Law and Human Behavior
2015, Vol. 39, No. 1, 1–14

© 2014 American Psychological Association
0147-7307/15/$12.00 DOI: 10.1037/lhb0000096

Double-Blind Photo Lineups Using Actual Eyewitnesses: An Experimental Test of a Sequential Versus Simultaneous Lineup Procedure

Gary L. Wells
Iowa State University

Nancy K. Steblay
Augsburg College

Jennifer E. Dysart
John Jay College of Criminal Justice

Eyewitnesses (494) to actual crimes in 4 police jurisdictions were randomly assigned to view simultaneous or sequential photo lineups using laptop computers and double-blind administration. The sequential procedure used in the field experiment mimicked how it is conducted in actual practice (e.g., using a continuation rule, witness does not know how many photos are to be viewed, witnesses resolve any multiple identifications), which is not how most lab experiments have tested the sequential lineup. No significant differences emerged in rates of identifying lineup suspects (25% overall) but the sequential procedure produced a significantly lower rate (11%) of identifying known-innocent lineup fillers than did the simultaneous procedure (18%). The simultaneous/sequential pattern did not significantly interact with estimator variables and no lineup-position effects were observed for either the simultaneous or sequential procedures. Rates of nonidentification were not significantly different for simultaneous and sequential but nonidentifiers from the sequential procedure were more likely to use the "not sure" response option than were nonidentifiers from the simultaneous procedure. Among witnesses who made an identification, 36% (41% of simultaneous and 32% of sequential) identified a known-innocent filler rather than a suspect, indicating that eyewitness performance overall was very poor. The results suggest that the sequential procedure that is used in the field reduces the identification of known-innocent fillers, but the differences are relatively small.

Keywords: eyewitness identification, sequential lineups, eyewitness error, eyewitness field experiments, lineup position effects

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

There is a growing awareness in the general American population that eyewitness identification evidence can be unreliable. Much of this awareness can be traced to media treatments of forensic DNA exonerations, for which over 75% of the first 250 were cases of mistaken eyewitness identification (Garrett, 2011). Best-selling books such as *Picking Cotton* (Thompson-Canino, Cotton, & Torneo, 2010), Jennifer Thompson-Canino, Ronald Cotton, and Erin Torneo's story of mistaken identification and redemption, have also had an impact. TV programs, such as CBS's *60 Minutes* and MSNBC's *Rock Center with Brian Williams*, have managed to not only present compelling personal stories of mistaken identification and wrongful conviction but also highlight the psychological science that attempts to understand and control eyewitness error. Entire states, such as New Jersey, Connecticut, Texas, Florida, and North Carolina, as well as individual law enforcement agencies, such as Boston, Minneapolis, St. Paul, Dallas, and Denver, have made procedural reforms in how they conduct lineups based on laboratory findings of researchers. In 2012, one of the most comprehensive treatments of eyewitness identification by any court was conducted by the Oregon Supreme Court in an attempt to blend the extant scientific literature with the law (*Oregon v. Lawson*, 2012).

In the legal system, the problems with eyewitness identification carry a nascent aura, as though they were only recently discovered. Eyewitness identification scholars know otherwise, as the issues predate even Munsterberg (1908). In fact, this journal (*Law and Human Behavior*) published the first issue of a scholarly journal devoted solely to eyewitness identification 33 years ago (Wells, 1980). The empirical literature has exploded since that time, much of it focused on what can be done to improve the reliability of eyewitness evidence, an approach commonly known as the system-variable perspective (Wells, 1978).

This article was published Online First June 16, 2014.

Gary L. Wells, Psychology Department, Iowa State University; Nancy K. Steblay, Psychology Department, Augsburg College; Jennifer E. Dysart, Psychology Department, John Jay College of Criminal Justice.

This research was supported by grants from the Laura and John Arnold Foundation, the Open Society Foundations, the JEHT Foundation, and the National Institute of Justice. The American Judicature Society, the Innocence Project, and the Police Foundation were key partners in coordinating this project. We thank the Austin (Texas) Police Department, the Travis County (Texas) District Attorney's Office, the Charlotte-Mecklenburg (North Carolina) Police Department, the Mecklenburg County (North Carolina) District Attorney's Office, the Tucson (Arizona) Police Department, the Pima County (Arizona) Attorney's Office, the San Diego (California) Police Department, and the San Diego County (California) District Attorney's Office for their cooperation and efforts in collecting these data.

Correspondence concerning this article should be addressed to Gary L. Wells, Psychology Department, Iowa State University, West 112 Lagomarcino Hall, Ames, IA 50021. E-mail: glwells@iastate.edu

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

Assisted by four police departments and various organizations, we conducted a field experiment examining what is perhaps the most controversial of all the system-variable reforms, namely the sequential lineup. There are many commentaries, reviews, and meta-analyses of the sequential lineup versus the traditional simultaneous lineup based on data from laboratory experiments (e.g., Clark, 2012; Steblay, Dysart, & Wells, 2011; Wells, Steblay, & Dysart, 2012). A sequential procedure is one in which the witness views lineup members one at a time and makes a decision on each before seeing the next. This contrasts with a simultaneous procedure in which all lineup members are available to be viewed at the same time. The sequential procedure was devised originally to prevent witnesses from making "relative judgments" in which witnesses compare lineup members with each other and then show a propensity to identify the person who looks most like their memory of the culprit *relative* to the other members of the lineup (Lindsay & Wells, 1985; Wells, 1984). Although a relative-judgment process is likely to work well if the actual culprit is in the lineup, relative judgments can lead to mistaken identifications of innocent people when the culprit is not present in the lineup. The sequential procedure was proposed as a method to encourage witnesses to engage in a more *absolute* process in which the witnesses compare each lineup member with their memory. Meta-analyses of lab data generally show that a sequential procedure reduces the chances of mistaken identification when the culprit is not in the lineup but also produces fewer identifications of the culprit when the culprit is in the lineup. Controversy about the simultaneous versus sequential procedure stems largely from this trade-off of accurate and mistaken identifications. Despite some trade-off of accurate identifications for fewer mistaken identifications, the diagnosticity of positive identifications (a ratio of accurate to mistaken identifications) is higher for a sequential procedure. Hence, when a positive identification is made with the sequential procedure, it can be better trusted to have been accurate. In support of this, a meta-analysis by Steblay, Dysart, and Wells (2011) showed diagnosticity ratios of 8.3 and 5.8 for sequential and simultaneous lineups, respectively. Likewise, Clark (2012) reported data on accurate and mistaken identifications collapsed across 51 studies that indicate diagnosticity ratios of 4.8 and 3.6 for sequential and simultaneous lineups, respectively.

Palmer and Brewer (2012) analyzed a large number of simultaneous-sequential experiments using a Signal Detection Theory approach. Palmer and Brewer found no differences in discriminability (ability to detect differences between a guilty and innocent lineup member) and instead noted that the sequential lineup appears to achieve the better diagnosticity ratio because it moves the witness from a generally sloppy decision criterion to a more conservative criterion. The Palmer and Brewer results are consistent with the idea that relative judgments are a loose criteria and that the sequential might reduce reliance on such loose criteria. Using receiver operating characteristics curves, however, some recent experiments produced results suggesting that discriminability might be higher for the simultaneous procedure than for the sequential procedure (Gronlund et al., 2012; Mickes, Flowe, & Wixted, 2012).

Purposes and Characteristics of the Current Field Experiment

It is not our purpose to rehash the laboratory data and interpretations of the simultaneous versus sequential issue. In fact, the current field experiment cannot provide a definitive test of the psychological processes that distinguish simultaneous and sequential procedures. Instead, the current field experiment addresses the question of what happens when actual witnesses to serious crimes view lineups using a simultaneous procedure versus a sequential procedure that is actually used in the field (which is not exactly the procedure used in all lab studies). This is an important issue because a large number of law enforcement agencies in the U.S. now use a sequential lineup procedure (e.g., the entire states of Connecticut, North Carolina, and New Jersey as well as major cities such as Dallas and Boston). A 2011 survey of law enforcement agencies in the U.S. indicated that 32% are using a sequential procedure (Police Executive Research Forum, 2013).

