1-1-2020

Assessment of bias in police lineups.

Nancy K. Steblay
Augsburg College, steblay@augsburg.edu

Gary L. Wells
Iowa State University

Psychology, Public Policy, and Law

Assessment of Bias in Police Lineups

Nancy K. Steblay, Gary L. Wells

Funded by:

- National Science Foundation

Assessment of Bias in Police Lineups

Abstract

Materials from five extant field studies were analyzed to determine the level of structural bias in police lineups. Depending on the jurisdiction, between 33% and 68% of lineups sampled from 1,548 real police lineups scored as suspect-biased using mock witness proportion score. The suspect did not draw a fair portion of mock-witness selections in 20% of field lineups (reverse-biased lineups). Lineup fairness measures revealed that a point estimate (mean) for a set of lineups can mask significant problems in lineup construction, and that any single lineup should not be assumed to be fair based on an aggregate score. A sample of 190 lineups from the field data of Wells, Steblay, & Dysart (2015) was used to conduct four new studies that examined the relationship between lineup structure and real eyewitness decisions. Our primary hypothesis—that real eyewitness decisions could be predicted by lineup bias measures—was partially supported. Suspect identifications from simultaneous (but not sequential) lineups were disproportionately linked to suspect-biased lineups. Suspect identifications from suspect-biased lineups were almost twice as frequent for simultaneous than for sequential lineups. Additional experimental tests of these field lineups using mock-witness measures demonstrated that detailed descriptions produced higher lineup bias scores than did brief descriptions, indicating that brief descriptions can hide substantial lineup bias. Mock-witnesses were able to find the suspect in some lineups via cues not readily apparent in the description alone. The data support the idea of using a framework of descriptors to improve the fairness of lineups. Study outcomes demonstrate the realities and nuances of field lineup structure that necessitate updated consideration of the use of mock-witness measures and a blended approach to lineup construction.

Assessment of Bias in Police Lineups

The core rule for effective lineup construction is that there should be only one suspect per lineup along with at least five fillers who do not make the suspect stand out based upon physical appearance or other contextual factors. This essential recommendation dates from the earliest days of eyewitness identification research (Lindsay & Wells, 1980) and has been carried through science-based reforms to the present as a means to collect eyewitness memory evidence using a fair non-suggestive strategy and to reduce eyewitness identification error (Wells, et al., 2020). The lab-based evidence that lineup structure matters is well established (Fitzgerald, et al., 2013). Eyewitness identification decisions are directly affected by the quality of fillers chosen to surround the suspect. When an innocent lineup member fits the description but the fillers do not, an innocent person's risk of being mistakenly identified increases dramatically compared to when the fillers also fit the description (Lindsay, et al., 1999; Lindsay & Wells, 1980; Wells, et al., 1993). The reason that good fillers protect an innocent suspect is because fillers who fit the description siphon a large share of the mistaken identifications away from the innocent suspect and toward the fillers, who will not be charged because fillers are known-innocents (Smith, et al., 2018).

Actual Eyewitnesses Versus Lab Experiments

Lab experiments constitute almost the entire scientific literature on eyewitness identification memory and lineup procedure, for good reason. Definitive ground truth (an identification as accurate or mistaken) and control of variables to establish cause-and-effect are more readily accomplished in the lab. On the other hand, lab experiments are not fully representative of the population and experiences of actual eyewitnesses to crime. Lab participants usually know that they are not witnessing an actual crime and that a mistaken identification will not result in wrongful conviction of an innocent person. Lab researchers show lineups to all participants following a witnessed event whereas in actual cases lineups may be shown only to those eyewitnesses who are motivated to assist the investigation and those who feel that they were close enough and paid enough attention to the perpetrator to attempt an identification. Lab studies are often based on a witnessed video rather than live events (for review, see Wells & Quinlivan, 2008). Moreover, laboratory researchers often take great pains to build fair lineups and to assess lineup structure effectiveness.

Field studies present interpretive challenges of their own. The term *field study* (or *field lineups)* in the current article refers to the study of actual police lineups that were constructed by police and presented to eyewitnesses in the course of criminal investigations. Unlike a lab experiment, ground truth regarding a suspect identification is unknown in real cases; if the eyewitness identifies the suspect, it might be an accurate identification or might be a mistaken identification. But this does not mean that we can never know if an eyewitness made a mistake in a field study. If a lineup is set up properly—with fillers who are known innocent persons, not suspects—then an eyewitness filler identification is a mistaken identification. For this and other reasons, eyewitness identification decisions in real investigations can supplement knowledge of psychological processes gained from lab experiments (Steblay, 2018). Moreover, field lineups offer ecologically-valid eyewitness identification experiences that stem from a variety of crime types and circumstances, such as differences in illumination, distance, exposure time, and culprit disguises. and they involve a wide-ranging set of witnesses and perpetrators, descriptions of the offenders, and importantly, lineup structures imposed by case detectives.

Eyewitness experts have accumulated photos of live lineups or copies of photo-lineups for decades, typically from defense attorneys for whom they are providing consultation or giving expert testimony. These lineups likely come to the attention of the experts precisely because they appear to the defense to be biased against the client and because the eyewitness has picked the suspect (rather than picking a filler or making no identification at all). However, analyses of lineup outcomes in actual cases indicate that approximately 60% of lineups do not result in the suspect being identified (Wells, et al., 2020, Table 1). All of this suggests that police lineups that come to the attention of eyewitness experts might be highly unrepresentative of police lineups in general.

What is the prevalence of lineup bias in real world lineups? Does lineup bias affect real eyewitness decisions? The answers to such questions require comprehensive documentation of field data: preserved lineup photos, police records regarding the basis for filler selections (such as an eyewitness description of the culprit), and the decisions of the real eyewitnesses who viewed the lineups. As might be imagined, such data are difficult to come by. In striking contrast to the vast published lab literature, there are a mere 11 published studies of aggregate data for lineup decisions by actual eyewitnesses to serious crimes in police jurisdictions (Wells, et al., 2020). Only three of the 11 published field studies retained the materials necessary to assess lineup structure fairness (Klobuchar, et al., 2006; Wells, et al. 2015; Wixted, et al., 2015).

The Unique Data Set of Wells, Steblay, and Dysart (2015)

A large set of materials from 494 real lineups was collected by Wells, et al. (2015). All of these lineups were conducted by police using double-blind procedures. The centerpiece of the current project is a sample of 190 of these 494 lineups (see Appendix 1). For Study 1, the sample of 190 lineup photos and witness descriptions from Wells, et al., was assessed for fairness of lineup structure using a lab-based mock-witness procedure. Subsequently, in Study 2, the real eyewitness’s decision for each of the 190 lineups was linked to the fairness measures of that lineup. This allowed testing for the relationships among lineup structure, fairness measures, and real eyewitness decisions.

Three methodological aspects of the Wells, et al. study (2015) are especially relevant to the understanding and interpretation of this unique set of data. First, Wells, et al. conducted a field experiment in which the police employed double-blind lineup administration and an automatic computer system that made a contemporaneous record of every response for each eyewitness. The focus of our study is eyewitness lineup decisions. Hence, the integrity of our analyses depends on the supposition that the lineup administrators were not steering witnesses toward suspects or away from fillers and that every lineup and every identification decision of every eyewitness (whether the selection was a suspect, a filler, or no identification) was faithfully entered into the record. It is increasingly clear from both actual cases (e.g., Behrman and Davey, 2001) and lab experiments (e.g., Rodriguez & Berry, 2014) that lineup administrators who are not blind as to whether the witness picked the suspect or picked a filler make records when the witness identifies a suspect but commonly fail to do this when the witness identifies a filler. A 2013 national survey of U.S. police forces found that 37% of agencies openly admitted that they do not write reports on a lineup if the witness does not identify the suspect (Police Executive Research Forum, 2013). This poses a problem for archival studies in which researchers trace back through police files to find cases in which lineups were shown to eyewitnesses. If a witness made no identification or identified a filler and no report was written, then it would appear that no lineup was conducted. Accurate summary and assessment of eyewitness identification decisions require that the biases inherent in non-blind lineup administration and incomplete reporting be avoided.

Second, the photos and their presentation order were preserved in the police record for the Wells, et al. study, as was the description of the offender provided by the eyewitness. This allowed calculation of lineup fairness indices in our current study to see if such measures would predict identification decisions of the eyewitnesses. Lab researchers have artificially created lineup biases and shown that bias influences lab participants' identifications (e.g., Wells et al., 1993) but until now no data have been collected to determine whether variations in lineup bias are related to identification decisions in actual cases.

Finally, the Wells, et al. study experimentally manipulated, with true random assignment by lineup, the presentation of a lineup as sequential (one photo at a time) versus simultaneous (all photos displayed together). Importantly, random assignment to simultaneous versus sequential presentation was done only *after* the police selected the lineup fillers. Accordingly, lineup fairness/bias and simultaneous/sequential presentation were methodologically orthogonal. This critical methodological feature of the Wells et al. study allowed us in the current study to cleanly ascertain differences that may exist between the two lineup formats in how eyewitnesses respond to lineup bias.

Almost three decades ago, Lindsay, et al., (1991) provided laboratory data indicating that sequential lineup presentation offered some protection against biased lineup construction. The rationale for predicting that lineup biases have less impact for the sequential lineups than simultaneous lineups is rooted in the sequential lineup safeguard that witnesses cannot compare lineup photos side-by-side. Therefore, lineup bias should be less salient in a sequential display. Moreover, unlike a simultaneous lineup, a sequential lineup is back-loaded (i.e., the witness does not know how many photos are yet to be viewed). Even if the suspect is the only one who fits the description, back-loading prevents the eyewitness from knowing whether there might yet be another (not yet viewed) lineup member who also fits the description. Hence, there is a conceptual basis to expect that lineup bias is less likely to influence the eyewitness in a sequential lineup than in a simultaneous lineup. This is the first test of that hypothesis with actual eyewitnesses.

Relevant Questions for the Current Project

This project assessed lineup construction fairness in real police lineups and examined how eyewitness identifications decisions in the police investigations aligned with lineup fairness measures from mock-witness tests. Note that we refer to the real witnesses as eyewitnesses and lab participants as mock-witnesses. The project is relevant and timely for multiple reasons, as addressed through the following questions.

(1) *What is the prevalence of biased lineups in the field?* Fair lineup structure is central to best practices in the collection of eyewitness identification evidence. Therefore, it is imperative to investigate the extent of biased lineups in police practice, a problem that may persist even with best training efforts. This is important because there is now solid evidence that using a pristine lineup (a lineup absent of lineup bias) is a necessary condition for a reliable confidence-accuracy relationship (Wixted & Wells, 2017; Sauer, et al. 2019). Moreover, some researchers assume that field lineups are fair in their development of lineup models and make strong claims based on that assumption (e.g., Wixted, et al., 2015; Seale-Carlisle, et al., 2019). This begs the question: do real police investigations adhere to the requirement of fair lineup construction?

(2) *Does lineup bias predict individual eyewitness behavior?* This compelling question taps into whether principles developed from laboratory studies will receive support in cases of actual eyewitnesses to serious crimes. Two specific hypotheses are tested in this study. First, standard measures of lineup fairness should vary across lineups and be associated with identification decisions of real eyewitnesses. Specifically, we expect disproportionate suspect selections from biased lineups. Second, lineup bias effects on eyewitness decisions should be less pronounced for sequential than for simultaneous lineups. The rationale for these hypotheses is more fully developed in the studies below.

(3) *Do standard measures adequately assess field lineup fairness?* Existing measures of lineup fairness have come under recent scrutiny. Mansour, et al., (2016) examined the properties and usefulness of mock-witness measures. Their experimental lab research probed the unique, although not fully independent, information provided by two forms of lineup fairness measures: lineup bias and lineup size. These researchers cautioned that current fairness measures unfortunately suffer from low reliability and may not adequately meet expectations for research or real-world cases. Sauer, et al., (2019) advised that aggregate lab data may not speak well to an individual lineup in a legal case because lineup bias can be indeterminable and hidden even in well-intentioned police practice. It is reasonable to question whether extant lineup quality measurements are meaningful for field lineups and whether triers-of-fact in real cases can reasonably assume that a confident suspect identification rests on a unbiased lineup.

(4) *Is there practical value in a "blended approach" to lineup construction*? The recent eyewitness white-paper (Wells, et al., 2020) addressed how police should build a fair lineup. A blended approach for choosing fillers is recommended that uses aspects of both a match-to-description strategy (fillers based on similarity to the *eyewitness's description* of the culprit) and a resemble-suspect strategy (fillers based on similarly to *appearance* of the suspect). The complexity of real eyewitness experience and police investigations makes it likely that there will be nuances in cases and exceptions to basic principles. For example, police may develop a suspect who does not match the witness's description, or the witness's description may be vague. The current field data can reveal instances of challenges that arise in the field, and perhaps the analyses can suggest means to avoid or address such problems.

