used in making your identification*, participants gave very similar responses. For both questions, over 80% of participants reported having based their lineup choice(s) on the verbal description they were given, most often mentioning specific descriptors or saying "he fit the description." Again for both questions, 7% of respondents made stereotype attributions (e.g., "he looked like a criminal"). Participants also made reference to the target's general physical appearance (12% for question 1 and 1% for question 2). Responses from fewer than 4% of participants for both questions included the mental picture generated based on the description, a process of elimination strategy, facial expression, first impression, checklist strategy, instinct, and guessing. When asked (3) *If you answered "yes" to more than one person in the lineup, please explain why in as much detail as possible*, 92% reported that the verbal description influenced their choices and 6% made a stereotype attribution. Less than 1% of participants referenced the general physical appearance, facial expression, instinct, and guessing. Last, when asked (4) *What do you think had more of an influence on the decision you made*? and presented with three possible

responses, 38% endorsed "The verbal description of the offender," 36% endorsed "The image of the offender's face that you generated from the verbal description," and 27% endorsed "The pictures you were shown," $\chi^2(2, N = 480) = 9.91$, $p < .01$.

DISCUSSION

Considering mockwitnesses who made *one* lineup choice, the data support hypothesis one in that structural bias was detected by a significant proportion of participants when presented with a biased sequential lineup, but for those shown a fair sequential lineup the identification rate of the target did not significantly differ from that of other lineup members.[10] Manipulations of lineup instructions and criterion instructions had no detectable effect.[11] Because a power analysis was used to determine sample size, we are confident that these null findings are not due to insufficient power. Although a positive interpretation of null findings is always somewhat uncertain, if real effects were present our methodology would have an excellent chance of detecting them. At the same time, we have been unable to find attributes of this study that would artifactually suppress these effects. Therefore, we interpret these null effects as an indication that these mockwitnesses' choices were not influenced by these manipulations. Overall, based on these single-choice data, the sequential mockwitness procedure appears to be comparable to the traditional simultaneous procedure in terms of detection of structural bias.

Interpretation of the data is less clear for participants who made *multiple* lineup choices. These mockwitnesses identified the target significantly more often than is expected by chance when presented with *both* fair and unfair lineups.[12] It is not surprising that the targets would be among the lineup members identified in any combination of multiple selections since those individuals best matched the description information given to participants. Although not easy to disentangle, we believe the interaction effect is largely driven by the main effects for lineup structure and for target. Similar to participants who made a single choice, multiple-choice participants correctly identified the target more often when the lineup was unfair than fair. Also, these participants made more correct identifications when viewing Target 1 than Target 2 (see below). Manipulation of lineup instructions appeared to have no meaningful effect, and manipulation of criterion instructions had no detectable effect.

Clearly, these data show that participants approach the mockwitness task with individual orientations. Some adopted a seemingly higher decision criterion when making their lineup decision in choosing only one individual in the lineup. Others approached the task with what appears to be a much more lax decision criterion, as is evident by the large proportion (nearly half) of multiple-choice participants who selected three or more lineup members. Presumably, those who selected multiple lineup members would have adopted a more stringent approach to the task if

[10] $z = 7.63$, $p < .001$ and $z = 1.65$, $p > .05$, respectively.
[11] $p > .05$, $\Phi = .11$ and $p > .05$, $\Phi = .003$, respectively (for single choice participants), $p > .05$, $\Phi = .11$ and $p > .05$, $\Phi = .05$, respectively (for multiple choice participants).
[12] $z = 6.41$, $p < .001$ and $z = 14.50$, $p < .001$, respectively.

experimenters constrained them to make only one choice. Ultimately, the complexity of these results and conclusions with multiple-choice participants is undesirable in terms of developing a useful lineup evaluation tool for sequential lineups.

The instructions and the nature of the mockwitness task appear to orient mockwitnesses to the physical description of the offender as the primary source of identifying information. Participants generally reported that the description information provided a basis for their lineup choice(s), and other sources of information had little influence. Yet, a particularly notable response given by a few participants was that their lineup choice(s) was driven by their own stereotypes or schemas for criminal appearance. The idea that "criminality" stereotypes and schemas can have an impact on decision making when faced with a lineup task has direct applications in the eyewitness domain and should be further explored in both mockwitness and eyewitness paradigms (see MacLin, Malpass, & Herrera, 2000).