A second purpose of this field experiment was to collect the first data with eyewitnesses to actual crimes using procedures that are double-blind, guarantee uniformity of prelineup instructions (via automation), randomize the position of the suspect in the lineups, and guarantee that every witness response is fully documented. The double-blind aspect of the data is very important because nonblind administrators can influence who the witness identifies (e.g., Greathouse & Kovera, 2009). In addition, nonblind administration can bias documentation of the results. For example, researchers using archival methods have noted the failure of lineup administrators' reports to document filler identifications (e.g., Behrman & Davey, 2001; Tollestrup, Turtle, & Yuille, 1994). Documentation of the witnesses' responses, of course, should be independent of whether they identified the suspect or identified a filler. And yet, as shown recently in controlled experiments, nonblind administrators tend to faithfully report all suspect picks but not filler picks whereas double-blind administrators faithfully report both suspect and filler picks (Rodriguez & Berry, 2014). Moreover, a recent national survey of U.S. law enforcement agencies indicated that 37% of the agencies do not prepare lineup reports if the witness does not pick the suspect (Police Executive Research Forum, 2013). Hence, accurate records of suspect picks, filler picks, and nonidentifications are particularly problematic for field studies that use archival methods (going back through case files) and for other field studies that did not involve nonblind administration.

Obviously, one of the great advantages of lab-based eyewitness identification experiments is that "ground truth" is established in the sense that it is known whether the lineup's suspect is the culprit or an innocent person. In actual cases, in contrast, we cannot be sure that the identification of a suspect is accurate. However, the current experiment used methods to ensure that every lineup contained only one possible suspect and the remaining lineup members were known-innocent fillers (something that archival studies cannot guarantee). Hence, whereas we cannot know with certainty that an identification of a *suspect* was accurate in this field experiment, we can safely assume that the identification of a *filler* was a mistake. Moreover, by using a methodology of true random assignment of each lineup to simultaneous versus sequential procedures, we know that the probability that the suspect is in fact the guilty party is equivalent for the two procedures.

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

We conducted this field experiment using the elements of a sequential that are used in the jurisdictions that have implemented sequential lineup procedures. This included back-loading, using the continuation rule, permitting witnesses to resolve their own multiple picks, permitting a second sequential viewing if the witness requested it, and permitting a "not sure" option. We briefly discuss each of these elements.

Back-Loading

Back-loading is an aspect of the sequential procedure in which the witness does not know how many individuals will be viewed in the lineup. Most lab studies, although not all, have used back-loading. Back-loading is standard practice in jurisdictions using a sequential procedure and was used in the current field experiment. In fact, back-loading was a central characteristic described as an important part of the sequential procedure from the outset and was used in the original simultaneous versus sequential experiment (Lindsay & Wells, 1985). Moreover, as early as 1991 it was shown that the sequential procedure works better if witnesses do not know how many faces are to be viewed (Lindsay, Lea, & Fulford, 1991). Finally, lab data indicate that back-loading eliminates order effects (i.e., influence of the suspect's position) in the sequential lineup (Horry, Palmer, & Brewer, 2012).

Continuation Rule

Some lab studies have used a "stopping rule" with the sequential lineup, such that the lineup ends when the witness makes an identification even if there are more photos in the lineup. But jurisdictions that have adopted the sequential procedure do not use a stopping rule and instead use a continuation rule that requires the witness to view the remaining photos even if they identified someone earlier in the sequence. The rationale for the continuation rule appears to be a sound one. First, suppose a witness chose the first (or second) person in the sequence and that ended the procedure. If it was the suspect who was chosen, the defense would ridicule the police for having shown the witness only one or two photos. Alternatively, suppose that the lineup member chosen in the first or second position happened to be a filler and the procedure were stopped without the witness ever seeing the photo of the suspect. In that case, the case detective would be understandably upset that she or he never got to see how the witness would have reacted to seeing the suspect. In fact, the police and prosecutors who cooperated in this experiment would have never have allowed us to conduct this study using a stopping rule. Accordingly, as with other jurisdictions that use a sequential procedure, a continuation rule was used in the current experiment. Moreover, in order to prevent witnesses from inferring that continuation might mean that they identified the wrong person, witnesses in the sequential condition were told in prelineup instructions that the procedure required them to continue to view all of the photos even if they identified someone.

Witnesses Resolve Their Own Multiple Picks

Witnesses sometimes pick more than one person from a lineup, even in simultaneous procedures. In actual investigations, it is up to the witness to resolve multiple picks. Lab studies using a sequential procedure, however, have used arbitrary rules to score who the witness identified. Usually, lab studies count only the first pick in a sequential procedure. For example, Gronlund, Carlson, Dailey, and Goodsell (2009) told witnesses that only their first "yes" would count. Other studies have counted only the last pick. For good reasons, neither of these arbitrary rules is used in actual practice with sequential lineups. Imagine an actual case in which a witness makes a tentative pick early in the sequence and then on a later photo says "This is the guy, not the other one." If this second person is the suspect, no jurisdiction is going to say "Too late. We are going to ignore your identification of the suspect." As with actual practices, in this experiment we let the witnesses' themselves resolve any multiple picks. As we note later in the results section, there were no differences between the simultaneous and sequential procedures in the rate of multiple picks and we let the witness's own words resolve the final decision for both simultaneous and sequential lineups.

The Second-Viewing Rule

Almost every lab experiment testing the sequential lineup has not permitted the witness to examine the photos a second time. In contrast, every jurisdiction that has adopted the sequential has permitted the witness a second time through the sequential (a second "lap") if the witness explicitly requests it. Although the witness is not told ahead of time that a second viewing is permitted, witnesses who ask to view a photo again after going through the entire sequence are sent back through the same sequence in the same order a second time. Of course, any second viewing is a matter of record that must be disclosed to the defense, and one could argue that a second lap makes the sequential a de facto simultaneous procedure. In keeping with actual practices, the current experiment permitted second viewings if the witness requested it. This allowed us to analyze the results both with and without the second viewing in the subset of cases for which a second viewing occurred.

Explicit Permitting of a "Not Sure" Option

Nearly all lab studies using sequential lineups force witnesses into making a yes or no decision before moving on to the next lineup member. As practiced in actual cases, however, witnesses can continue through the procedure by indicating that they are not sure for any specific photo. The argument from law enforcement is that providing a "not sure" option is more in keeping with how simultaneous lineups are conducted because witnesses are able to say that that they are not sure and make no identification with a simultaneous lineup. Interestingly, this represents a potentially very important gap between how almost all lab studies conduct lineups (regardless of whether they are simultaneous or sequential) and how lineups play out in actual practice. The literature on both simultaneous and sequential lineups is built on a procedure that forces witnesses to either make an identification or to indicate that the culprit is not in the lineup. Including a "not sure" option makes good sense. After all, what is an unsure witness supposed to do? We included the "not sure" option for both the simultaneous and sequential lineups to be consistent with actual practices. We revisit this "not sure" issue in the Discussion section.

These differences between the sequential procedure used in this field experiment and those used in the lab are potentially quite

significant. As described above, police and prosecutors had reasonable arguments about why they used these procedures (e.g., the continuation rule, permitting a second viewing on request, letting witnesses resolve their own decision from multiple picks) and we wanted to test a sequential procedure that reflects how the sequential is conducted in actual practice.