Study 1a: Measurement of Lineup Fairness in the *Wells 2015* field study

The conceptual foundation for understanding lineup fairness stems from the idea that lineup members who do not fit the verbal description that the eyewitness gave of the perpetrator are nominally but not functionally present in the lineup and hence suggest to the eyewitness which person to pick, the one who best fits the witness's description (Wells, et al., 1979). A fair lineup structure will not make it obvious as to which lineup member is the accused. Three intersecting aspects of this foundation are important to note: the starting point for structuring the lineup is the description provided by the eyewitness; lineup fairness diminishes when lineup members fail to meet the description of the culprit, that is, when the number of viable options are reduced (lineup size); and lineup fairness is undermined when the suspect stands out among lineup members (lineup bias).

Lineup quality is not a static property of a lineup, but rather a result of the interaction between the lineup members, the suspect, and the description provided by an eyewitness (Wells & Bradfield, 1999). To determine the fairness of a lineup, a useful laboratory paradigm has been developed that employs the eyewitness's description of the suspect along with the lineup. Mock-witnesses (people who did not see the crime or the perpetrator) are given the eyewitness's description of the offender, shown the lineup, and asked to pick the person they think is the suspect in the case (Doob & Kirshenbaum, 1973). Armed with this simple information, the mock-witness makes a selection based not on memory, but on inference from the physical description of the culprit and any other available cue from the lineup. With a fair lineup, these mock-witness choices should be equally distributed across the lineup members (1/6 of mock-witness picks should fall on each member of a six-person lineup). Conversely, if a disproportionate number of mock-witnesses can identify the suspect—based on no memory of the crime—the real witness's selection of the suspect may be challenged in that the identification may be based on non-memory factors instead of recognition. The rate of mock-witness selections of the suspect (*proportion score*) thereby can reveal unfair lineup structure.

The purpose of Study 1 is to assess the fairness of real lineups from field investigations, using proportion scores derived from mock-witness procedures. For this purpose, we introduce an analysis strategy to define lineup structures that are *fair*, *biased toward the suspect*, or *biased away from the suspect,* and we assess the prevalence of these three categories of lineup structure. This strategy for describing lineup quality is the necessary first step toward exploration in Study 2 of the relationship between lineup fairness and real eyewitness decisions.

Method

Participants

A total of 550 laboratory participants (university students, staff, and volunteers) provided data through a mock-witness procedure. Participants ranged in age from 18 to 70 years (*M* = 21.04, *SD* = 5.57, *Mdn* = 20.00), and 52.9% self-reported as female. Major segments of self-reported race/ethnicity included 57% white/Caucasian, 15% black/African-American, 11% Asian, 9% Hispanic/Latino, and 5% multi-racial. In all studies reported throughout this document, participants received class credit or a cash reward and were treated in accordance with APA standards.

Materials

A sample of 190 lineups was drawn from Wells, et al. (2015), using a sampling strategy described more fully in Study 2 below. The lineups of Wells, et al. (n = 494; see Appendix 1) were all stranger-lineups that employed a double-blind protocol with a witness who was seeing the suspect's photo for the first time. The lineups were six-person photo arrays from four United States cities: Austin, Texas; San Diego, California; Tucson, Arizona; and Charlotte-Mecklenburg, North

Carolina. Most photos were high-quality color from photo repositories (only Charlotte-Mecklenburg used black/white photos). Wells, et al. (2015) randomly assigned the suspect's photo in the lineup via computer program to positions 2-6 (the suspect was never placed in position 1), thereby distributing suspect position evenly across the lineup positions 2-6. Because our sample was a stratified sample taken from the larger Wells. et al. data set, we expected to see a similar equal distribution across the possible five possible suspect positions. As expected, suspect position did not significantly differ across the possible five positions, with approximately 20% of suspects in each of positions 2-6, X^2 (4) = 2.51, *p* = .64.

Design and Procedure

A standard mock-witness protocol was employed. Experimenters were blind as to which lineup member was the suspect and which were fillers. Each lineup was presented via computer as a simultaneous six-person display along with the eyewitness's verbal description of the culprit at the top of each lineup slide. Presentation order of the lineups was counterbalanced across participants. The key instruction to participants was "Your task for each lineup is to read the description, view the lineup, and make a choice, as best you can, as to which of the lineup members you think is the accused. That is, who do you think the suspect is?" (Wells & Bradfield, 1999). The experimenter left the participant's cubicle during the rating task. The 190 lineups were divided into 10 segments, each with 18-21 lineups. Each participant-rater rated one segment of lineups. Each of the 190 lineups of our sample was rated by 50-78 mock-witnesses.

Results and Discussion (Table 1)

This analysis emphasizes *proportion score* as a direct measure of lineup bias that focuses on the suspect: is the lineup biased toward *this specific lineup member*? This metric is straightforward to interpret; it is the proportion of mock-witnesses who picked the suspect.

Proportion scores in this sample ranged from .00 to .79, and the distribution skewed toward suspect-bias. A non-parametric statistic, X^2 Goodness of Fit, was used to categorize each lineup as *fair* (proportion of suspect picks not significantly different from chance), *suspect-biased* (a proportion significantly higher than chance), or *reverse-biased* (a proportion significantly lower than chance). Additional measures derived from these 550 mock-witnesses, such as functional size and effective size, are discussed in Study 2.

Fewer than half of the lineups (43%) scored as fairly constructed, and mock-witnesses were able to detect the suspect at a rate significantly above chance in a third of lineups (33%). There was also a substantial percentage (24%) of lineups in which the suspect did not receive a fair share of mock-witness selections. This surprising outcome pattern begged the question, are the *Wells* field lineups an anomaly? Fortunately, we were able to analyze additional field data in Study 1b. Attention to the full range of proportion scores (from .00 to 1.00) was justified by existing summary reports of wide-ranging lineup proportion scores. For example, Steblay (2011) reported mock-witness proportion scores in an Evanston, Illinois, sample that ranged from .00 to .64, and a sample of lineups from Houston, Texas, revealed proportion scores from .02 to .92 (W. Wells, 2014; Wixted, et al., 2015).

Study 1b: Lineup Fairness in Four Additional Field Studies

In Study 1b, the lineup fairness analysis from *Wells* is supplemented by data from four additional field studies. Science-based recommendations for lineup structure and delivery and a requirement that all lineup outcomes be recorded were employed in two published studies (Klobuchar, et al., 2006, lineups from Minneapolis, Minnesota, hereinafter, *Klobuchar;* and Wixted, et al., 2015, lineups from Houston, Texas, hereinafter, *Wixted*). The primary investigator for the Houston Police Department (HPD) study produced both a report for the HPD (W. Wells, 2014) and a subsequent published analysis of selected data (Wixted, et al., 2015).

Two additional sets of lineups were from unpublished work, but brought into our analyses via materials obtained through a successful Freedom of Information Act (FOIA) lawsuit. The FOIA-secured raw data included lineup photos, police records, and documentation of eyewitness decisions from two cities in Illinois (Evanston and Chicago). Methodological critiques of the unpublished study that generated the Illinois lineups (Mecklenburg, 2006) are available in prior publications (e.g., Schacter, et al., 2008; Steblay 2011; Steblay 2018). For the current project, analyses regarding lineup quality were conducted for Evanston and Chicage (hereinafter, *Steblay 2011* and *Steblay 2018*, respectively).

Method

Materials and Procedure (Table 2)

The case investigator in all jurisdictions constructed the lineup according to department standards. The standard mock-witness procedure of Study 1a was followed for lineup assessments of materials available from *Klobuchar*, *Steblay 2011*, and *Steblay 2018*. The standard mock-witness procedure is assumed for the *Wixted* data, although the document did not explicitly say so (W. Wells, 2014). Only tests of stranger-lineups are reported (no prior-familiarity between witnesses and offenders) as best as could be determined through the police reports. Lineups were excluded from testing if the police report failed to indicate which lineup member was the suspect, if no witness description was provided for a specific witness, when multiple witnesses viewed the same lineup (so only one lineup could be tested from the case), if photo clarity was poor, or if a description appeared to be based on public documents rather than witness information.

The *Wixted* study is the only one of the five field studies in this report for which we did not have direct access to the lineups and police reports. The HPD document (W. Wells, 2014) reported proportion scores for each of the 60 lineups and noted that “the proportion of mock witnesses who identified the suspect were beyond chance expectation in over half of the photo spreads” (p. 27).

Results and Discussion

As in Study 1a, the non-parametric statistic X^2 was used to categorize the lineups as fair, suspect-biased, or reverse-biased. It is clear that the *Wells* data are not an anomaly (Table 1). Variability in proportion scores was substantial within each of these five field studies, and the distributions of scores were often skewed toward suspect-bias.

How prevalent are biased lineups? These field data of 365 tested lineups sampled from 1548 lineups originating from five field studies involving eight U.S. cities—the best estimates researchers have to date—revealed 149 (41%) to be significantly biased against the suspect. The percentage of well-constructed lineups never exceeded 50% across the five studies (Table 1). Police in these studies knew that data were being collected from these lineups as part of a research project, police knew prescribed procedures for lineup construction, many were trained in best practices, and they presumably did their best (at the least followed usual practice), yet lineup construction bias was substantial. We do not infer negative police intent from these biased lineups; rather, construction of fair lineups appears to be a very difficult task in practice.

The reason(s) for the extremely high rate of biased lineups in one jurisdiction (Chicago, 68%) is unknowable, but these identification procedures are unique among the field tests in that they were all conducted live (Table 2). One possibility, raised by Steblay (2018), is that it might have been difficult for police to find live lineup fillers who adequately matched basic attributes of suspect hairstyle, height, and size. In fact, in two lineup reports, the investigators themselves noted that they had no adequate fillers for the live lineups (the subsequent suspect identifications came from lineups with functional sizes of 1.2 and 2.3). These live lineups also were typically of size five, a truncation of lineup size that may exacerbate the impact of lineup bias. The proportion score cutoffs employed with this sample were the same as with the six-person lineup samples. Had we used proportion cutoffs for a lineup of size five (chance at .20 instead of .167), the percentage of biased lineups in this sample would decrease only slightly (to 63%) and the percentage of reverse-biased lineups would increase slightly (to 17%) with no effect on the percentage of fair lineups (20%).

An important conclusion is that a point estimate (mean) of fairness scores across lineups can mask significant problems in lineup construction. An aggregate score collapsed over lineups does not necessarily offer meaningful assessment of lineup structure quality for the set of lineups or for any individual lineup. Furthermore, it is illogical and misleading to assume that reversed-biased lineups cancel out suspect-biased lineups in a set or to justify a conclusion that the overall group of lineups are fair. If half of the lineups are biased against the suspect and the other half are reverse-biased, then it would appear from an aggregate score that the lineups are fair. Moreover, we note that across these five studies, reverse-bias and suspect-bias are not evenly balanced; the bias was twice as likely to be toward the suspect. Researchers have previously reported conclusions that sets of field lineups are fairly constructed (Steblay, 2011; Wixted, et al., 2015), but these claims must be reconsidered in light of the more precise information from the current proportion-score categories.

It might be argued that reverse-biased lineups are fair in that they may draw eyewitness picks away from the suspect. Perhaps the combination of fair plus reversed-biased lineups is a better indication of fairness *for the suspect* and this brings the percentages of fair lineups closer to

2/3 in some jurisdictions. However, reverse-biased lineups bring up intriguing questions. How does reverse-bias occur in building a lineup? What are the implications for innocent and guilty suspects when reverse-bias is present? How do real eyewitnesses respond to lineups in which reverse-bias occurs? We return to the *Wells* data to address these issues.

Study 2: Lineup Fairness and Eyewitnesses Decisions

Study 2 is an examination of how real eyewitness identification decisions are associated with lineup quality (including the role of simultaneous and sequential lineup format) and an evaluation of the properties and effectiveness of the lineup fairness measures themselves. We predicted that lineup bias will be significantly related to eyewitness identification decisions, particularly that lineups that produced suspect identifications will be more biased than lineups for which eyewitnesses identified no one or chose a filler. We also predicted that this relationship between lineup bias and suspect selections will be stronger for simultaneous lineups than for sequential lineups.

Method

Participants

The 550 mock-witnesses described in Study1a provided the data for these analyses.