Stimulus Sampling, Target Distinctiveness, and Position Effects

For generalizability purposes, we designed our stimulus sets using multiple targets. The use of multiple targets is an important element in experimental eyewitness identification studies that is often neglected. Our data show greater accuracy with one of the two targets. This finding is of particular interest given that a manipulation check done in the early phases of this research showed the lineups to be nearly equivalent in terms of their fairness ratings across both targets. Still, our target photographs were not coded for *distinctiveness*; there is a body of literature that suggests a relationship between facial distinctiveness and subsequent recognition accuracy (Bruce, Burton, & Dench, 1994; Sarno & Alley, 1997). Future studies should attempt to control for this effect when choosing stimulus materials. Also, we counterbalanced the photographs in our lineups in order to explore any effects of position when administering the lineup sequentially. This is another often ignored consideration when designing eyewitness studies. Participants in this study showed a preference for position one in our sequential lineups, whereas the other five lineup positions were chosen at comparable rates. Taken together, these particular findings illustrate the need for an improved methodological approach to some aspects of experiments designed in this field of research.

Conclusions and Recommendations

Mockwitnesses are used only to estimate the degree to which eyewitness lineups are biased towards (and, potentially, away from) the suspect. Therefore it would be desirable that mockwitness lineup choices are influenced only by structural manipulations of the lineup—target-filler similarity—and not influenced by other factors that are known to effect eyewitnesses, such as instructional bias and decision criterion. For those participants who made only one choice, these conditions were fulfilled exactly.

In general, research has supported the use of mockwitness identifications to evaluate structural bias in simultaneously presented eyewitness lineups. Our findings suggest that this overall conclusion can be extended to the mockwitness evaluation

of sequential lineups, which affirms the mockwitness technique as a useful method for assessing lineup fairness. Based on our results, we offer several recommendations for use of the sequential mockwitness procedure: (1) mockwitnesses should be constrained to make only one lineup choice, (2) mockwitnesses should be given a forced-choice (biased) instruction so as to avoid many nonchoices, (3) mockwitnesses should be unaware of the number of photographs they will view, and (4) mockwitnesses should be instructed to choose the lineup member who best fits the description provided. In addition, experimental eyewitness and mockwitness studies should use multiple target individuals who have been evaluated a priori for distinctiveness, and lineup members should be counterbalanced so as to avoid position effects.

ACKNOWLEDGMENTS

This research was supported by grant number SBR-9730937 from the National Science Foundation. We thank Harmon Hosch and James Wood for their contribution to this research, Brian Cutler and Ray Bull for their valuable comments on drafts of the manuscript, and Kimberley McClure for her statistical advice. We also thank anonymous reviewers for their invaluable comments. Special thanks are extended to the El Paso Police Department for their help and resources in conjunction with this research.

REFERENCES

American Psychological Association. (2001). Ethical principles of psychologists and code of conduct. *Publication manual of the American Psychological Association* (5th ed.). Washington, DC: Author.

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., Ready, D. J., & Spier, S. A. (1990). Standards for evaluating the fairness of photograph lineups. *Basic and Applied Social Psychology*, *11*(2), 149–163.

Bruce, V., Burton, A. M., & Dench, N. (1994). What's distinctive about a distinctive face? *Quarterly Journal of Experimental Psychology*, *47A*(1), 119–141.

Buckhout, R., Rabinowitz, M., Alfonso, V., Kanellis, D., & Anderson, J. (1988). Empirical assessment of lineups: Getting down to cases. *Law and Human Behavior*, *12*(3), 323–331.

Bull, R. H., & Green, J. (1980). The relationship between physical appearance and criminality. *Medical Science and the Law*, *20,* 79–83.

Bull, R. H., & Rumsey, N. (1988). *The social psychology of facial appearance*. New York: Springer.

Clifford, B., & Bull, R. (1978). *The psychology of person identification*. London: Routledge & Kegan Paul.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. New Jersey: Erlbaum.

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

Cutler, B. L., & Penrod, S. D. (1988). Improving the reliability of eyewitness identification: Lineup construction and presentation. *Journal of Applied Psychology*, *73*(2), 281–290.

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

Dunning, D., & Stern, L. B. (1994). Distinguishing accurate from inaccurate eyewitness identifications via inquiries about decision process. *Journal of Personality and Social Psychology*, *67*(5), 818–835.

Goldstein, A. G., Chance, J. E., & Gilbert, B. (1984). Facial stereotypes of good guys and bad guys: A replication and extension. *Bulletin of the Psychonomic Society*, *22*(6), 549–552.

Gonzalez, R., Ellsworth, P. C., & Pembroke, M. (1993). Response biases in lineups and showups. *Journal of Personality and Social Psychology*, *64*(4), 525–537.