Gaining Experimental Control in the Field

Much has been written in recent years about the challenges and limitations of conducting eyewitness identification experiments in the field (eyewitnesses to actual crimes in active cases) and how that differs from controlled laboratory experiments (e.g., Steblay, 2008; Wells, 2008).

An important difference between laboratory experiments and field experiments concerns control over the scientific protocol, which includes the consistent running of procedures and systematic and complete recording of responses. Laboratory studies permit extensive training of experimenters as well as direct and indirect monitoring of conformity to protocol. In ongoing criminal cases, in contrast, eyewitness identification procedures take place in a large number of different environments (e.g., homes, places of business, police stations), and observational monitoring to examine protocol compliance is unrealistic. Other difficulties with conducting eyewitness identification experiments in the field include the fact that case detectives typically build their own lineups (develop the suspect, select fillers) and then administer their own lineups (e.g., give instructions, interact with the witness, make records of what the witness did and said). Unless critical controls are undertaken, this is potentially much more serious than simply a variance problem and could in fact introduce systematic biases into the outcomes.

Accordingly, we imposed critical control features on this experiment. First, as is normal in actual practice, case detectives continued to build their own lineups (getting a photo of the suspect, selecting filler photos to use) in this experiment. However, they did not know whether the procedure would entail a simultaneous lineup or a sequential lineup when they built the lineup (because that was not randomly assigned until the laptop was turned over to the witness). Hence, detectives could not intentionally or unintentionally select fillers to build "easier" or more "difficult" lineups as a function of whether the lineup was going to be simultaneous versus sequential. Second, we required that lineups be administered by someone who does not know which lineup member is the suspect and which are fillers, thereby making administration of the lineups double-blind. We required that the functional administration of the lineup be controlled by laptop computers. After the laptop was turned over to the witness by the blind administrator, the laptops gave all prelineup instructions to the witness (which allowed for perfect consistency), randomly arranged the order of the lineup photos, and randomly assigned the witness to a simultaneous or sequential presentation. The software then walked the witness through the procedure. Thus, detectives could not inadvertently nudge witnesses away from fillers and toward the suspect. The witness used a mouse to make all identification and nonidentification decisions, which were recorded by the software. Hence, not only were the lineup administrators blind as to which lineup member was a suspect, the procedure itself was administered by the laptop computer, which ensured that all witness responses were faithfully recorded.

Hypotheses

An unambiguous prediction in this field experiment concerns filler identifications. The lab-based literature is very clear in finding that the rate of filler identifications should be lower for the sequential procedure than for the simultaneous procedure (see review by Clark & Godfrey, 2009 and meta-analysis by Steblay et al., 2011). For the same reason, that is, the extant laboratory literature, we predicted that suspect identifications would be lower for the sequential than for simultaneous lineups.

Complicating the predictions somewhat, however, is the fact that the sequential procedure used in the current experiment is not exactly the procedure used in the laboratory. We did not have firm predictions, for example, about how the "not sure" responses would be distributed. Moreover, we did not have firm predictions about how often witnesses would make more than one identification, about how letting the witnesses resolve their own multiple identifications would affect the results, nor as to whether the rate and outcomes of multiple picks might differ between simultaneous and sequential procedures. And, although we were able to analyze what the sequential data would have looked like if a second viewing were not counted (and then what happens when any second viewings are counted), we did not have clear expectations for how often witnesses would request a second viewing or how that second viewing might change the results.

Method

Sample

The final data set consisted of 494 double-blind lineups from witnesses who were attempting to identify a stranger and who were seeing the suspect's photo for the first time. Crimes ranged from credit card theft to murder. Lineups were conducted in four U.S. cities: Charlotte-Mecklenburg, North Carolina (metro population: 2,297,000), Tucson, Arizona (metro population: 992,000), San Diego, California (metro population: 3, 187,000), and Austin, Texas (metro population: 1,834,000). The number of lineups contributed by the four sites varied substantially for a variety of reasons. For example, only a small percentage of the 494 lineups (9.7%) came from the Charlotte-Mecklenburg Police Department because this department had to abruptly discontinue the study when North Carolina mandated that all lineups use the sequential method. Tucson (15.4%) and San Diego (5.9%) experienced technical problems with software and photo database interfaces that limited data collection. As a result, most lineups (69%) came from the Austin Police Department.

Materials: Computer Software and Data Files

With considerable input from detectives, initially at the Charlotte-Mecklenburg Police Department and later at the other three police department sites, the scientists and partners used the expertise of SunGard Public Sector, Inc. in High Point, North Carolina to construct software to administer photo lineups to eyewitnesses.

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

The software was designed to have several key characteristics. First, the software allowed the witness to self-administer the lineup with minimal assistance from a lineup administrator. Second, the software ensured standardized photo display and instructions (both oral and written) to the witness. The software interfaced with the police department's electronic photo repository, thereby allowing uniform photo quality. The presented images of the lineup members were exactly the same size on the screen regardless of whether the display was simultaneous or was sequential.

Importantly, the software randomly scrambled the order of the photos to be presented and randomly assigned the session to follow the simultaneous or sequential procedure. Finally, every mouse click made by the witness was recorded by the software, as was an audio file of the identification procedure, including the conversation between the lineup administrator and witness.

At completion of the identification session, the complete electronic file was uploaded to the police department's server. This record of all the lineup information (all photos, responses, response latencies, order of photos, whether the lineup was simultaneous or sequential, witness information, case information, and so on) was immediately available as a .pdf document on the laptop for the case detective or others to view the results. The uploaded file could always be retrieved from the police department server. The audio recording file (a WAV file) was maintained as a separate file that was also uploaded to the police department server. These files could be readily retrieved from the police department server by any of several means, such as via the case number, witness name, or suspect name. The researchers retrieved those files from the servers via a police department coordinator on the project.

Procedure

Case detective. Detectives from four police departments were trained to use the software program on laptop computers provided by the research team. Detectives were instructed to use whatever their usual criteria were for selecting filler photos from their department electronic photo repository. To prepare a lineup, the case detective loaded the lineup photos via the software onto the laptop.

For each lineup, the case detective also entered witness and case information into the software program. This information became a part of a single electronic file for each lineup that was yoked with the lineup photos and all of the identification data. Such information included the witness's viewing distance, period of observation, status as a bystander versus a victim, whether the witness knew the perpetrator (and if so, how well), and so on. Likewise, information was obtained about the type of crime, whether a weapon was involved (and, if so, what kind), whether there was violence involved, and the witness's description of the perpetrator(s). Detectives ordinarily collect this type of information and document it in some number of police reports, but the program helped make sure this information was electronically recorded, centralized, and readily available for the case detective.

Once the detective created the lineup and entered case information, the lineup file was uploaded to the police department's server, where it became available for download to any of the laptops programmed with the presentation software. When it came time to administer the lineup, the case detective recruited a second person, who did not know which of the lineup members was the suspect and which were fillers, to administer the lineup to the eyewitness.

Lineup administrator. A second (blind) detective conducted the lineup using the laptop computer. The administrator opened the designated lineup program and entered information including his or her own name, the name of the eyewitness, the date, time and location where the lineup was shown, and names of any other persons present during the showing.

The administrator began the lineup procedure with the witness by cueing up the program and turning the computer over to the eyewitness. Once the witness hit the "START LINEUP" button, the computer randomly assigned the lineup to the simultaneous or sequential procedure and randomized the order of the photos with the caveat that the suspect never appeared in Position 1. The decision to not place the suspect in Position 1 was to allay concerns that prosecutors in these jurisdictions expressed about potential defense arguments if the witness identified the first photo they saw in a sequential procedure.