Materials (Samples I and II)

As noted previously, simultaneous versus sequential lineup format was randomly manipulated in the *Wells* study only *after* the lineup had been constructed. Therefore, we can test for differences between sequential and simultaneous lineups formats with a stratified random sample. To do so, Sample I required equal numbers of sequential (60) and simultaneous (60) lineups, with equal numbers of suspect, filler, and non-identifications (20 each within each lineup format; see Appendix 1), the strategy proposed in the NSF grant that funded this study. Eligible lineups for mock-witness testing were those that had a record of the eyewitness's description of the offender (63% of the 494 pristine lineups available from Wells, et al., 2015) and high photo quality (black-and-white photos were excluded from Sample I). This intended sampling strategy was restricted, however, by the low rate of filler identifications in sequential lineups. A simple modification was to limit filler identifications to 18 lineups each for sequential and simultaneous lineups rather than the prescribed 20, resulting in 116 total lineups. Power was at .70 for the primary analyses (assuming medium effects), however power to reveal differences between subsets of the data (e.g., between fillers in sequential vs. simultaneous lineups) was very limited (.30). The benefit of Sample I was to allow direct and fair comparisons between randomly selected sequential and simultaneous lineups.

Subsequently, to make better use of these field data, and to increase sample size and power of the analyses (boosting power to .80 for most comparisons), 74 additional lineups were tested with the mock-witness procedure and added to make Sample II, a total of 190 lineups that was used in Study 1a (Appendix 1). This larger sample included black and white photos from Charlotte-Mecklenburg as well as additional lineups from the other three cities. Approximately equal numbers of suspect identifications (total *n* = 77) and of non-identifications (*n* = 66) from sequential and simultaneous were maintained, but the low number of sequential filler picks (*n* = 18) resulted in additional fillers picks coming only from simultaneous lineups (*n* = 29).

Measures

Six lineup fairness measures were derived from mock-witness responses. Four measures are standard in the literature: *proportion score*, *functional size*, *effective size,* and *defendant bias* (see, e.g., Malpass & Lindsay, 1999). Two new measures of lineup fairness were *mean rank* and *explicit lineup member rejections*. *Proportion score* is the proportion of mock-witnesses who selected the suspect from the lineup. *Functional size* (Wells, et al., 1979) is calculated as the total number of mock-witnesses divided by the number who picked the suspect, an index of lineup size and bias relevant to the suspect. In a lineup that is unfair to the suspect, functional size will be lower than the nominal lineup size (six). Functional size can also range well above the nominal size of the lineup; if only one of 50 mock-witness selects the suspect, the functional size of the lineup is 50. When functional size exceeds nominal size, it is an indication of reverse bias.

Effective size (Malpass, 1981; Tredoux, 1998) indicates the number of lineup members that drew their fair share of mock-witness identifications to fulfill nominal chance expectation (e.g., 16.7%). Effective size (commonly *Tredoux's E* statistic) is an estimate of the number of plausible, adequately-functioning lineup members. This index does not take into consideration which person is the suspect and in fact the suspect may not even be among the lineup members drawing a fair share. Thus, effective size is not a measure of the suspect's risk. *Defendant bias* (Malpass, 1981) is calculated as the difference between the proportion of identifications by mock-witnesses of the suspect and the proportion of suspect identifications expected by chance in a lineup of a given effective size. If a defendant-bias score is negative, it is an indication of reverse bias.

Mean rank is a measure of lineup fairness that we introduce in this study. Mean rank is based on the mock-witnesses' rankings of which lineup member is most likely to be the suspect. A single-choice method (mock-witness chooses one person, the #1 rank) is the default procedure used in the extant literature. As far as we can tell, no previous studies have asked mock-witnesses to rank-order all six lineup members according to the likelihood that each was the suspect. We believe that these additional (ranking) data on each lineup member might add something to the estimates and increase the power of mock-witness data to predict lineup outcomes. The mean rank for the suspect should be 3.5 in a totally unbiased lineup (the average of the ranks of 1-6). An average rank below 3.5 indicates bias toward the suspect and an average rank above 3.5 indicates reverse bias. This new ranking method can always be converted back to the traditional measures (e.g., functional size or defendant bias) by just treating the number one rank of each mock witness as their choice. But the ranking method has the potential to tell us more about the lineup, by incorporating mock-witness judgments about top-scoring versus low-scoring lineup members.

Finally, mock-witnesses were instructed to mark any individual lineup member that they believe could not be the suspect. We included this measure to flag possible instances in which mock witnesses, although forced to rank all six lineup members, were not just giving low ranks but actually thought that none of the lower ranked individuals (e.g., ranked numbers 4, 5, and 6) could be the suspect. This measure is intended to directly tap witnesses' elimination of lineup members (truncation of lineup size), and indicates the proportion of mock-witnesses who made *explicit lineup member rejections*.

Results

The analyses below first briefly consider the smaller (stratified random) Sample I (n = 116) and then move to the larger Sample II. Confidence intervals are reported at the 95% confidence level, and p-values are two-tailed throughout the results. It can be noted at the outset that Samples I and II produced very similar lineup fairness statistics (Table 3).

Sample I (116 lineups)

Assessment of lineup quality. The mean proportion score (M = .20) did not differ significantly from the value of .167 for a six-person lineup, X^2 (1, N = 116) = .70, p = .40. However, proportion scores ranged from .00 to .75, a positively skewed distribution (Mdn = .14). Using the proportion-score cutoffs of Study 1, 45% (52/116) of the lineups were fair, 28% of the lineups (33/116) were suspect-biased, and 27% (31/116) were reverse-biased.

Measures of lineup quality and eyewitness decisions. If lineup bias plays a role in identification behaviors in actual cases, then lineups in which eyewitnesses identify the suspect should be more suspect-biased than those for which the eyewitness identifies a filler or identifies no one. For example, we expected that mean proportion score (as a dependent variable) should differ significantly between three groups of lineup outcomes (the independent variable). The three groups compared through a one-way ANOVA were those in which lineups had produced suspect identifications versus filler identifications versus non-identifications. Contrary to expectation, *proportion score* was not significantly higher for suspect-identified lineups (*M* = .24) compared to filler-identified lineups (*M* = .16) and non-identifications (*M* = .20), although the means were in the predicted direction. Just two measures differentiated between lineups that had produced suspect versus filler picks. Suspect identifications were from lineups of lower *functional size* (*M* = 7.39, *SD* = 8.23, *CI*: 4.76 - 10.03) than filler identifications (*M* = 14.18, *SD* = 15.66, *CI*: 8.89 - 19.49), *F* (2, 113) = 3.32, *p* < .05, *d* = .54, and lower *mean rank* (*M* = 3.11, *SD* = .83, *CI*: 2.85 - 3.37) than filler identifications (*M* = 3.57, *SD* = .75, *CI*: 3.32 - 3.83), *F* (2, 113) = 2.99, *p* < .05, *d* = .58. There were no significant differences in fairness measures for lineups that had produced non-identifications compared to suspect or to filler identifications.

Lineup format. We predicted that suspect identifications would stem disproportionately from biased lineups and that this pattern would be less pronounced for sequential than for simultaneous lineups. Our stratified random sample of 116 lineups was drawn based on eyewitness decisions (equal numbers of suspects, of fillers, and of non-identifications) from 58 simultaneous and 58 sequential lineups. In other words, we started with a set of eyewitness decisions, so that we could subsequently examine factors that are associated with those decisions. Here, we examined the relationship between *eyewitness decisions* (suspect, filler, non-identification) and *lineup fairness* (suspect-biased, fair, reverse-biased) separately for sequential and for simultaneous lineups.[1] In line with our prediction, an intriguing pattern emerged for witness decisions under the two lineup formats. Approximately half of suspect identifications came from fairly constructed lineups, with no significant difference in percentages of suspect identifications between sequential (55%) and simultaneous (50%) lineups. However, for simultaneous lineups, the remaining suspect identifications came largely from suspect-biased lineups (45% of suspect selections, 9 of 20), whereas only 25% of suspect selections from sequential lineups came from suspect-biased lineups (5 of 20). This 20% difference is not significant in this small sample, X^2 (1, *n* = 40) = 1.76, *p* = .18, but we were able to examine this matter again with the larger Sample II. For suspect identifications, proportion score was higher for the 20 simultaneous lineups (*M* = .29, *SD* = .19, *CI*: .21, .37) than for the 20 sequential lineups (*M* = .19, *SD* = .12, *CI*: .10, .28), *t* (34) = 2.06, *p* = .047, *d* = .63). This difference in lineup proportion score for suspect identifications underscores the fact that suspect identifications from simultaneous lineups come from more biased lineups than do suspect identifications form sequential lineups.

We also compared lineup fairness measures across the three witness decision categories (suspect, filler, non-identification), analyzed separately by lineup format. Simultaneous lineups produced significant differences across decision categories with respect to functional size, *F* (2, 56) = 3.18, *p* = .049 and mean rank, *F* (2, 56) = 4.95, *p* = .01, and marginally significant differences for proportion score, *F* (2, 56) = 2.71, *p* = .08 and effective size, *F* (2, 56) = 2.58, *p* = .09. Post-hoc tests showed that, compared to filler selections, suspect selections from simultaneous lineups came from lineups with significantly lower functional size (*M* = 4.56, *SD* = 3.60, *CI* [2.56, 6.55] vs. *M* = 13.41, *SD* = 15.36, *CI* [6.41, 20.40], *p* = .02, *d* = .79), lower mean rank (*M* = 2.68, *SD* = .77, *CI* [2.25, 3.11] vs. *M* = 3.57, *SD* = .81, *CI* [3.20, 3.94] *p* = .003, *d* =

1.13), higher proportion scores, (*M* = .34, *SD* = .19, *CI* [.24, .45] vs. *M* = .20, *SD* = .18, *CI* [.12, .28], *p* = .03, *d* = .76), and smaller effective size (*M* = 3.37, *SD* = 1.02, *CI* [2.80, 3.94] vs. *M* = 4.14, *SD* = 1.03, *CI* [3.67, 4.61], *p* = .03, *d* = .75).

Unlike with the simultaneous lineups, lineup quality did not vary with eyewitness behaviors (suspect pick, filler pick, or no selection) for sequential lineups. In sum, there is evidence in this stratified random sample of a stronger link between eyewitness behaviors and lineup bias for simultaneous lineups than for sequential lineups

Sample II (190 lineups)

Assessment of lineup quality. The mean proportion score of Sample II (*M* = .21) was near identical to Sample I (*M* = .20) and did not differ significantly from the chance value of .167 for a six-person lineup, *X2* (*N* = 190) = 2.33, *p* = .13 (Table 2). Proportion scores ranged from .00 to .79, a positively skewed distribution (*Mdn* = .16). Recall from Study 1 that 43% of these Sample II lineups were fairly constructed, 33% were suspect-biased, and 24% were reverse-biased (Table1). These percentages are not significantly different from Sample I, 45%, 28%, and 27% respectively, X^2 (2, *N* = 190) = .63, *p* = .73. The 30 black-and-white lineups from Charlotte-Mecklenburg produced a similar proportion score (*M* = .23, *SD* = .12) to the sample of 160 color photos (*M* = .21, *SD* = .17), *t* (188) = .80, *p* = .43.

Measures of lineup quality and eyewitness decisions. As with Sample I, one-way ANOVAs were used to compare the three categories of witness decisions—suspect identifications, filler identifications, and non-identifications—on each lineup fairness measure. In this larger sample there were no significant differences on any measures among the three witness decision categories. Similarly, logistic regression analysis using only witness choosers (suspect or filler picks) produced no significant results (all *p*s > .20). As an example, there was no difference in proportion score for the selected lineup member between suspect identifications (*M* = .22, *SD* = .15) and filler identifications (*M* = .20, *SD* = .14) *t* (122) = .57, *p* = .57. Contrary to our hypothesis, an individual witness's identification decision is not easily predicted by lineup fairness measures. However, following up on Sample I outcomes, these same relationships next were examined separately for the two lineup formats.

Lineup format (Table 4). The primary focus in this analysis is on 77 suspect identifications (38 from sequential lineups; 39 from simultaneous lineups). We examined the association between *eyewitness decision* and *lineup bias*. As in Sample 1, there was not a significant difference in percentages of suspect identifications coming from fair lineups, simultaneous (49%) versus sequential (39%), X^2 (1, *n* = 77) = .67, *p* = .41. However, for simultaneous lineups, the remaining suspect identifications came largely from suspect-biased lineups (46% of suspect selections, 18 of 39), whereas only 24% of suspect selections from sequential lineups came from suspect-biased lineups (9 of 38), X^2 (1) = 4.27 *p* = .04, a significant difference of 22% that echoed the smaller Sample I results. Suspect identifications were significantly fewer from simultaneous lineups when the lineup was biased away from the suspect, (2/39, 5%) than from sequential lineups (14/38, 37%), X^2 (1) = 11.76, *p* < .001 (Figure 1).