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

Lindsay, R. C., & Bellinger, K. (1999). Alternatives to the sequential lineup: The importance of controlling the pictures. *Journal of Applied Psychology*, *84*(3), 315–321.
Lindsay, R. C., Lea, J. A., & Fulford, J. A. (1991). Sequential lineup presentation: Technique matters. *Journal of Applied Psychology*, *76*(5), 741–745.
Lindsay, R. C., Lea, J. A., Nosworthy, G. J., Fulford, J. A., Hector, J., LeVan, V., et al. (1991). Biased lineups: Sequential presentation reduces the problem. *Journal of Applied Psychology*, *76*(6), 796–902.
Lindsay, R. C., & Wells, G. L. (1985). Improving eyewitness identifications from lineups: Simultaneous versus sequential lineup presentation. *Journal of Applied Psychology*, *70*(3), 556–564.
Lipton, J. P. (1996). Legal aspects of eyewitness testimony. In S. Sporer, R. Malpass, & G. Koehnken (Eds.), *Psychological issues in eyewitness identification* (pp. 7–22). Mahwah: Erlbaum.
MacLin, M. K., Malpass, R. S., & Herrera, V. (2000, June). *Physical descriptors and facial stereotypes: Implications for eyewitness identification*. Poster session presented at the American Psychological Society, Miami, FL.
Maclin, O. H., McQuiston, D. E., Kroeger, B., & Malpass, R. S. (2001, June). *A fair comparison of laboratory administered simultaneous and sequential lineups.* Paper presented at the fourth biennial meeting of the Society for Research in Memory and Cognition, Kingston, Ontario.
Malpass, R. S. (1981). Effective size and defendant bias in eyewitness identification lineups. *Law and Human Behavior*, *5*(4), 299–309.
Malpass, R. S., & Devine, P. G. (1980). Realism and eyewitness identification research. *Law and Human Behavior*, *4*(4), 347–358.
Malpass, R. S., & Devine, P.G. (1981). Eyewitness identification: Lineup instructions and the absence of the offender. *Journal of Applied Psychology*, *66,* 482–489.
Malpass, R. S., & Devine, P. G. (1983). Measuring the fairness of eyewitness identification in lineups. In *Evaluating eyewitness evidence* (pp. 81–102). London: Wiley.
Malpass, R. S., & Devine, P. G. (1984). Research on suggestion in lineups and photospreads. In G. L. Wells & E. F. Loftus (Eds.), *Eyewitness testimony: Psychological perspectives* (pp. 64–91). Melbourne: Cambridge University Press.
Malpass, R. S., & Lindsay, R. C. L. (1999). Measuring line-up fairness. *Applied Cognitive Psychology*, *13,* 1–7.
McQuiston, D. E., Malpass, R. S., & MacLin, O. H. (2002). *Sequential vs. simultaneous lineups: A critical review of methods and theory*. Manuscript in preparation.
Melara, R. D., DeWitt-Rickards, T. S., & O'Brien, T. P. (1989). Enhancing lineup identification accuracy: Two codes are better than one. *Journal of Applied Psychology*, *74*(5), 70–713.
Sarno, J. A., & Alley, T. R. (1997). Attractiveness and the memorability of faces: Only a matter of distinctiveness? *American Journal of Psychology*, *110*(1), 81–92.
Shoemaker, D. J., South, D. R., & Lowe, J. (1973). Facial stereotypes of deviants and judgments of guilt or innocence. *Social Forces*, *51*(4), 427–433.
Sobel, N. R. (2001). *Eyewitness identification: Legal and practical problems* (2nd ed.). New York: Clark Boardman Callaghan.
Sporer, S. L. (1993). Eyewitness identification accuracy, confidence, and decision times in simultaneous and sequential lineups. *Journal of Applied Psychology*, *78*(1), 22–33.
Steblay, N. M. (1997). Social influence in eyewitness recall: A meta-analytic review of lineup instruction effects. *Law and Human Behavior*, *21*(3), 283–297.
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), 459–473.
Technical Working Group for Eyewitness Evidence. (1999). *Eyewitness evidence: A guide for law enforcement*. Washington, DC: National Institute of Justice (NCJ 178240).
Tredoux, C. G. (1998). Statistical inference on measures of lineup fairness. *Law and Human Behavior*, *22*(2), 217–237.
Wells, G. L., & Bradfield, A. L. (1999). Measuring the goodness of lineups: Parameter estimation, question effects, and limits to the mockwitness 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., Malpass, R. S., Lindsay, R. C. L., Fisher, R. P., Turtle, J. W., & Fulero, S. M. (2000). From lab to the police station: A successful application of eyewitness research. *American Psychologist*, *55,* 581–598.
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., Small, M., Penrod, S., Malpass, R. S., Fulero, S. M., & Brimacombe, C. A. E. (1998). Eyewitness identification procedures: Recommendations for lineups and photospreads. *Law and Human Behavior*, *23*(6), 603–647.

Wells, G. L., & Windschitl, P. D. (1999). Stimulus sampling and social psychological experimentation. *Personality and Social Psychology Bulletin*, *25*(9), 1115–1125.

Yarmey, A. D. (1993). Stereotypes and recognition memory for faces and voices of good guys and bad guys. *Applied Cognitive Psychology*, *7,* 419–431.