The lineup administrator also asked cued questions and documented the witness's verbal responses. For example, if the witness clicked the "yes" button to indicate that a person or people appeared familiar, the computer asked the witness to make a statement about the identified person ("How do you know this person?"). If the witness reported that lineup member was familiar as "related to the crime," the administrator prompted the witness to use his or her own words to say how sure she or he is that this is the person who committed the crime. The lineup administrator, who could also hear the voice of the computer, wrote down the witness's answers to these questions. The software continued to return the witness back to the lineup until the witness indicated that no one else in the lineup was familiar (with the simultaneous procedure) or until the witness had gone through all the lineup members (with the sequential procedure). This information was documented as part of the audio record and retained as a .pdf written report connected to the lineup file.

The lineup administrator also was the only means of initiating a second viewing ("lap") of the sequential lineup. Witnesses were not told that they could view the sequential lineup a second time. However, if a witness requested a second viewing of the sequential lineup after having gone through the all of the photos, the lineup administrator could initiate a second lap through a password-secure procedure; for any second lap the lineup was shown with photos in the same order. Additional laps beyond two were not permitted.

When the lineup ended, the lineup administrator resumed control over the laptop computer and entered responses to two key questions: whether any aspect of the protocol could not be followed and whether she or he (the lineup administrator) knew which person in the lineup was the suspect (yes/no toggle boxes). A text box was provided to explain any aspects of the procedure that could not be followed. The blind administrator uploaded the file to the police department server.

Instructions to the witness. The computer provided prelineup instructions to the witness, in both written form and via a prerecorded audio using a female speaker. Each instruction was on its own screen and required the witness to acknowledge that she or he understood the instruction before proceeding to the next screen. The rationale for each instruction is summarized below.

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

First, the witness was informed that a lineup contains only one possible suspect and that if the witness indicates that someone is familiar she or he will be asked to indicate whether the person is familiar for reasons related or unrelated to the crime. This instruction was important and consistent with best practices for two reasons. First, it made it clear to the witness that there was only one suspect in the lineup. Accordingly, if the crime was a multiple-perpetrator offense, the witness would know to not look for any more than one of the perpetrators in any given lineup. Second, because witnesses might see someone in the lineup that they know for other reasons (such as someone from their neighborhood), the instructions made clear that they should indicate that fact so that "yes" responses to the familiarity question would yield a record of what they meant by indicating is familiar. The instructions also made it clear to witnesses who were about to view a simultaneous lineup that if they identify someone, they will be returned to the lineup and asked whether another other individuals in the lineup were familiar. Witnesses who were about to view a sequential lineup were told that if they identify someone, they will continue to view all the photos in the lineup. Stating this up front, before the procedure began, was essential so that the witness did not think that returning to the photos after making an identification was some type of "feedback" indicating that their first choice was wrong.

The witness was next instructed that there is no particular order to the photos and to take as much time as needed. This instruction is an important means to help nullify any implicit witness assumption about particular positions in the lineup, and it also makes clear that the pace of progression of the lineup is controlled by the witness, not by the computer.

Three instructions followed recommendations by the NIJ Guide (National Institute of Justice Eyewitness Technical Working Group on Eyewitness Evidence, 1999). The standard admonition that the person who committed the crime may or may not be included in the lineup was employed (Malpass & Devine, 1981; Steblay, 1997), as was a reminder that some features, such as facial hair, can be easily changed and that complexion colors may look slightly different in photos (see also Charman & Wells, 2007). Witnesses were instructed that they did not have to make an identification and that the investigation will continue even if they do not identify someone. The last instruction screen required the witness to click the "continue" button to start the lineup.

When the witness completed the lineup, the blind administrator took control of the laptop. In order to end the session, the lineup administrator had to indicate whether she or he knew which person in the lineup was the suspect and indicate whether there were any aspects of the protocol that could not be followed. In order to count as a double-blind lineup, the lineup administrator had to say that she or he did not know which lineup member was the suspect and that all other aspects of the experimental protocol had been followed. As further assurance that a double-blind procedure was used, we checked the name of the lineup administrator against the name of the case detective to make sure that they were different people.

Results

Overview

The law enforcement agencies downloaded the electronic files from the police department server in the form of both Excel data files and .pdf documents and provided those files to the researchers. Only lineups that were conducted using the established protocol were used in the current work. We began by analyzing aspects of this dataset to be certain that some broad assumptions were met. Subsequently we checked for any lineup position effects for the simultaneous lineups and for the sequential lineups. We compared simultaneous versus sequential differences in rates of suspect identifications, filler identifications, lineup rejections (no to all photos), and "not sure" responses. Finally, we analyzed various estimator variables to see if they moderated the effect of the simultaneous versus sequential manipulation. All statistical comparisons are reported as two-tailed. The statistic r is the effect size indicator, and reported with 95% confidence intervals. Small, medium, and large effect sizes for r are .10, .30, and .50, respectively.

The Data Set

Detectives in the agencies were encouraged to use the software in all cases, even if the case did not meet the protocol requirements, so that they would remain familiar with the software. However, in order to be included in the data set, the lineup had to be conducted by a blind administrator, the eyewitness could not have prior familiarity with the suspect (i.e., stranger cases only), and the witness could not have seen the suspect's image any time after the crime and before the lineup (e.g., no prior show-up or other identification procedure for the witness with this suspect, no exposure to the suspect's photo in news media).

Although it was always clear who the witnesses had identified from the photo lineups, there were 30 cases that fully met the protocol for inclusion but involved witnesses who made more than one identification from a single lineup. In these cases, two raters, blind to the purpose of the study as well as to position of the suspect in the lineup, listened to the audiotapes. Their task was to determine from the audiotape whether the witness had resolved the multiple pick (e.g., "I thought it might have been Number 3, but now I am sure it is Number 5"). The raters were unable to agree on three of the 30 cases and these three lineups were removed from the data prior to analysis (were not part of the $N = 494$). The blind coders agreed on the remaining 27 lineups. The coders' decisions were that 13 of the 27 witnesses had no preference that was discernible from the audios or that the quality of the audio was too poor to discern any preference (these 13 were not included in the subsequent analysis of 494 lineups). The coders also agreed that for the remaining 14 lineups the witnesses had clear preferences; these lineups were included in the final analysis. In six of these 14 cases (five simultaneous and one sequential), the multiple identifications were all of fillers and, therefore, were scored as filler identifications. The remaining eight multiple-identification cases (seven sequential and one simultaneous) involved the identification of one or more fillers along with an identification of the suspect. Witnesses resolved these eight in favor of the suspect in all but one case (a sequential lineup in which the witness resolved the decision with a filler).

Assumptions Tests on the 494 Lineups

Normally, random assignment to conditions is a design characteristic that does not require analyses to make sure that it "worked." However, for this field experiment, there were two

important reasons to examine the random-assignment assumption. First, random-assignment feature of the computer-administered lineup software was developed specifically for this experiment. Although we had confidence in the software, we wanted to make sure that there was no evidence of significant deviation from random assignment in photo lineup position or for simultaneous versus sequential presentation. Second, the only other attempt to conduct a field experiment with simultaneous and sequential lineups (the Illinois Study reported by Mecklenburg, 2006) claimed to have used random assignment (actually, a quasi-random method of every-other-case) in its Evanston site. In fact, however, later analyses of the data (obtained using a Freedom of Information Act request) showed that the sequential lineups were assigned to the most difficult witnessing conditions (e.g., sequential assigned to more cross race cases, longer delay between the crime and the lineup, more likely to be victims, and other factors, see Steblay, 2011).

A first expectation based on random assignment is that approximately 50% of the lineups should have been presented as simultaneous and 50% as sequential lineups. In fact, 47.8% were sequential lineups and 52.2% were simultaneous lineups. As expected, these two rates do not differ significantly, $z = .98$, $p = .32$, $r = .04$, 95% CI [−.05, .13].