Lineup format is relevant in this larger sample, as it was in Sample 1, when comparing categories of eyewitness decisions (suspect, filler, non-identifications) for measures of functional size, *F* (2,101) = 4.34, *p* = .02, and mean rank, *F* (2, 101) = 2.82, *p* = .07. Post-hoc tests indicated that eyewitness suspect versus filler selections differed in (only) simultaneous lineups for functional size (*M* = 5.39, *SD* = 3.46, *CI* [4.27, 6.52] vs. *M* =11.48, *SD* = 13.60, *CI* [6.31, 16.65], *p* = .004, *d* = .61) and mean rank (*M* = 2.93, *SD* = .65, *CI* [2.71, 3.14] vs. *M* = 3.37, *SD* = .87, *CI* [3.05, 3.70], *p* = .02, *d* = .57). In sum, compared to filler selections, simultaneous lineup suspect picks are associated with lineups of lower functional size and lower mean rank (greater lineup bias).

Observations regarding the validity of measures of lineup fairness. A set of one-way ANOVAs compared the three lineup fairness categories derived from proportion score for each of the five additional lineup fairness measures (Table 5). Outcomes from three of the five measures—functional size, defendant bias, and mean rank—aligned well with the proportion score categories. These data suggest that the new defined fairness categories are meaningful metrics of lineup fairness and they illustrate how boundaries for fair versus biased lineups might be drawn for other fairness measures. For example, a perfectly fair lineup should produce a mean rank of 3.5. Our category of fair lineups produced a mean rank of 3.41, for suspect-biased lineups a mean rank of 2.43, and for reverse-biased lineups a mean rank of 4.10, significantly different values among the three categories of lineup fairness.

Mansour, et al. (2016) reported two primary dimensions of mock-witness lineup fairness measures—lineup size and lineup bias. To follow-up on this notion, we conducted a principle component factor analysis with varimax rotation with our six mock-witness measures. Correlations among measures were obtained for the 190 lineups along with loadings for two orthogonal components (Appendix 2). Component 1, the primary source of variance (52%), appears to address lineup bias. Component 2, accounting for 26% of the variance, appears to address lineup size. These two factors underscore convergent validity for measures of lineup bias—proportion score, functional size, defendant bias, mean rank—and discriminant validity for measures of lineup size—effective size and explicit rejections. Neither factor score was a significant predictor of eyewitness decision.

Our new measure of explicit rejections of lineup members was significantly correlated with effective size (*r* = -.35): a greater percentage of mock-witnesses who ruled out a lineup member was associated with reduced effective size. All of the lineups prompted some rejections of lineup members, with percentages of mock-witnesses making lineup member rejections ranging from 8% to 86% across the 190 lineups.

Filler-biased lineups. Eyewitness researchers who use mock-witness tests for evaluating lineup fairness primarily focus on implications for the suspect. But, the identification of a filler is the one identification decision in actual cases for which there is ground truth – when an eyewitness identifies a filler, it is a ground-truth error. In a fair lineup, one would expect mock-witness picks to land on fillers 83.5% of the time (five fillers each receiving a fair share of mock-witness picks), and some fillers may receive a slightly higher or lower proportion of selections just by chance. In this sample, the highest proportion score for a filler across lineups ranged from .11 to .84 (*M* = .35, *SD* = .13). A filler received a higher proportion score than did the suspect in 138 lineups (73%). Even in a fair or a suspect-biased lineup, a filler may draw more picks than the suspect. Indeed, these 138 lineups included 33% reverse-biased lineups, 59% fair lineups, and 8% suspect-biased lineups.

How did real eyewitnesses respond when a filler had a greater mock-witness score than did the suspect? One might worry that this filler-bias would exacerbate lineup errors by drawing eyewitnesses to fillers. However, only 16 eyewitnesses (8%) picked the highest-scoring filler. There was no significant relationship between this aspect of lineup structure (whether a filler or the suspect has the greatest proportion score) and eyewitness decision, X^2 (2) = 1.43, *p* = .49. That is, most eyewitnesses did not appear prone to identify the filler selected by mock-witnesses. The original Wells, et al., field study only generated 15% filler selections (75 of 494 lineups), of which

47 were tested here in Sample II. Under the best practices employed in the Wells, et al. field study, filler selections were relatively few.

Discussion

The *Wells* field data demonstrate suspect-bias in a substantial portion (33%) of field lineups rendered under seemingly well-intentioned evidence collection procedures across four jurisdictions. The aggregate data masked significant problems in lineup structure. Alarmingly, suspect-bias means that mock-witnesses—with no memory of the crime—were able to identify the suspect from a group of six lineup members, with identification rates ranging from 25% of mock-witnesses (1 in 4) to a high of 79% (4 in 5). These outcomes pose an alternative explanation for a real eyewitness's suspect-identification: the lineup composition suggested which member was the suspect (i.e., which member to choose). In these cases, the lineup structure undermines the pristine lineup conditions necessary for reliable memory evidence.

Did the lineup bias (as measured here) make a difference in eyewitness decisions? Without ground truth, we cannot be certain that lineup structure led to greater or lesser eyewitness accuracy. However, an interesting point emerged from our comparison of suspect identification decisions under two lineup formats. Sequential and simultaneous lineup formats produced similar levels of suspect identifications from fairly constructed lineups. But, suspect identifications were almost twice as frequent from suspect-biased lineups for simultaneous than for sequential lineups. As predicted, witnesses viewing the lineup in sequential format were seemingly less likely to be influenced by suspect-bias and less likely to be inhibited in their suspect selection by a lineup that was biased away from the suspect.

Two orthogonal components of lineup fairness were revealed through factor analysis, findings that echo the Mansour, et al. (2016) lab results of convergent validity for lineup bias measures and discriminant validity between size and bias measures. Four measures—proportion score, functional size, defendant bias, and mean rank—allow detection of bias both toward and away from the suspect. Fairness measures that focus on lineup size (*effective size, explicit rejections*) did not adequately discern between a lineup structure that puts the suspect at risk and a structure that may implicate a filler more than the suspect. Early lab work (Lindsay, et al., 1999) also reported that lineup bias measures better predict witness decisions than do lineup size measures. Given the reality of substantial numbers of reverse-biased lineups in the field, bias measures appear to be the more illuminating assessment.

We are left with the puzzle that these measures did not effectively predict overall eyewitness decisions of suspect versus filler versus non-identification. Of course, mock-witness responses cannot reflect the memory strength possessed by eyewitnesses. It is perhaps reassuring that mock-witnesses picks do not too easily mimic eyewitnesses. But also, we see that mock-witness picks (the basis of the predictors) favor fillers that match the description of the offender more so than does the suspect and that real eyewitnesses were not similarly drawn to such fillers. Moreover, mock-witness suspect proportion scores close to zero strongly hint that something is fishy about the lineup structure (more about this later, in Study 4). This is an arguably different scenario from laboratory lineups that underscores a need to assess lineup bias in a manner that addresses the realities of field lineups.

Study 3: Sources of lineup bias

Mock-witness decisions are based on the real eyewitness's description of the culprit. Any weaknesses in ascertaining the witness's best memory for the culprit's attributes or in fully implementing this description in lineup construction may mislead measurements of lineup fairness, as well as the police collection of eyewitness evidence. We next consider the role of eyewitness description in the creation of lineup bias, examining the two extremes of the *Wells* lineups: 20 lineups with the highest proportion scores (Study 3) and 19 lineups with the lowest proportion scores (Study 4).

Suspect-bias may be due to police failure to effectively implement a match-to-description strategy. It is also possible that a suspect may stand out for reasons related to the display (e.g., atypical photo size, background or markings) or some aspect of suspect demeanor, facial expression, a tattoo, unkempt appearance, clothing, or a stereotypical criminal appearance (Flowe & Humphries, 2011; Lindsay, et al., 1987). Study 3 tested the hypothesis that an eyewitness's description of the culprit is a key factor underlying lineup bias. Our assessment also allowed for the possibility that a suspect may stand out based on non-description aspects of the lineup. A simple two-group within-subject experiment was created. The original full description provided by the eyewitness was edited for each of 20 lineups to produce a new limited description. The independent variable was manipulated within-subject by the level of descriptive detail provided with a lineup (*Full* or *Limited*). The key dependent measure was the proportion score derived from a mock-witness procedure.

The rationale and predictions were as follows. (1) The proportion score for the *Full-Description* condition will be greater than chance, replicating the high bias scores in the Study 1 mock-witness data. (2) Description detail is expected to be the major reason for bias. Therefore, the lineups with *Limited* description should produce a lower level of lineup bias than the same lineups with the *Full* descriptions. (3) Non-description influences may produce lineup bias. Therefore, some lineups may produce bias scores above chance even in the *Limited* description condition.

Method

Participants

Participants were 77 volunteers, who self-reported as 40.3% female, 71% Caucasian, with ages ranging from 18-62 (*M* = 22.2 *SD* = 7.7).

Materials and Procedure. We selected the 20 most highly suspect-biased lineups from 190 lineups of Study 1 (proportion scores ranged from .40-.79, *M* = .55, *SD* =.11). The decision to choose 20 lineups was based on previous experience of the number of lineups that could be rated by a mock-witness within a lab session of 20-30 minutes. All lineups had associated real eyewitness descriptions that included attributes beyond race and gender. For each participant, half of the 20 viewed lineups retained the original description (*Full*) and the other half had descriptions reduced to offender race and gender (*Limited*). Order of lineup presentation was counterbalanced across subjects, and full versus limited description was counterbalanced within lineup and subjects. Each witness-participant viewed all 20 lineups and attempted to determine the accused in each lineup using the standard mock-witness procedure. The two experimenters were blind as to which lineup member was the suspect.

Results and Discussion (Table 6)

In support of Hypotheses 1 and 2, the full-description condition was only slightly lower in proportion score (.46) than the score for these same 20 lineups in Study 1 (.55), and the limited-description condition produced proportion scores at a level not significantly different from chance. All five measures of lineup bias indicated significantly more suspect-bias in lineups based on the full versus the limited descriptions (all *p*- values < .01). When a full description was provided, the mock-witnesses identified the suspect at a significantly higher rate (*M* = .46, *SD* = .17) than when the description was limited to race and gender (*M* = .18, *SD* = .11), *t* (38) = 6.41, *p* < .05, *d* = 2.08. Simply put, increased description detail produced increased suspect-bias.

An additional research question was whether mock-witnesses were able to identify the suspect in *any* of the lineups in the limited-description condition. In five of 20 lineups (25%), even the limited description produced a proportion score greater than chance. With four of these lineups, it is difficult to ascertain how the mock-witnesses arrived at the suspect, a finding that implies that suspects can stand out for (non-obvious) reasons beyond a failure to match the description.

The reason for significant bias in one limited-condition lineup (proportion score of .33) was quite straightforward. The description of the offender as “Hispanic Female” could be arguably applied to only three lineup members (as evidenced by a functional size of 3). That is, even the limited descriptors were not met by half the lineup members. The problem grew worse with the addition of a distinctive feature in the full description (“thin long black hair”). The lineup was further truncated in that it included two brown-haired women (resulting proportion score = .60, functional size = 1.66). The addition of a single distinctive feature (“thin long black hair”) diminished functional size (-1.34) more than two times as much as effective size (-.59).

The example above suggested a post-hoc analysis. We calculated discrepancy between the two measures across the 20 lineups (functional size minus effective size). When the description was limited (gender, race), functional size was higher than effective size (+.24). With the full description, this relationship reversed, functional size was now significantly lower than effective size (-.68), *t* (38) = 2.68, *p* =.01, *d* = .87. The functional size measure is more sensitive to problems for the suspect as description detail increases.

In summary, the description detail used to build the lineup and test the mock-witnesses made a significant difference for all lineup fairness measures, at least at these ends of the description continuum (very brief description versus more detailed description). These findings, of course, focus on the extremely brief descriptions (gender and race) that occur in real cases. In

Study 4, we further examine low proportion scores for a set of lineups that included 2-4 descriptors.

Finally, a small set of lineups produced significant suspect-bias with only a generic description, that is, mock-witnesses seemed to pick up on some non-obvious cue in the lineup structure. This finding lends support to a notion that a suspect may be more plausible (also to the eyewitness) because of factors that are not readily apparent vis-a-vis the description alone.

Study 4: Vague or Brief Descriptions

Default values are descriptors of the perpetrator that the eyewitness surely noticed but did not think to mention (assuming perhaps that most offenders are young) or that were not clarified or recorded by police (Lindsay, et al., 1994). This circumstance is tied to the natural limitation of the match-to-description method: eyewitness free-recall that is often vague and limited. An incomplete description (“white male, shaved head”) from an eyewitness who simply failed to report a defining default value (“young” or “clean-shaven”) may hinder an investigation and later prompt a lineup that is highly suggestive for the young male in a lineup of older or bearded men. Lindsay, et al, (1994) speculated that lineups could be more fairly constructed if a framework of basic descriptors was asked of each eyewitness.

Study 4 addressed the possibility that vague descriptions of offenders may disrupt the relationship between mock-witness measures and eyewitness decisions by hiding biased lineup structure. The hypothesis is that lineups with descriptions based on a new descriptor framework would score as more biased than lineups based on the original limited eyewitness description.

Method

Participants (Steps 1 and 2). The study involved two separate steps, a rating task and a mock-witness task. The 20 participants at Step 1 self-reported as 60% male, 45% Caucasian, with mean age of 18.84 years. At Step 2, 73 participants provided responses to each lineup in an experimental two-group within-subjects design using the standard mock-witness procedure. This group self-reported as 60% female, 54% Caucasian, with a mean age of 19.5 years (*SD*=2.71).

Method and Materials. Nineteen lineups were drawn from the set of 190, specifically those in which the suspect was picked less than 1/6 of the time, and in which no more than four descriptors were provided by the witness (2-4 descriptors regarding age, race, height, sex, or hair color).

Procedure. Participants at Step 1 were asked to describe each suspect (the photo removed from the lineup) using six categories: race, sex, age, hair, face, and size (height, weight, build). Estimates of height, weight, and build are difficult when viewing head-and-shoulders photos; participants were asked to make a general assessment of "size" as best they could surmise. The presentation order of the suspects was counterbalanced, and participants had unlimited time to complete the task. The resulting descriptions were then synthesized to create a new modal "framework" description based on a consensus strategy (details available from the author).

Step 2 was an experimental study of the impact of the description provided with a lineup on dependent measures of mock-witness lineup fairness scores. For each lineup, either the original (*limited*) description or the new framework description was provided with the lineup. Each mock-witness viewed all 19 lineups, with order of the lineups and assignment of lineup descriptions (limited vs. framework) counter-balanced.

Results and Discussion (Table 7)

As predicted, framework descriptions produced significantly higher suspect-bias scores than did the original limited descriptions. For example, lineups with the original limited descriptions produced a mean proportion score of 11%. The same lineups with framework descriptors produced a significantly higher mean proportion score (43%), *t*(36) = 6.89, *p*< .05, *d* = 2.30. Admittedly, there is a circular reasoning in this exercise, in that raters who described attributes of a single photo will set up that photo to be identified by mock-witnesses who use that same description. Nonetheless, these results aid our understanding of reverse-biased lineups and shed light on a reason for the failure of mock-witness responses to predict eyewitness decisions.

A central point is that police reports may not fully reveal offender features that will drive lineup bias. Unstated default values then may result in a mismatch between the mock-witness task and the real eyewitness's expectations. Proportion scores that indicate a fair or reverse-biased lineup (with a very brief witness description) may hide a lineup in which the suspect clearly stands out to the eyewitness, but not to mock-witnesses. In these field lineups, the most frequent framework descriptors that appeared to influence mock-witness decisions included general size (as ascertained from the head-and-shoulders photo array), facial hair, and skin tone (complexion). These are very basic attributes that can be asked of eyewitnesses early in an investigation. A better framework for descriptors may improve the predictive power of our lineup fairness measures.

There are also practical implications from this study regarding how police gather eyewitness evidence and construct lineups. It has long been argued that the match-to-description strategy may be ineffective in selecting lineup fillers when the eyewitness's description of the culprit is sparse; the less detailed the description given by the eyewitness, the greater the possibility that the lineup will include fillers who differ substantially from the appearance of the accused (Lindsay, Martin, & Webber, 1994). Moreover, because lineup fairness measures typically rely on the eyewitness's description to assess lineup bias, the exact same lineup can look fair if the description was sparse but look biased if the description was detailed (Mansour, et al., 2017). In specifying how lineup fillers should be selected by police, Wells, et al., (2020) stated that a match-to-description method should only be used by police for selecting lineup fillers if the description is reasonably complete or detailed. When the description is a sparse or fails to capture a critical feature(s) of the suspect's appearance, then the fillers should share that feature(s). In a sense, this means that a modified version of the resemble-suspect strategy for selecting fillers should be used when the description is sparse.

It is important to keep in mind that the sparse descriptions we found in these cases could have resulted from three different problems. First, it is possible that many of the records we had of the descriptions were themselves not as complete as an earlier description the eyewitness had given in a previous statement. Second, it is possible that the eyewitness could have given a much more complete statement if the eyewitness had been more fully interviewed prior to the lineup (along the lines that the Wells et al., 2020 eyewitness system-variables article describes). A third possibility is that large shares of the descriptions were quite sparse because eyewitnesses are very poor person describers, especially when stress, fear, and poor witnessing conditions impinge on witnesses.

In effect, our novel use of framework descriptors for assessing lineup fairness was an attempt to correct for the prevalence of sparse descriptions. But the use of framework descriptors raises other possible concerns. One potential concern is that the framework-descriptor approach might suffer from the problem of having no clear stopping point, a problem that has long been associated with the resemble-suspect strategy (Luus & Wells, 1991). Our participants at Step 1 approximated the scenario in which a detective searches a photo repository using a resemble-suspect method or asks colleagues to take a look for a quick evaluation of lineup fillers. As an example, in one lineup the eyewitness's description did not include anything about a mole. The majority of our participants, however, described the suspect as having a small mole under his eye. This raises some intriguing questions. Is the presence of a small mole significant enough to qualify as a default feature that should have been matched in the fillers? How large or small would the mole need to be to qualify versus not qualify as a feature? If the suspect is innocent, would permitting the suspect to be the only one with a mole on his face have helped the eyewitness reject the suspect ("not this person, because the guy I saw did not have a mole")? Should a mole be added electronically to each filler photo (see Wells, et al., 2020)? This entire issue of default descriptors and the criteria for including versus not including them as a basis for lineup construction needs further development.

Study 5: Lineup bias and description quality

A distinctive facial characteristic or a unique combination of attributes may be difficult for police to replicate in lineup members. In such cases, both real eyewitnesses and mock-witnesses may be able to readily select the suspect. Mock-witnesses should be particularly sensitive to distinctive characteristics, given that the culprit's description defines their deliberative task. A failure to build a lineup that encompasses distinctive perpetrator attributes can be expected to both reduce the effective size of the lineup and to increase suspect-bias. We predicted that description distinctiveness will be negatively correlated with lineup fairness.

Method and Procedure

Six coders (for each lineup, at least two plus the PI) rated eyewitness descriptions of perpetrators for 190 lineups using a coding rubric developed for the task. The lineups were not visible during the coding process. First, a total number of descriptors was tallied for each lineup. Next, a more nuanced coding of descriptors was developed to capture description distinctiveness. For example, an eyewitness who recalled "shoulder-length black hair in dreadlocks" is offering more distinctive information than "black hair." *Basic* descriptors included seven features: gender, race, age range, hair color, eye color, and indistinct statements of weight and height ("medium" or "average"). *Distinctive* descriptors were scored as one point each for more specific information ("6' tall", "200 pounds") and for mention of additional features ("clean-shaven" "dark complexion") or adjectives ("*long* black hair"). A two-point score was given for more fine-grained descriptions (e.g., specific hairstyle or facial hair, scar, tattoo). Thus, in the example above, "black hair" was coded as one point of basic description, "shoulder-length" was coded as one point of a more specific (distinctive) description, and hair style ("dreadlocks") added two more points to the distinctiveness score.

Each description score was tallied on two aspects of physical (non-clothing) features: (1) seven basic/generic descriptors (typically easy to replicate in a lineup); and (2) distinctive descriptors (more difficult to replicate) as described above. Coders were trained prior to the task. Disagreements were resolved by returning to the coding sheet for clarification and often by revising the coding system, thereby also necessitating a revisiting and clarifying the scores of all prior descriptions until agreement was 100%. A final test for interrater reliability was conducted by the PI at the time of this report preparation. The PI again coded each description and then compared these ratings with the previous group decisions. Agreement in the total *number* of descriptors was 100%. The proportion of agreement for coding categories was 92%. Instances of disagreement derived from some ambiguous witness descriptions ("taller than the other one") that did not clearly fall into a category of basic descriptor ("medium height") or a distinctive descriptor ("six-feet tall"); these disagreements were resolved with a rule for consistent classification (in this example, as "distinctive").

Results and Discussion

Total description details. The number of reported offender descriptors ranged from 2 to 13, including both physical features and items of clothing (*M* = 5.59, *SD* = 2.18, *Mdn* = 6). Only one lineup fairness measure was significantly related to the simple number of descriptors: as the number of descriptors increased, the proportion of mock-witnesses who rejected at least one lineup member also increased, *r* (*N* = 190) = .22, *p* = .003. It might be surmised that real eyewitnesses who are able to supply detailed descriptions may have better memories of the offenders, in which case a positive relationship between number of descriptors and eyewitness suspect identifications would be expected. However, eyewitness decision (suspect, filler, no pick) was not associated with total number of descriptors, *F* (2, 187) = .34, *p* = .72. Early laboratory research dispelled the notions that eyewitnesses who are good "describers" are also good "identifiers" (Wells, 1985) and that the quality of eyewitness description is a good predictor of identification accuracy—research that directly challenges the U.S. Supreme Court's 1977 position in *Manson v. Braithwaite* (Wells & Quinlivan, 2008). Outcomes of these real eyewitness cases support the lab finding.

Description distinctiveness. Each of the 190 lineups was coded for a *basic descriptor score* (*M* = 2.87, *SD* = 1.01, ranging from 0-7), a *distinctive descriptor score* (*M* = 3.17, *SD* = 2.37, range 0-10) and a *total score*, (*M* = 6.04, *SD* = 2.64, range 2-15). A significant relationship was evident only between explicit rejections and descriptor level. A higher descriptor score was associated with more mock-witnesses who rejected at least one lineup member. This relationship held for *basic* descriptors (*r* = .19, *p* = .01), *distinctive* descriptors (*r* = .22, *p* = .002) and for *total* descriptor score (*r* = .27, *p* < .0001). The proportion of mock-witnesses who explicitly rejected at least one lineup member was significantly higher when both basic and distinctive descriptor scores were at or above the medians (*Mdn* = 3) (*M* = .46, *SD* = .14, CI: .43 to .50, n = 69), compared to lineups with both basic and distinctive descriptors below median (*M* = .30 *SD* = .13, CI: .26 to .35, *n* = 31) or just one score above median (*M* = .41, *SD* = .14, CI: .39-.45, n = 88), *F* (2, 185) = 13.28, *p* < .0001, *n* = 187). The three levels of descriptor scores (as defined above, based on median splits) were not associated with eyewitness decisions, X^2 (4, *N* = 190) = 5.47, *p* = .24.

In summary, these data support the relevance of lineup description for the mock-witness decision and, by extension, to the decision process of real eyewitnesses who have limited memory of the offender beyond their recalled description. The measure of explicit rejections was informative in this context. Mock-witnesses were sensitive to the level of descriptors, seemingly using the perpetrator description as a basis for eliminating lineup members from consideration. However, there was no significant relationship between distinctive descriptors and real eyewitness decisions. This finding perhaps speaks to a potency of recognition memory that surpasses a witness's ability to describe an offender.

Three caveats to these findings should be noted. First, most lineups at face value appear to be fairly constructed with respect to basic offender features (gender, race, age) so a smaller effect may be found here compared to a set of more poorly constructed lineups. Second, our descriptor coding system may have not effectively measured the variable of distinctiveness. Finally, the descriptions from police records may not have effectively captured the witness's memory at the time of the crime especially for distinctive features. Whatever the case, the core requirement that lineups be structured around the witness's description of the perpetrator remains essential to both the mock-witness task and to the assessment of lineup fairness in real field lineups.

General Discussion

This analysis of field lineups—the most complete set of data available to date—prompts recommendations for lineup quality measurement using the mock-witness procedure and informs the discussion of effective police procedures for collection of eyewitness evidence. The research demonstrates the realities and nuances of field lineup structure that differ from the laboratory and that necessitate an updated and blended approach to lineup construction and reconsideration of the use of mock-witness measures.

Prevalence of Lineup Bias in Field Lineups

Five field studies revealed a substantial level of lineup bias against suspects. Mock-witnesses were able to detect the suspect at rates significantly higher than chance in 41% of lineups. Moreover, aggregate lineup fairness scores masked significant bias among the lineups. One of the important contributions of this research is the way it underscores the fact that the average proportion of mock-witnesses who pick the suspect is not necessarily sound evidence of fair lineup construction. This leads to a straightforward recommendation. Meaningful reporting of aggregate lineup fairness measures must go beyond a static point, that is, must include range and standard deviation, along with proportions of reverse- and suspect-biased lineups.