Second, we expected the suspects to be placed equally often in each of the lineup positions. The suspect was never placed in Position 1 for either the simultaneous or sequential lineup, therefore the suspect could be in any of five positions, namely 2, 3, 4, 5, or 6, with an expectation of about 20% placement in each position. The percentage of suspect placements in each position did not differ from that expected by chance for either the simultaneous, χ^2 (4, $N = 258$) = .53, $p = .97$, Cohen's $w = .05$, 95% CI [−.07, .17], or the sequential lineups $\chi^2(4, N = 236) = 2.72$, $p = .61$, Cohen's $w = .10$, 95% CI [−.03, .23].

We next tested for differences between the simultaneous and sequential lineups for aspects of the crime and investigation. For example, we examined the number of days that intervened between the witnessed event and the time of the lineup. These data were not distributed normally because the absence of an upper limit skews the numbers in the high direction of the distribution. Hence, we calculated the medians. The median number of days delay for the simultaneous procedure was 13 days and for the sequential procedure the median number of days delay was 14 days, indicating relative equivalence between lineup conditions. Similar results were found for other variables. A weapon was reported to have been used in 46.1% of the simultaneous cases and 48.3% of the sequential cases, a difference that is not statistically significant, z ($N = 494$) = .44, $p = .66$, $r = .02$, 95% CI [−.07, .11]. The percentage of witnesses that were victims rather than bystanders was 71.3% for the simultaneous cases and 69.1% for the sequential cases, also not a statistically significant difference, z ($N = 494$) = .61, $p = .54$, $r = .03$, 95% CI [−.06, .12]. Hence, we found no evidence of differences that exceeded what would be expected based on random assignment.

Main Identification Results

We conducted an initial analysis to see if the pattern of witness decisions (suspect, filler, no identification) varied by site. The San Diego and Charlotte sites were combined because their sample sizes were relatively smaller. The analysis yielded a nonsignificant result for the site factor, $\chi^2(4) = 3.40$, $p = .49$. Hence, for all subsequent analyses we collapsed the data across sites.

Our design permitted us to examine only first lab results (the standard lab practice for the sequential), second lap results, and both laps combined. We began with analyses of first-lap results.

First-lap only results. As predicted, lineup choosing rates were significantly lower with the sequential lineup than simultaneous lineup (34.5% vs. 43.8%), z ($N = 493$) = 2.25, $p = .02$, $r = .10$, 95% CI [.02, .18]. Suspect identifications, however, were only 2.6% lower in the sequential lineup, not a significant difference from the simultaneous lineup, (23.4 vs. 26.0%), z ($N = 493$) = .67, $p = .50$, $r = .03$, 95% CI [−.06, .12]. As anticipated, filler identifications were significantly lower for the sequential lineup (11.1%) than for the simultaneous lineup (17.8%), a significant difference, z ($N = 493$) = 2.33, $p = .02$, $r = .10$, 95% CI [.02, .19]. Importantly, this pattern of first lap results is similar to lab studies, which almost uniformly report only first lap decisions (that is, the sequential procedure produces somewhat lower suspect identification rates and appreciably lower mistaken identification rates compared to the simultaneous lineup; see Figure 1.)

Second-lap behaviors. If the witness requested a second viewing of the lineup, this additional lap was allowed. We now examine in detail the decisions of the 37 witnesses (15.7% of sequential witnesses) who requested a second lap. For 20 of these 37 witnesses their final decision was unchanged by a second lap. Hence, for only 7.2% of the total witnesses who viewed a sequential lineup (17 of 236) did the sequential second-lap policy result in a decision change. Moreover, for two of the 17, the change was from selecting one filler to selecting a different filler. Hence, the second lap policy changed results for only 15 of the 236 sequential witnesses (6.4%). The frequencies are too small for traditional inferential significance tests, but we report descriptive statistics for the 15 witnesses who made a consequential decision change between their first and second laps. Thirteen of the 15 witnesses made no identification on the first lap. Among these 13, 10 identified the suspect on the second lap and three identified a filler.

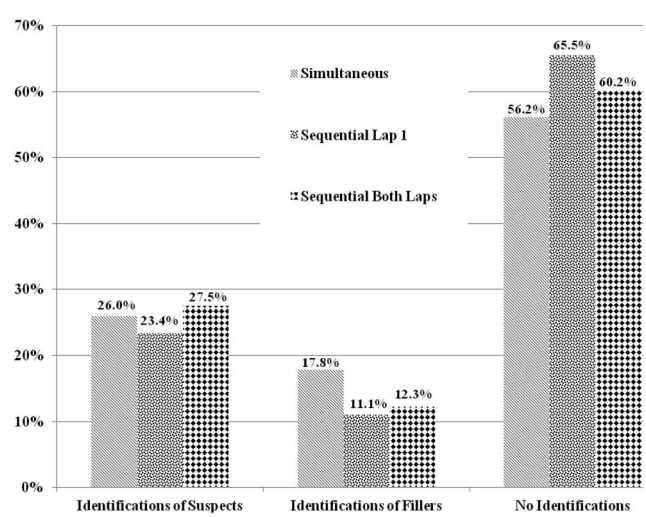

Figure 1. Percentages of suspect, filler, and no identifications as a function of simultaneous versus sequential procedure ($N = 494$).

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

Among the remaining two witnesses, one identified the suspect on the first lap and then identified a filler on the second lap and the other witness identified a filler on the first lap and the suspect on the second lap.

First and second lap results combined. With second laps included as final decisions, sequential lineups yielded a 1.5% higher rate of suspect identification (27.5%) compared to simultaneous lineup procedures (26.0%), a difference that is not statistically significant, z (N = 494) = .38, p = .70, r = .02, 95% CI [−.07, .11]. Filler identifications remained lower for the sequential (12.3%) than for the simultaneous (17.8%), but the difference was no longer statistically significant, z (N = 494) = 1.72, p = .08, r = .08, 95% CI [−.01, .17].

Suspect and filler rates among choosers only. A forensically relevant way to look at these data is to consider rates of suspect and filler identifications among only those witnesses who made identifications (choosers). Figure 2 shows the percentages of all final identification decisions for the simultaneous and sequential lineups. These data are instructive for reasons we will describe in the Discussion section. In particular, the data in Figure 2 show that these witnesses to serious crimes identified an innocent filler almost four of every 10 times that they made an identification (36.2% when collapsed across procedures). And, although the sequential lineup witnesses identified fewer fillers than did the simultaneous witnesses, even the sequential procedure witnesses identified a filler nearly three of every 10 times that they made an identification.

Lineup Rejections Versus "Not Sure" Responses

Nonidentifications were made by 58.1% of the witnesses (287 of the 494). The nonidentification rate was slightly higher for the sequential (60.2%) than it was for the simultaneous procedure (56.2%), but this difference was not statistically significant, z = 1.00, p = .16, r = .04, 95% CI [−.04, .13]. As discussed in the beginning of this article, there are two distinct types of nonidentifications, lineup rejections and "not sure" responses. A lineup rejection is when the witness gives a "no" response to every photo in the lineup. A "not sure" lineup, in contrast, is when the witness never says "yes" to any photo but says "not sure" to at least one photo with the sequential procedure or says "not sure" when viewing the simultaneous procedure.

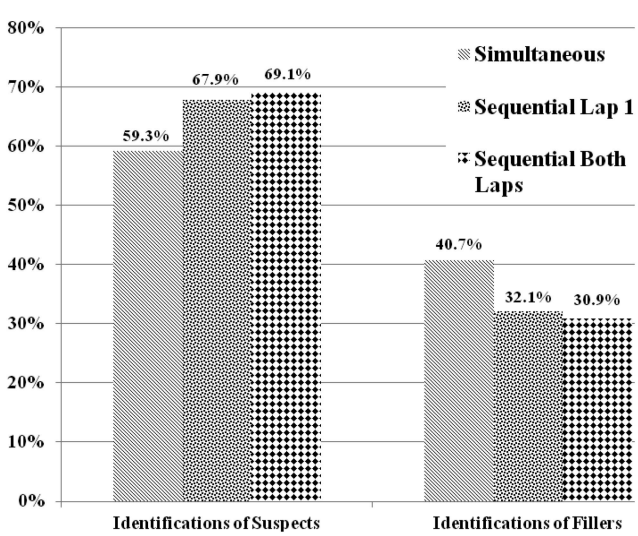

Figure 2. Percentages of suspect identifications and filler identifications among witnesses who made an identification (N = 207).

Analysis of "not sure" versus lineup rejections produced very large differences between the simultaneous and sequential procedures. As shown in Figure 3, for the simultaneous procedure, 80.7% of the nonidentifications were lineup rejections and only 19.3% of witnesses gave a "not sure" response. For the sequential procedure, in contrast, 53.5% of the nonidentifications were lineup rejections and 46.5% included "not sure" responses. The difference in these rates was statistically significant, z (N = 287) = 4.82, p < .001, r = .22, 95% CI [.17, .28].

Hence, compared with the simultaneous procedure, witnesses using the sequential procedure were less likely to reject the lineup altogether when they did not make an identification. An examination of the "not sure" in sequential lineups shows that 28.8% of those who made a "not sure" response specifically indicated "not sure" to the suspect's photo rather than to a filler. For the simultaneous lineups, determining the percentage of time that the witnesses indicated "not sure" specifically to the suspect's photo was more difficult to determine because the "not sure" responses in simultaneous conditions was a global option for whether anyone in the lineup was familiar. Accordingly, audiotapes of "not sure" responses in the simultaneous conditions were examined to see if the witnesses mentioned the suspects' photo as one for which they were unsure. There were only four instances that could be verified of the "not sure" witnesses explicitly mentioning the suspect's photo. However, several of the audio files were not clear enough to determine what the witness said. Therefore, we do not draw a comparative conclusion beyond the fact that witnesses with a simultaneous lineup were less likely to use the "not sure" response option than witnesses with a sequential lineup.

Position Effects?

A position effect means that there is a tendency for a witness to be more or less likely to pick a suspect as a function of where the suspect's photo is placed in an array. One potential concern that has been raised is that the sequential lineup may pose position effects that do not exist for the simultaneous array. Because the suspect was randomly assigned to Positions 2–6 for both the simultaneous and sequential lineups, it was possible to look for position effects in this set of data. Table 1 shows the frequencies with which witnesses selected individuals in Positions 1–6 as a function of the actual position of the suspect (Positions 2–6) for both the sequential and the simultaneous lineup procedures. A position effect would be evident to the extent that the percentage that the suspect is identified deviates from the expected percentage based on the overall rate of suspect identifications. The boldfaced frequencies across the diagonal in Table 1 represent selections of the suspect by position. For the sequential procedure, the suspect was selected between 20.8% of the time (when in Position 4) and 34.8% of the time (when in Position 3). For the simultaneous procedure, the suspect was selected between 20.0% of the time (in Position 3) and 36.5% of the time (in Position 5). Examination of the frequencies suggest that, if anything, the simultaneous procedure appeared to show more evidence for a position effect than did

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

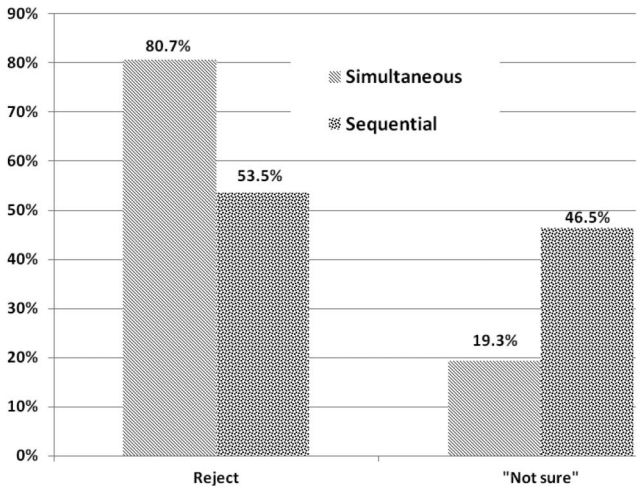

Figure 3. Percentages of lineup rejections and "not sure" responses among witnesses who made no identification ($N = 287$).

the sequential. However, chi square analyses of the frequencies of identifying the suspect show no statistically significant deviation in suspect identification rates as a function of suspect position either for the simultaneous lineups, $\chi^2(4) = 4.92$, $p = .30$, Cohen's $w = .14$, or for the sequential lineups, $\chi^2(4) = 2.88$, $p = .58$, Cohen's $w = .11$.

Estimator Variables

Our interest in the estimator variables that were collected in this experiment was restricted to the question of whether the simultaneous versus sequential variable interacts with the estimator variables, a question that can be important for purposes of generalization of the results. Hierarchical loglinear analyses were used to test for any interactions between lineup procedure, final witness decision, and each of six variables—the presence or absence of a weapon, whether the witness was a bystander or a victim-witness, whether the crime involved violence, the delay between the crime and the lineup (retention interval), the distance between the witness and the perpetrator, and exposure time to the perpetrator. No significant interactions emerged, all $ps > .26$. The analysis had acceptable power ($> .85$) to detect a medium effect at $p = .05$, but only 30% power to detect a small effect.

Table 2 documents proportions of errors (filler picks) among lineup picks for various levels of six estimator variables, between sequential and simultaneous lineups for 207 witnesses who were choosers, that is, made a pick from the lineup. (Recall that rates of witness choosing between lineup procedures for final decisions did not differ significantly; 40% for sequential lineups, 44% for simultaneous lineup.) The proportions of errors for simultaneous versus sequential show considerable variation across the moderator variables. Notably, cell sizes become quite small for some comparisons, thus creating wide confidence intervals. Nevertheless, the direction of the simultaneous/sequential difference was the same (lower error rates for the sequential) for all levels on each of the six estimator variables. In other words, there is no evidence for a reversal of the simultaneous versus sequential difference for any of the 15 comparisons reported in Table 2.

Table 1
Frequencies of Identification Decisions as Functions of Lineup Position of the Suspect for Sequential and Simultaneous Conditions

Sequential	Witness pick							
Suspect position[a]	No ID	1	2	3	4	5	6	Totals
2	35	4	**14**	0	0	1	2	56
3	26	3	0	**16**	1	0	0	46
4	33	0	2	1	**10**	0	2	48
5	25	0	3	0	2	**11**	0	41
6	22	2	2	0	3	1	**14**	44
Total	141	9	21	17	16	13	18	**235**[c]

Simultaneous	Witness pick								
Suspect position[a]	NoID	1	2	3	4	5	6	Unres[b]	Totals
2	28	3	**13**	0	3	1	1		49
3	34	3	2	**11**	0	2	2	1	55
4	32	0	2	1	**10**	2	1	1	49
5	22	3	2	1	3	**19**	1	1	52
6	29	3	4	1	2	1	**13**		53
Total	145	12	23	14	18	25	18	3	**258**

[a] The suspect never appeared in Position 1. Frequencies in boldface across the diagonal are identifications of the suspect. [b] These three cases are unresolved (Unres) for purposes of the position of the identified person because each picked two fillers and there was no final determination from the audiotapes of which filler they preferred. [c] One additional case could not be classified in this table because, although it is known that the witness picked a filler, the position of the suspect was not recorded.