Our approach of classifying individual lineups as unbiased, suspect biased, and reverse-biased revealed nuanced and critical information for a practical appreciation of field lineups. A reverse-biased lineup (bias away from the suspect) cannot cancel out a totally different lineup that is biased against the suspect. After all, a police department that uses a lineup in which a filler is a better match than is the suspect in one case (e.g., a robbery) does nothing to protect an innocent suspect in a totally different case (e.g., a sexual assault). Moreover, as we noted earlier, the distribution of proportion scores is typically skewed, and among biased lineups twice as many (two-thirds) are biased against the suspect than against a filler.

A faulty assumption about lineup fairness will undercut other conclusions derived from that assumption. Fairness of an individual lineup might be erroneously construed from the aggregate, a concern voiced by Sauer, et al. (2019). Previous reports of unbiased lineups in Steblay (2011) and Wixted et al. (2015) are directly challenged by our new analysis and corrected here. And, given the level of lineup bias in these five studies, the use of witness confidence to inform researchers, investigators, and triers-of-fact regarding eyewitness identification accuracy in the field may be premature (Sauer, et al., 2019; Wixted & Wells, 2017).

Lineup Bias as a Predictor of Individual Eyewitness Behavior

Mock-witness measures did not effectively predict eyewitness decisions (suspect vs. filler vs. non-identification), an indication of poor criterion validity. Several reasons were evident for this failure of fairness scores to align with eyewitness decisions. Mock-witnesses must rely on eyewitness descriptions of the offenders, and these descriptions were sparse for many lineups. In some instances, the police suspect did not match the description. Mock-witness selections were frequently siphoned off to fillers that met the description better than did the suspect (“super-fillers” as discussed by Lee and Penrod, 2019). Perhaps some investigators somehow over-compensated in their filler selections in order to meet expectations for unbiased lineup construction in these field studies. However, the most sensible reason for the disparity between mock-witness and eyewitness decisions is that real eyewitnesses have some memory of the offender that will drive the identification decision in ways that are different from mock-witnesses. This was apparent in the fact that very few real eyewitnesses chose the filler who had the highest mock-witness proportion score. A filler-biased lineup should not reduce guilty culprit picks from real eyewitnesses who have a memory of the culprit because the culprit is the best match to holistic memory even if a filler is a better match to the description. Real eyewitnesses can have memory strength that allows holistic recognition whereas mock-witnesses use a deliberative strategy of reasoning to make a lineup selection (much like a witness with limited memory).

Although eyewitness behavior was not predictable from mock-witness measures overall, a subset of the data—simultaneous lineups—supported our hypothesis that suspect identifications would stem disproportionately from suspect-biased lineups. Also as predicted, eyewitnesses who viewed sequential lineups (half of the lineups) were seemingly less influenced by structural bias than eyewitnesses who viewed a simultaneous display. It is important to note that the number of suspect identifications from the two procedures was the same; the significant difference was in the quality of lineup from which suspect identifications were made. These findings support the early lab findings of Lindsay, et al. (1991) that sequential lineups are most beneficial when lineups are structurally biased and of later reports of sequential advantage for accuracy when using biased lineups (Carlson et. al., 2008; Clark, et al., 2008; Lee & Penrod, 2019) and when suggestive behaviors of lineup administrators are in play (Kovera, & Evelo, 2017). Lindsay, et al., (1991) claimed that the sequential lineup may protect against bias when eyewitnesses are unaware of the number of lineup members to be presented and allowed to view each lineup member only once. It is important to note that a sequential lineup's protection against bias has limits. A sequential lineup that is not back-loaded or that allows the witness a second lap through the lineup may undermine this protection. Current sequential lineup policies typically allow a second lap through the lineup at the witness's request and require that the witness's decisions during both laps be recorded. The use of caution in lapped procedures is underscored by these current data. We also do not suggest that an egregiously biased lineup construction (e.g., with a single lineup member who possesses a key attribute) can be ameliorated by a sequential presentation.

Policy-makers should take note of a recent examination of sequential and simultaneous lineup performance using a new signal detection theory-based framework (Lee & Penrod, 2019). These researchers concluded that a preference for one or the other lineup format will depend on the bias levels in actual lineups as well as the weights given by policymakers to various outcomes. The high level of bias exhibited in the current set of lineups is germane to policy decisions.

Fairness Measures Applied to Field Lineups

The mock-witness procedure represents the worst-case scenario for eyewitness evidence—a witness who has no memory whatsoever of the culprit. Lineup fairness scores reflect a decision process that occurs in the absence of memory. Our mock-witnesses did as we expect; they eliminated non-contenders from the lineup (evidenced by explicit rejections of lineup members), and they relied heavily on the witness's description to find the best-fitting lineup member. A mock-witness procedure can provide useful evidence regarding the core element of lineup fairness, that is, whether the suspect stands out from other lineup members and so can be spotted in the lineup by a person without a clear memory of the offender. We also found support for convergent validity within measures of lineup bias (proportion score, functional size, defendant bias, mean rank) and discriminant validity between size and bias measures. This echoes the lab study of Mansour, et al. (2016).

The current study alerts us to unique circumstances in the field that challenge assessment via a mock-witness strategy. Twenty percent of field lineups in the *Wells* data were significantly biased away from the suspect (reverse-biased), lineups in which the suspect failed to pull a fair share of mock-witness selections. A majority of lineups included a filler who drew more mock-witness picks than did the suspect. These scenarios are quite unlike our lab lineups. Future assessment of field lineup fairness must include a measure indicative of specific risk to the suspect and that can alert us to reverse-bias. We recommend proportion score as the primary mock-witness measure. All lineup bias measures (functional size, mean rank, defendant bias) aligned significantly with lineup fairness categories based on proportion score, and these lineup bias measures accounted for the majority of variance in mock-witness responses. Measures oriented to lineup size (effective size, explicit rejections) cannot signal when a suspect stands out nor discern bias toward versus away from the suspect. An effective size measure alone (as is often reported in literature) will miss the greater contribution of lineup bias in mock-witness reports and is less sensitive to eyewitness descriptors that make a lineup vulnerable to biased structure.

More than a third (37%) of 494 lineup records in the Wells, et al., (2015) field experiment reported no description of the offender. Crime suspects are developed from many sources, for example, gang affiliation, similar crime type, possession of stolen goods, a citizen tip-line, or by strong resemblance to a composite sketch or a figure on surveillance tape (Wixted & Wells. 2017). In a substantial percentage of cases, then, a meaningful description of the culprit is unavailable or police will structure the lineup using a resemble-suspect strategy. The match-to-description that underlies the mock-witness method will not suffice to assess lineup fairness for these lineups. A challenge for future research is to develop a mock-witness procedure that employs a description of the suspect derived from non-witness sources. A framework-descriptor system (Study 4) may be the starting point for such a mock-witness test, as it also may be for an effective means to build a lineup using a resemble-suspect method.

Implications for Lineup Practice: Uncertainty Reduction

What do our findings mean in the context of real lineup practice? Prior to the lineup, the eyewitness evidence consists of only a verbal description of the offender. Yet, an eyewitness to a crime may have a useful holistic image of the offender’s face in memory, a basis for recognition that goes beyond the level of the reported description. Luus and Wells (1991) explained that a fair lineup can allow police to learn something more from the witness and specifically to test the police hypothesis that the suspect is the guilty perpetrator. Wells and Bradfield (1999) refer to this as the uncertainty-reduction function of a lineup. An eyewitness’s suspect selection from a fair lineup may increase investigator certainty about the offender’s identity, as with the 44% of suspect identifications in the *Wells* dataset. Conversely, uncertainty lingers for the 35% of suspect picks from suspect-biased lineups because there is an alternative reason for the identification: the suspect could be detected in a biased lineup even in the absence of eyewitness memory.

Does reverse-bias reduce investigator uncertainty about the identity of the culprit? At first blush, a reasonable assumption is that a witness who makes a suspect identification in the face of a reverse-biased lineup (21% of suspect identifications in *Wells*) should strengthen the evidence against the suspect. In some cases, this may be so. The eyewitness's decision has confirmed the police hypothesis about the suspect-as-perpetrator. The eyewitness has avoided the fillers and selected the suspect despite someone else better matching the description. The suspect's image presumably triggered memory for a facial configuration that was apparent to the eyewitness but not to mock-witnesses. That is, recognition memory won out over the witness's earlier vague or incorrect description. Study 5 also brings out the point that recognizing someone is not the same task as describing someone. This is how the uncertainty-reduction feature of a lineup is supposed to work. Indeed, a reverse-biased lineup may be tantamount to a highly-protective lineup for a suspect (the suspect is receiving mock-witness picks well below chance), not a bad thing.

This situation may be a bit more complicated, however. Fair lineup structure controls for simple recall of descriptors by ensuring that all lineup members meet the eyewitness's verbal description to a similar extent. A reverse-biased lineup indicates that not all lineup members met the description of the culprit, a violation of fair lineup construction; the lineup is not pristine. A suspect selection from a reverse-biased lineup is likely coming from a lineup of truncated size. Moreover, mock-witness measures tell us only about the product of the lineup construction, not about the process itself or about why the suspect does or does not stand out. A description noted in a police report may not accurately reflect the full description provided by the witness. In real cases, lineup construction quality is vulnerable to inconsistent information gathered from multiple eyewitnesses or to changes in a culprit's appearance (e.g., hair style or color; facial hair, age) that render a suspect photo different from the eyewitness's description. A lineup constructed on the basis of one witness's memory may be inadequate for other witnesses who have divergent recollections of offender attributes. Furthermore, police may arrive at a suspect based on other leads and build the lineup with that suspect in mind rather than the eyewitness's description. These problems muddy interpretation of any lineup outcome, but are especially relevant for reverse-biased lineups.

Consider one of our lineups that garnered a mock-witness proportion score of zero for the suspect, that is, a reverse-biased lineup. (Details here are edited for brevity and confidentiality.) The eyewitness's description of the offender was "white female, light brown or blonde hair," yet the suspect in the lineup had dark brown hair. This can happen when police find a suspect that does not quite match the witness's description–or if suspect appearance has changed–but the police still use the description to locate fillers. The logic of match-to-description was undercut in a circuitous manner. The lineup fillers indeed included white females with blonde or brown hair, but reverse-bias was incurred: two fillers with blonde hair drew most of the mock-witness picks and no mock-witness chose the suspect. The difficulty is that an eyewitness's choice of the suspect under these conditions would present a troublesome inconsistency between appearance of the suspect versus the earlier description. A lineup rejection may be evidence of suspect innocence, of eyewitness memory failure, or due to the difficulty of recognizing a guilty offender who has changed appearance. A pick of the favored filler may increase uncertainty about the guilt of the suspect or impugn the reliability of the witness who has made a known error. Whatever the eyewitness's decision, this lineup does little to reduce uncertainty about the identity of the culprit. Mostly, a reverse-biased lineup can signal that something went wrong in the construction of the lineup. Unfortunately, we have no basis in this data set for resolving questions about what may have happened between the initial interview with a witness and the lineup construction. A solution for future lineups is to videotape the pre-lineup interviews that elicit the descriptions. This is a new and highly detailed recommendation of the Wells, et al. (2020) white paper.

How might this have gone better? Note that the mismatch in hair color was known before the lineup procedure. Wells, et al. (2020) advise that in such cases, the fillers should match the suspect's lineup appearance (resemble-suspect), not the witness's description. However, there is an additional problem with the lineup. A framework description built for this lineup through Study 4 included age: "white female, 40-50 years of age, brown hair." Suspect *age* was particularly salient in a lineup with four noticeably younger members, and with this descriptor included, 43% of mock-witnesses now chose the suspect. This example illustrates three core points. First, the initial mock-witness measure (of a reverse-biased lineup) was misleading if the measure is taken to mean that the lineup was well-constructed. Lee & Penrod (2019) noted that the situation in which a filler resembles the perpetrator more than does the suspect may not be rare in the real world. This speculation is supported in these field data. Second, offender age may well be a default value, an attribute that the eyewitness noticed, but just did not mention. Third, such problems may have been avoided with a more thorough eyewitness interview up-front and/or detected through a reasonable resemble-suspect screening of the lineup prior to use. The Wells, et al. (2020) white paper includes a pertinent recommendation for a video-recorded prelineup interview with the eyewitness as soon after the crime as practicable as a means to facilitate, in part, productive collection of an offender description.

The Blended Approach to Lineup Construction

The recent white paper of eyewitness identification recommendations (Wells, et al. 2020) prescribes a blend of match-to-description and resemble-suspect strategies for selecting lineup fillers, a notion of a hybrid approach to lineup construction that is not entirely new (see e.g., Wells, et al., 1998; Brigham, et al., 1999). The current field data underscore recent calls for such a comprehensive strategy (e.g., Wells & Wixted, 2017; Wells, et al., 2020).