Discussion

We launched this research to find out how the sequential lineup procedure performs relative to the simultaneous lineup procedure using actual eyewitnesses and using characteristics of the sequen-

Table 2
Estimator Variable Analysis by Lineup Procedure: Error Proportion Among 207 Choosers

	Sequential	Simultaneous	*r*	95% CI
Weapon				
Present	.22	.39	.18	[−.03, .37]
Absent	.29	.54	.26	[.03, .45]
Witness status				
Victim	.31	.42	.11	[−.06, .27]
Bystander	.32	.46	.15	[−.15, .42]
Crime				
Violent	.30	.38	.09	[−.07, .25]
Nonviolent	.24	.57	.35	[.07, .59]
Delay between crime and lineup				
0–1 day	.20	.45	.26	[−.19, .62]
2–14 days	.28	.40	.12	[−.11, .34]
15–30 days	.27	.53	.27	[.03, .49]
Distance between witness and perpetrator				
0–1 foot	.27	.48	*.21*	[−.10, .48]
2–3 feet	.35	.42	.07	[−.21, .34]
4–10 feet	.20	.29	.10	[−.23, .41]
Exposure time to perpetrator				
< 1 min	.37	.59	.22	[−.06, .46]
2–5 min	.34	.37	.03	[−.24, .29]
6–20 min	.25	.42	.17	[−.30, .57]

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

tial procedure that are actually used in practice. Examining only first-lap data, the results followed the pattern found in laboratory data, namely that the sequential procedure produced a lower rate of filler identifications than did the simultaneous procedure but also a lower rate of identifying the suspect. Only the difference in filler rates was statistically significant. Less than 16% of sequential witnesses requested a second lap and only 6.4% of sequential witnesses made a consequential change in their identification decision with a second lap. Including these relatively rare second laps had little effect on the pattern of results.

Current Results Versus Lab Studies on Simultaneous/Sequential Differences

Although the results showed an advantage for the sequential procedure overall in reducing the rate of known (filler) errors, it is also clear that the differences were small. We used the effect size *r* to estimate the size of the effect (Cohen, 2008). The value of *r* for the difference in filler rates favoring the sequential was .10 for the first lap and .08 for the first and second laps combined. Cohen defines a small effect for the *r* statistic as .10, a medium effect as .30, and a large effect as .50. This is a smaller effect of the sequential for suppressing incorrect identifications than what tends to be reported in lab experiments (e.g., Steblay et al., 2011). A direct quantitative comparison between these results and lab results cannot be made because the lab estimates are based on two known conditions, namely suppression when the culprit is present in the lineup and suppression when the culprit is absent from the lineup. The Steblay et al. meta-analysis shows that suppression of filler identifications is greater when the culprit is not in the lineup than when the culprit is in the lineup (see Table 3, p. 113 of Steblay et al., 2011). In the current experiment, we do not and cannot know what percentage of the time the culprit is in the lineup, so a quantitative comparison between the lab and the current experiment on these effect sizes is not appropriate.

Although precise quantitative comparisons between lab studies and the current field experiment are precarious, the *pattern* of the results in this field experiment parallels the lab studies in two ways. Specifically, both the current study and lab studies show that (a) witnesses are less likely to make an identification with the sequential procedure, and that (b) witnesses are less likely to identify a filler with the sequential.

The ratio of suspect identifications to filler identifications was higher for the sequential than for the simultaneous in this field test. In the lab, a ratio calculated of accurate culprit identification to filler identifications also favors the sequential procedure. However, we cannot presume that all suspect identifications in the field were accurate. Hence, we cannot definitively say that these results parallel lab results on this point.

We should not be surprised that the magnitude of the effect of the simultaneous versus sequential treatment variable was considerably smaller than what is found in lab studies. In lab studies, the same event is shown to all eyewitnesses, all witnesses view the same lineup faces, the delay to identification is the same for all witnesses, witnesses are drawn from a homogeneous class (e.g., college students), and so on. In other words, in a lab study there is minimal noise and everything is nearly identical except for the focal, manipulated variable. In a field experiment, in contrast, eyewitnesses view different events some of which are stressful and some are not, some witnesses have good views and some have poor views, some view a lineup within 24 hr whereas others do not view a lineup for months, some witnesses are considerably older and others younger, witnesses in different cases are shown different lineup with different fillers and different suspects, and so on. Hence, unlike the lab, a manipulated variable in a field experiment has to overcome a great deal of noise that is bound to dilute the effect of the manipulated variable.

Comparisons to Field and Archival Studies on Identification Behavior Rates

One of the purposes of this work was to examine suspect, filler, and nonidentification rates in actual cases when the data were collected with double-blind methods and full records of all witness responses were guaranteed. How do our experimental results compare and contrast with field and archival studies of actual eyewitnesses that have not followed these protocols (e.g., not double-blind, no random assignment of lineup position, etc.)? There have been nine published field and archival studies that have reported filler and suspect identification rates (these studies are: Behrman, & Davey, 2001; Behrman & Richards, 2005; Horry, Halford, Brewer, Milne, & Bull, 2014; Horry, Memon, Wright, & Milne, 2012; Klobuchar et al., 2006; Memon, Havard, Clifford, Gabbert, & Watt, 2011; Valentine, Pickering, & Darling, 2003; Wright, & McDaid, 1996; Wright, & Skagerberg, 2007). The data from the current field experiment agree closely with these published field and archival studies in some respects but not others. In particular, we note that the percentage of identifications that are of fillers (about 36% overall) is quite close to the average (32%, range 22%–49%) of the nine field and archival studies that have been published in peer-reviewed journals. However, the choosing rate is much lower in this field experiment (42%) than the average in the nine studies (67%).

We believe that there are several important differences that might account for the apparently lower choosing rate. First, archival studies rely on searching case files. We speculate that case file methods might be underestimating the frequency with which eyewitnesses view a lineup and identify no one because there might be no record in the file for many of the nonidentifications. In a 2012 national survey using a stratified random sample of 619 law enforcement agencies in the United States, 37% openly admitted that they do not document nonidentifications (i.e., write no report when the witness makes no identification; see Police Executive Research Forum, 2013). This is a remarkable admission by over one third of U.S. law enforcement agencies because nonidentifications can have exculpatory diagnostic value (see Clark & Wells, 2008) and a failure to make records and disclosures to the defense of nonidentifications of a defendant can be considered a violation of U. S. constitutional law (a "Brady violation" a la *Brady v. Maryland*, 1963). Given the obvious inappropriateness of failing to make records of nonidentifications, we fear that the 37% of departments admitting to this might be an underestimate of those who do not write reports for nonidentifications. Therefore, there could be many case files in archival studies that could appear (erroneously) to not have involved a lineup at all because the eyewitness did not make an identification (resulting in no report of a lineup). In other cases, there might have been multiple witnesses shown lineups, but the only report appearing in the file is the one

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

in which the witness made an identification. This would affect the denominator of the calculated rate of choosing for archival studies, but it would not affect the rate of filler identifications as a function of all identifications. As a result, archival studies might be underestimating how often witnesses make no identification. This could be an important difference between archival studies that rely on case files versus field experiments such as the current one that automatically makes records of the outcome for every lineup.

Another possible reason why the overall choosing rates are lower in this field experiment than is typical of field and archival studies is that our field experiment used double-blind administration procedures. Nonblind lineup administrators might be more likely to push witnesses to make an identification. Consistent with this, the only other field study to use double-blind administration (Klobuchar et al., 2006) produced a choosing rate (46%) similar to the current experiment (42%), which is well below the average for the other eight published field studies that were not blind (70% choosing, range 59%–86%).

An additional reason for lower choosing rates in this experiment is that computer administration of prelineup instructions guaranteed that each and every witness was clearly given the prelineup admonitions (e.g., culprit might not be present, you do not have to make an identification, and so on), which is not something that is guaranteed in the other field studies. In fact, no archival study has ever taken measures to assess whether such admonitions were given at all. In addition, many of the archival studies have failed to ensure that there was not prior familiarity between the witness and the perpetrator or guarantee that the witness had not previously made an identification of the suspect using another procedure (e.g., a show-up).