The first step of this blended approach is collection of accurate and complete descriptors of the offender. A minimal requirement is that all fillers meet the description that the eyewitness gave of the culprit. Wells, et al. (2020) recommend that investigators avoid facial feature checklists, as such lists may subtly pressure a witness to complete the list with less accurate details and best guesses. Instead, the prompting of open-ended responses followed by specific (but non-leading) probes about appearance details (e.g., height, build, age, sex, complexion) and any distinguishing characteristics is useful. The witness can be prompted to provide general, coarse-grained information first (Brewer, et al., 2018) that may address what may otherwise become omitted default values. The Person Description Interview (PDI; Demarchi & Py, 2009) similarly requires the witness to provide general information prior to moving to specific features of the face. The PDI also moves the witness's description of the offender's face from lower region (chin, lips) to upper region (hair, eyes). A recent revision of the Cognitive Interview that focuses on perpetrator-related descriptors shows promise in eliciting increased and accurate details from eyewitnesses in a 20-minute interview (Satin & Fisher, 2019). Gabbert, et al. (2009) also have developed a self-administered interview tool to generate more correct details than free recall alone.

As these techniques illustrate, police investigators can control the depth and direction of descriptions provided by eyewitnesses. Truncation of eyewitness information may result from the urgency of a crime scene, shortcuts in the process, investigator assumptions about the likely offender that preempt effective questioning, and/or a description from the witness that is otherwise filtered through the lens of the interviewing officer. For example, an investigator who already has a lead from another source may short-change the interview process for a given witness to the potential detriment of the investigation's later identification procedures. Wells, et al. (2020) also underscore the importance of videotaping the eyewitness interview to provide a record for the investigation and triers of fact. At the least, then, framework descriptors should be elicited from a eyewitness and used to draw fillers from a photo repository, with fillers then selected to match the description.

The second step is to screen the lineup and verify that the suspect does not stand out among lineup members. There is a need for an a-prior strategy for filler selection when no witness description is available. In either case (eyewitness description available or not), an intentional lineup screening seems in order (e.g., by another investigator) prior to administration of the lineup to the witness. Match-to-description *and* resemble-suspect procedures are complementary rather than competing versions of lineup construction.

Conclusion.

Field data provide a useful supplement to laboratory knowledge. These field studies demonstrate that police lineups are prone to biased structure and that current measures are frequently inadequate to detect poor structure. Mock-witness measures of lineup bias (e.g., proportion score) can signal unfair bias against a suspect, yet do not readily predict eyewitness decisions. One reason for this lack of predictive power in the current data is that witnesses who viewed sequential lineups appear to be less influenced by structural lineup bias than witnesses who viewed simultaneous lineups. A second reason is that incomplete or inadequate descriptors are often used in lineup construction. The quality of lineup structure depends on an effective description provided by a witness under a match-to-description strategy and a rigorous resemble-suspect strategy to develop and screen a lineup prior to the identification procedure. Further research on a blended approach to lineup construction is recommended.

References

Behrman, B.W., & Davey, S.L. (2001). Eyewitness identification in actual criminal cases: An archival analysis. *Law and Human Behavior, 25*, 475–491. https://doi.org/10.1023/A:1012840831846

Brewer, N., Vagadia, A.N., Hope L., & Gabbert, F. (2018). Interviewing witnesses: Eliciting coarse-grain information. *Law and Human Behavior 42*(5), 458-471. https://doi: 10.1037/lhb0000294

Carlson, C.A., Gronlund, S.D., & Clark, S.E. (2008). Lineup composition, suspect position, and sequential lineup advantage. *Journal of Experimental Psychology: Applied*, *14*(2), 118-128. https://doi: 10.1037/1076-898X.14.2.118

Clark, S.E., Howell, R.T., & Davey, S.L. (2008). Regularities in eyewitness identification. *Law and Human Behavior, 32*(3), 187-218. http://doi:10.1007/s10979-006-9082-4

Doob, A.N., & Kirshenbaum, H.M. (1973). Bias in police lineups—partial remembering. *Journal of Police Science and Administration*, *1*, 287-293. https:// doi:10.1016/0022-1031(73)90062-0

Demarchi, S., & Py, J. (2009). A method to enhance person description: A field study. In R. Bull, T. Valentine, & T. Williamson (Eds.), *Handbook of psychology of investigative interviewing: Current developments and future directions* (p. 241–256). Wiley-Blackwell. https://doi.org/10.1002/9780470747599.ch14

Fitzgerald, R. J., Price, H. L., Oriet, C., & Charman, S. D. (2013). The effect of suspect-filler similarity on eyewitness identification decisions: A meta-analysis. *Psychology, Public Policy, and Law*, *19*(2), 151–164. https://doi.org/10.1037/a0030618

Flowe, H.D., & Humphries, J.E. (2011). An examination of criminal facial bias in a random sample of police lineups. *Applied Cognitive Psychology, 25*(2), 265-273. https://doi.org/10.1002/acp.1673

Gabbert, F., Hope, L., & Fisher, R. P. (2009). Protecting eyewitness evidence: Examining the efficacy of a self-administered interview tool. *Law and Human Behavior, 33*(4), 298–307. https://doi.org/10.1007/s10979-008-9146-8

Innocence Project. (2020). https://www.innocenceproject.org/cases/rickie-johnson/

Klobuchar, A., Steblay, N., & Caligiuri, H. (2006). Improving eyewitness identifications: Hennepin County's Blind Sequential Lineup Pilot Project. *Cardozo Public Law, Policy & Ethics Journal*, *4(2*), 381-413.

Kovera, M. B., & Evelo, A. J. (2017). The case for double-blind lineup administration. *Psychology, Public Policy, and Law, 23*(4), 421–437. https://doi.org/10.1037/law0000139

Lee, J., & Penrod, S. D. (2019). New signal detection theory-based framework for eyewitness performance in lineups. *Law and Human Behavior, 43*(5), 436–454. https://doi.org/10.1037/lhb0000343

Lindsay, R. C. L., Lea, J. A., Nosworthy, G. J., Fulford, J. A., Hector, J., LeVan, V., & Seabrook, C. (1991). Biased lineups: Sequential presentation reduces the problem. *Journal of Applied Psychology, 76*(6), 796–802. https://doi.org/10.1037/0021-9010.76.6.796

Lindsay, R. C. L., Martin, R., & Webber, L. (1994). Default values in eyewitness descriptions: A problem for the match-to-description lineup foil selection strategy. *Law and Human Behavior, 18*(5), 527–541. https://doi.org/10.1007/BF01499172

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, https://doi:10.1002/(SICI)1099-0720(199911)13:1+3.3.CO;2-O

Lindsay, R. C. L., Smith, S. M., & Pryke, S. (1999). Measures of lineup fairness: Do they postdict identification accuracy? *Applied Cognitive Psychology, 13*(Spec Issue), S93–S107. https://doi.org/10.1002/(SICI)1099-0720(199911)13:1+<S93::AID-ACP633>3.3.CO;2-O

Lindsay, R. C., Wallbridge, H., & Drennan, D. (1987). Do the clothes make the man? An exploration of the effect of lineup attire on eyewitness identification accuracy. *Canadian Journal of Behavioural Science / Revue canadienne des sciences du comportement, 19*(4), 463–478. https://doi.org/10.1037/h0079998

Lindsay, R. C. L., & Wells, G. L. (1980). What price justice? Exploring the relationship of lineup fairness to identification accuracy. *Law and Human Behavior, 4*(4), 303–313. https://doi.org/10.1007/BF01040622

Luus, C. A. E., & Wells, G. L. (1991). Eyewitness identification and the selection of distracters for lineups. *Law and Human Behavior, 15*(1), 43–57. https://doi.org/10.1007/BF01044829

Malpass, R. S. (1981). Effective size and defendant bias in eyewitness identification lineups. *Law and Human Behavior, 5*(4), 299–309. https://doi.org/10.1007/BF01044945

Malpass, R. S., & Lindsay, R. C. L. (1999). Measuring line-up fairness. *Applied Cognitive Psychology, 13* (Special Issue), S1–S7. http://doi.org/10.1002/(SICI)1099 -0720(199911)13:1+<S1::AID-ACP678>3.0.CO;2-9

Manson v. Braithwaite, 432 U.S. 98, 1977

Mansour, J. K., Beaudry, J. L., Kalmet, N., Bertrand, M. I., & Lindsay, R. C. L. (2017). Evaluating lineup fairness: Variations across methods and measures. *Law and Human Behavior, 41*(1), 103–115. http://doi.org/10.1037/lhb0000203

Police Executive Research Forum (2013). A national survey of eyewitness identification procedures in law enforcement agencies. Retrieved January 8, 2020 from http://www.policeforum.org/assets/docs/Free_Online_Documents/Eyewitness_Identification/a%20national%20survey%20of%20eyewitness%20identification%20procedures%20in%20law%20enforcement%20agencies%202013.pdf

Rodriquez, D.N., & Berry, M.A. (2014). The effect of line-up administrator blindness on the recording of eyewitness identification decisions. *Legal and Criminological Psycho*logy*19* (1), 69-79. https://doi: 10.1111/j.2044-8333.2012.02058x

Satin, G. E., & Fisher, R. P. (2019). Investigative utility of the Cognitive Interview: Describing and finding perpetrators. *Law and Human Behavior, 43*(5), 491-506. http://dx.doi.org/10.1037/lhb0000326

Sauer, J. D., Palmer, M. A., & Brewer, N. (2019). Pitfalls in using eyewitness confidence to diagnose the accuracy of an individual identification decision. *Psychology, Public Policy, and Law, 25*(3), 147–165. http://doi.org/10.1037/law0000203

Schacter, D. L., Dawes, R., Jacoby, L. L., Kahneman, D., Lempert, R., Roediger, H. L., & Rosenthal, R. (2008). Policy forum: Studying eyewitness investigations in the field. *Law and Human Behavior, 32*(1), 3–5. https://doi.org/10.1007/s10979-007-9093-9

Seale-Carlisle, T.M., Colloff, M.F., Flowe, H.D., Wells, W., Wixted, J.T., & Mickes, L. (2019). Confidence and Response Time as Indicators of Eyewitness Identification Accuracy in the Lab and in the Real World. *Journal of Applied Research in Memory and Cognition, 8*(4), 420-428. https://doi.org/10.1016/j.jarmac.2019.09.003

Smith, A. M., Wells, G. L., Smalarz, L., & Lampinen, J. M. (2018). Increasing the similarity of lineup fillers to the suspect improves applied value of lineups without improving memory performance. *Psychological Science*, *29* (9), 1548-1552. https://doi.org/10.1177/0956797617698528

Steblay, N.K. (2011). What we know now: The Evanston Illinois lineups, *Law and Human Behavior*, *35,* 1, 1-12. https://doi: 10.1007/s10979-009-9207-7

Steblay, N.K. (2018). All is not as it seems: Avoidable pitfalls in the interpretation of lineup field data. *Psychology, Public Policy, and Law*, *24*(3), 292-306. https://doi: 10.1037/law0000171

Steblay, N.K., Dysart, J. E., & Wells, G.L. (2011). Seventy-two tests of the sequential lineup superiority effect: A meta-analysis and policy discussion. *Psychology, Public Policy, and Law*, *17*(1), *99-139.* https://doi: 10.1037/a0021650

Tredoux, C. G. (1998). Statistical inference on measures of lineup fairness. *Law and Human Behavior*, *22* (2), 217–237. https://doi:10.1023/A:1025746220886

Wells, G.L. (1985). Verbal descriptions of faces from memory: Are they diagnostic of identification accuracy? *Journal of Applied Psychology*, *70*(4), 619-626. https://doi:10.1037//0021-9010.70.4.619

Wells, G. L. (1993). What do we know about eyewitness identification? *American Psychologist*, *48*(5), 553–571. https://doi.org/10.1037/0003-066X.48.5.553

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*(Spec Issue), S27–S39. https://doi.org/10.1002/(SICI)1099-0720(199911)13:1+<S27::AID-ACP635>3.3.CO;2-D

Wells, G.L., Kovera, M.B., Douglass, A. B., Brewer, N., Meissner, C., & Wixted, J. (2020).

Policy and procedure recommendations for the collection and preservation of eyewitness identification evidence. *Law and Human Behavior*, *44* (1), 3-36. https://doi.org/10.1037/lhb0000359

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. https:// doi:10.1007/BF01039807

Wells, G.L., & Quinlivan, D.S. (2008). Suggestive eyewitness identification procedures and the Supreme Court's reliability test in light of eyewitness science: 30 years later. *Law and Human Behavior, 33*(1), 1-24. https://doi: 10.1007/s10979-008-9130-3

Wells, G. L., Rydell, S. M., & Seelau, E. P. (1993). On the selection of distractors for eyewitness lineups. *Journal of Applied Psychology, 78(5)*, 835-844. https://doi:10.1037/0021-9010.78.5.835

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*, *22*(6), 603-653. https://doi:10.1023/A:1025750605807

Wells, G.L., Steblay, N.K., & Dysart, J.E. (2015). Double-blind photo-lineups using actual eyewitnesses: An experimental test of the sequential versus simultaneous lineup procedure. *Law and Human Behavior*, *39 (1)*, 1-14. http://dx.doi.org/10.1037/lhb0000096

Wells, W. (2014). The Houston Police Department Eyewitness Identification Experiment: Analysis and Results. Last viewed January 8, 2020 at http://www.lemitonline.org/research/documents/Final%20Results%20Report_HPD%20eyewitness%20experiment_Final.pdf

Wixted, J. T., Mickes, L., Dunn, J. C., Clark, S. E., & Wells, W. (2015). Estimating the reliability of eyewitness identifications from police lineups. *Proceedings of the National Academies of Science, 113(2),* 304–309. https://doi.org/10.1073/pnas.1516814112

Wixted, J. T. & Wells, G. L. (2017). The relationship between eyewitness confidence and identification accuracy: A new synthesis. *Psychological Science in the Public Interest, 18* (1) 10–65. http://doi.org/10.1177/1529100616686966

Footnotes

1. Footnote 1. A point of clarification is useful here. This analysis does not test the relationship between lineup format and lineup fairness. For such a test, it would be necessary to secure a random sample of sequential and simultaneous lineups without regard to eyewitness decision. Furthermore, because random assignment to lineup format in the Wells, et al. field study was not made until *after* the lineup was constructed, lineup format and lineup fairness are, by design, methodologically orthogonal in this data set.

2. We thank an anonymous reviewer for insights into DNA-exoneration cases that reflect the problems of distinctive features discussed here. Rickie Johnson was identified even though he had a prominent gold tooth which was never part of the victim's description of her attacker. Larry Mayes was identified from a photo array by a victim who did describe her assailant as having a gold tooth. Mayes had a gold tooth; the record is unclear as to whether the lineup fillers matched this attribute (Innocence Project, 2020).

Table 1

Mock-Witness Proportion Scores from Five Field Studies of Real Lineups (Strangers)

								Proportion of suspect selections		
								< .10	**.10 - .24**	**>.24**
Field study		Study *N*	Test (*n*)	*M*	*SD*	*Mdn*	Range	Reverse-Biased	Fair	Suspect-Biased
Wells	2015	494	190	.21	.17	.16	[.00 .79]	24% (46)	43% (82)	33% (62)
Klobuchar	2006	178	37	.29	.22	.22	[.03 .91]	16% (6)	38% (14)	46% (17)
Steblay	2011	61	19	.18	.12	.17	[.00 .64]	21% (4)	47% (9)	32% (6)
Steblay	2018	315	59	.37	.23	.36	[.02 .84]	12% (7)	20% (12)	68% (40)
Wixted*	2015*	500	60	.23	.17	.20	[.02 .92]	18% (11)	42% (25)	40% (24)
Total		1548	365					74	142	149

*Wixted, et al., (2015) examined a smaller set of 30 lineups with blind administration (setting aside 30 blinded lineups), in which 50% of lineups were fair, 27% suspect-biased, and 23% reverse-biased. Wixted et al. concluded that the lineups were unbiased overall, an assumption that was crucial for the other conclusions that they reached from these data. But, as we note in this manuscript, finding reverse bias in one lineup should does not cancel suspect bias in a totally different lineup. Hence, we disagree the broad conclusion that the Wixted lineups were unbiased.

Table 2

Characteristics of the Tested Lineups (n = 365)

Field study	Lineups	Raters	Sampling	Color/B&W	(Double)Blind/Format	Photo/Live	6-person
	tested	per lineup	strategy		Simultaneous/Sequential		lineup
Wells	190	50-78	Stratified Random*	84% color	Blind 55% SIM 45% SEQ	photo	100%
Klobuchar	37	104	not random	51% color	Blind 100% SEQ	Photo	100%
Steblay (2011)	19	78	not random	Black/White	36% Non-blind SIM 52% Blind SEQ 2% Other	Photo	95%
Steblay (2018)`	59	52-62	not random	Color	63% Non-blind SIM 37% Blind SEQ	Live	74% size 5 range 4-8
Wixted	60	49	random	Color	25% Blind SEQ 25% Blinded SEQ** 25% Blind SIM 25% Blinded SIM**	Photo	100%

*see detail in Study 2

**Blinded conditions allowed the primary investigator to administer the lineup but prevented the investigator from knowing the suspect's position in the lineup and which photo the witness was viewing.

Table 3

Descriptive statistics for Samples I and II fairness measures

	(n = 116)			*(n = 190)*		
Measure	Mean	(SD)	95%CI	Mean	(SD)	95%CI
Proportion	.20	(.17)	[.17, .23]	.21	(.17)	[.19, .23]
Functional size	10.54	(11.70)	[8.41, 12.67]	9.83	(11.07)	[8.2, 11.4]
Mean rank	3.33	(.84)	[3.18, 3.48]	3.25	(.79)	[3.14, 3.36]
Defendant bias[a]	-.08	(.16)	[-.11. -.05]	-.07	(.17)	[-.09, -.05]
Effective size	3.87	(1.02)	[3.69, 4.05]	3.91	(1.01)	[3.77, 4.08]
Explicit reject	.44	(.16)	[.41, .47]	.41	(.15)	[.39, .43]

a. A negative score indicates bias away from the suspect; positive score indicates suspect bias.

Table 4

Eyewitness Decisions by Lineup Fairness Categories (Sample II, n = 190)

	Eyewitness Decision						
Fairness category	Suspect		Filler		No pick		
Fair	34	.44	23	.49	25	.38	82
Suspect bias	27	.35	12	.25	23	.35	62
Reverse bias	16	.21	12	.26	18	.27	46
	77		47		66		

Sample II, by lineup format

	Sequential lineups						Simultaneous lineups					
	Suspect		Filler		No pick		Suspect		Filler		No pick	
Fair	15	.39	11	.61	13	.43	19	.49	12	.41	12	.33
Suspect bias	9	.24*	3	.17	8	.27	18	.46*	9	.31	15	.42
Reverse bias	14	.37*	4	.22	9	.30	2	.05*	8	.28	9	.25
Total	38		18		30		39		29		36	

* Significant differences in source of suspect selections (p < .05)

Table 5

Mock-Witness Measures by Categories of Lineup Fairness (Sample II)

	Lineup Fairness Category								
	Suspect-biased			Fair			Reverse-biased		
	(n = 62)			(n = 82)			(n = 46)		
	M	(*SD*)	[*CI*]	*M*	(*SD*)	[*CI*]	*M*	(*SD*)	[*CI*]
Proportion	.41 ^{a}	(.14)	[.38, .45]	.15 ^{b}	(.04)	[.14, .16]	.05 ^{c}	(.02)	[.05, .06]
Functional size	2.65 ^{a}	(.76)	[2.5, 2.8]	7.06^{b}	(1.81)	[6.7, 7.5]	24.45 ^{c}	(14.3)	[20.3, 28.6]
Mean rank	2.43 ^{a}	(.45)	[2.3, 2.5]	3.41 ^{b}	(.41)	[3.3, 3.5]	4.10 ^{c}	(.57)	[3.9, 4.3]
Defendant bias	.12 ^{a}	(.06)	[.10, .13]	-.10 ^{b}	(.09)	[-.12, -.08]	-.25 ^{c}	(.13)	[-.29, -.21]
Effective size	3.59 ^{a}	(1.01)	[3.4, 3.8]	4.24 ^{b}	(.96)	[4.0, 4.5]	3.75 ^{a}	(1.08)	[3.5, 4.1]
Explicit rejections	.43	(.14)	[.40, .47]	.39	(.14)	[.36, .42]	.44	(.17)	[.39, .49]

Different subscripts in rows indicate significant differences between groups

Table 6

Study 3: Lineup Fairness Measures by Original vs. Limited Description for 20 Highly – Biased Lineups

	Limited description	Original description		
	M (SD)	*M (SD)*	*t* (38)	*d*
Proportion score	.18 (.11)	.46 (.17)	6.41*	2.08
Functional size	9.39 (8.71)	2.48 (1.10)	3.52*	1.14
Effective size	4.87 (.84)	3.16 (1.08)	5.73*	1.86
Mean Rank	3.39 (.59)	2.53 (1.0)	3.31*	1.07
Explicit rejections	.33 (.13)	.49 (.16)	5.63*	1.83

* *p* < .05

Table 7

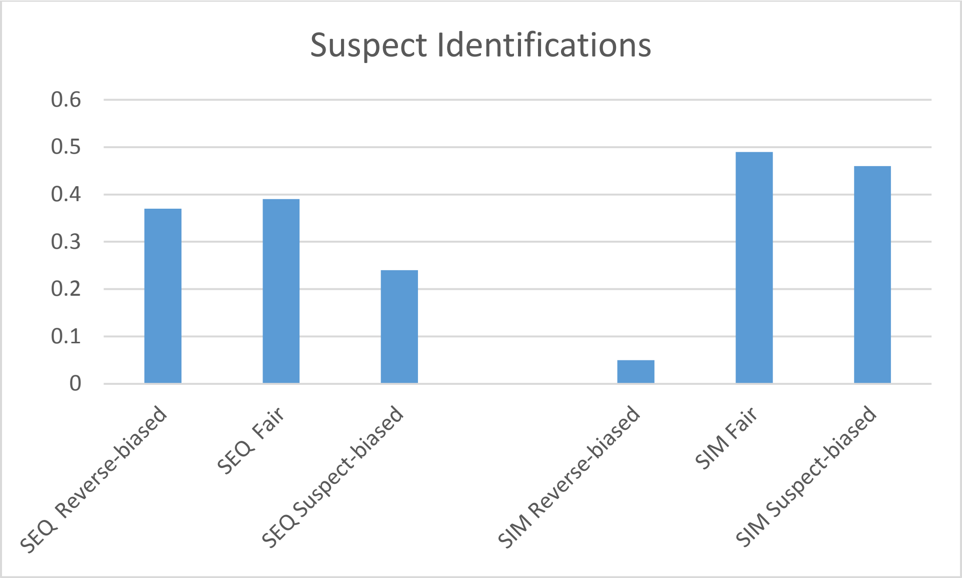
Study 4: Lineup Measures: Original vs. Framework Descriptions (n = 19)

	Original	Framework		
	M (SD)	*M (SD)*	*t (36)*	*d*
Proportion score	.11 (.07)	43 (.19)	6.89 *	2.30
Functional size.	14.78 (11.90)	3.04 (1.95)	4.25 *	1.42
Effective size	4.07 (.91)	3.38 (1.20)	2.01 (p = .05)	.67
Mean rank	3.8 (.60)	2.49 (.57)	7.00 *	2.33
Explicit rejections	.38 (.14)	.54 (.15)	3.20 *	1.07

* *p* < .05

Figure 1

Suspect selections by lineup format



Appendix 1

Summary of Eyewitness Decisions from Wells, et al. 2015,
and current study Sample 1 and Sample II

	Lineup format					
	Simultaneous		Sequential		Combined	
Eyewitness decision						
	Wells, Steblay, & Dysart, 2015 (N = 494)					
Suspect	67	(26.0 %)	65	(27.5%)	132	26.7%
Filler	46	(17.8%)	29	(12.3%)	75	15.2%
No ID	145	(56.2%)	142	(60.2%)	287	58.1%
Total	*258*		*236*		*494*	
	Sample I (n = 116)					
Suspect	20	(34.5 %)	20	(34.5%)	40	34.5%
Filler	18	(31.0%)	18	(31.0%)	36	32.0%
No ID	20	(34.5%)	20	(34.5%)	40	34.5%
Total	*58*		*58*		*116*	
	Sample II (n = 190)					
Suspect	39	(37.5 %)	38	(44.2%)	77	40.5%
Filler	29	(27.9%)	18	(20.9%)	47	24.7%
No ID	36	(34.6%)	30	(34.9%)	66	34.7%
Total	*104*		*86*		*190*	

Appendix 2

Factor analysis of lineup fairness measures

Correlation matrix (Bold = significant, p < .05)

	Prop	FS	Rank	Def Bias	ES	Exp. rejects
Proportion Score		**-.60**	**-.84**	**-.82**	**-.32**	.10
Functional Size			**.67**	**-.70**	-.14	.10
Mean rank				**-.67**	**.33**	- .07
Def bias					**.23**	-.13
Effective size						**-.36**

Factor loadings

	Component 1	Component 2	
Proportion score	-.92	-.26	
Functional size	.83	-.22	
Mean rank	.89	.27	
Defendant bias	-.92	.25	
Effective size	.06	.89	
Explicit rejections	.05	-.73	
% Variability	52%	26%	Total 78%