We do not argue that these field experiment data necessarily represent a superior database for the outcomes of lineups in actual criminal cases compared to what has been obtained with archival approaches. Nevertheless, these data do raise questions about whether nonblind administration, the inability to document prelineup instructions, the absence of guarantees that nonidentifications are documented, and failures to weed out witnesses who had prior familiarity with the suspect, might be skewing some of the results in the nine published field studies, especially those that rely on archival methods.

"Not Sure" Responses

The use of an explicit "not sure" option (or the conceptually related "don't know" option) is not new to the eyewitness identification area, its use dating back over 30 years (Warnick & Sanders, 1980). Moreover, recent work (Perfect & Weber, 2012; Weber & Perfect, 2012) shows that giving eyewitnesses an explicit "don't know" option reduces errors (both incorrect identifications as well as incorrect rejections) with no significant reduction in accurate identification rates. Steblay and Phillips (2011) found that the use of the "not sure" option reduced incorrect identifications more for a sequential than for simultaneous procedures. But, for some reason, providing eyewitnesses with an explicit "not sure" or "don't know" option has never been considered to be a part of standard way of conducting lineups. Instead, the extant literature on simultaneous versus sequential lineups is based almost exclusively on forcing witnesses to make a choice between a positive identification or rejecting the lineup.

Although we could have expected a somewhat higher rate of the "not sure" response for sequential arrays (based on Steblay & Phillips, 2011), we were somewhat surprised at the magnitude of the simultaneous-sequential difference. With the simultaneous procedure 81% of the nonidentifications were rejections and only 19% were "not sure" responses. In contrast, with the sequential procedure only 53% of the nonidentifications were rejections and 47% were "not sure" responses. An argument can be made that the higher rate of clear rejections for the simultaneous procedure than for the sequential procedure is an advantage for simultaneous because a clear rejection is more likely to free an innocent suspect from further suspicion than is a "not sure" response. On the other hand, an argument can be made that the higher rate of "not sure" responses for the sequential than for the simultaneous is an advantage for sequential because it will not necessarily deflect an investigation from a guilty suspect if the witness indicates that he or she is not sure in response to that suspect (nearly 30% of the sequential procedure's "not sure" responses were to a suspect). At this point we take no position on the question of whether the higher rate of "not sure" responses for the sequential is a positive or a negative aspect of lineup performance.

Why did witnesses use the "not sure" response at a higher rate for the sequential than for the simultaneous? A somewhat trivializing explanation might be that the sequential procedure requires six decisions (one for each lineup member), each permitting a "not sure" response, whereas the simultaneous procedure requires only one. But this type of explanation strikes us as inadequate and perhaps even specious. After all, in order to count as a "not sure" response to the sequential lineup, the witness must not only say "not sure" to one of the six photos, but must never say "yes" to any of the six photos for which they also had six chances to do so. If the mere number of chances to say "not sure" is the difference, then why would there not also be more "yes" responses to the sequential lineup, which there were not?

We strongly suspect that the higher rate of "not sure" responses for the sequential procedure is related to the qualitatively different nature of the experience that is inherent in a simultaneous versus a sequential identification procedure. A "not sure" response presumably occurs when the witness sees enough resemblance to single out that photo. The witness is not sure enough that this is the culprit to say "yes" but also not sure enough to definitively rule out the photo by saying "no" and therefore places a "marker" of sorts on that photo (by indicating "not sure"). We speculate that the sequential lineup makes this dilemma more likely to occur because the witness cannot be certain that there is not perhaps another not-yet-seen photo (later in the sequence) that resembles the culprit even more than the photo currently being viewed. According to this account, it is not the sequential procedure's greater number of opportunities to indicate "not sure" that accounts for the higher rate of "not sure" responding; it is the fact that witnesses have to make their judgment without knowing whether there is someone else yet to be seen that resembles the culprit even more than the one being currently viewed. According to this view, sequential witnesses are simply being more careful about reporting decisions. As one reviewer of this work noted, *if* it is important to get witnesses to make clear rejection decisions (e.g., to free an innocent suspect from suspicion), then maybe there is value to following a sequential procedure with a simultaneous procedure in the circumstance when a witness ends the sequential lineup without a

positive identification or a clear lineup rejection. The underlying assumption of this suggestion—that the simultaneous presentation will allow witnesses to make a clear reject versus "not sure" decision without also eliciting increased mistaken identifications—is yet to be explored.

The Interpretation of Filler Identification Rates

When conducting research with lineups in actual cases, filler identifications are the only eyewitness behaviors that can be coded for accuracy because they are, by definition, mistaken identifications. But, a mistaken identification of a filler in itself is of no particular forensic consequence to the selected lineup member: a filler identification is a known wrong answer, and that filler is unlikely to face subsequent investigation or scrutiny. Hence, in what sense are filler identification rates in actual cases important?

We argue that filler identifications are an indication of the riskiness of a procedure for innocent suspects. When an eyewitness identifies a filler, it means that the eyewitness is accusing an innocent person of being the culprit. Although the filler has a hidden "trump card" (filler status) of sorts to avoid the jeopardy of being charged with a crime, the fact that this innocent person was identified indicates a propensity for mistaken identifications to occur. Had the filler been an innocent suspect, some severe consequences could follow. Hence, if the simultaneous procedure inflates rates of filler identifications relative to a sequential procedure, it logically follows that it also inflates risk to an innocent suspect.

Although filler identification rates are a proxy measure to assessing risk to innocent suspects for purposes of the current field experiment, filler identification rate differences between simultaneous and sequential do not permit *quantitative* estimates of how much more at risk an innocent suspect is under one procedure versus the other. Quantitative estimates are not possible because the question of risk to an innocent suspect presumes that the perpetrator is not in the lineup (i.e., the sole suspect in the lineup is innocent). And, both logic and data tell us that the rate of filler identifications is higher when the suspect is innocent than when the suspect is guilty (e.g., Clark & Davey, 2005; Wells, 1993). In this field experiment we do not know what percentage of the time the suspect was innocent.

Another reason that filler identification rates likely underestimate risk to an innocent suspect is that innocent suspects are often more likely to be mistakenly identified than is the average filler. This will be true when the suspect "stands out" for some reason, which seems to be a fairly common problem. Occurrences of an innocent suspect being more likely to be identified than a filler seems to be a problem not just for actual cases but also lab experiments in which the researchers are trying to construct fair lineups. In the controlled lab studies that compare show-ups with lineups, for example, the rate of choosing the innocent suspect from a perpetrator-absent lineup is over three times the rate of the choosing the average filler (see meta-analysis by Steblay, Dysart, Fulero, & Lindsay, 2003). Likewise, in their analysis of 94 studies examining lab-based regularities in eyewitness identification, Clark, Howell, and Davey (2008) show data indicating that that an innocent suspect is more than two times as likely to be identified from an absent lineup than is the average filler (see Table A1, pp. 212–214).

Yet another reason to consider the filler identification rates to be conservative estimates of identification error is the fact that we cannot consider all of the suspect identifications to have been accurate identifications. The only identifications that we can know with certainty were mistaken are the filler identifications. When witnesses made an identification, they identified a filler 42% of the time with the simultaneous procedure and identified a filler 31% of the time with the sequential procedure. Given those figures, it seems likely that some portion of the identifications of suspects were also mistaken identifications.

Finally, we note that filler identification rates are a concern not only because of what they tell us about the propensity for a procedure to produce identifications of innocent suspects, but also because of the fact that a filler identification "burns" the witness for any later identification task. Suppose, for instance, that a detective puts together a lineup focused around his suspect "Joey" and the witness views the lineup and identifies a filler. Later, an anonymous tip comes in indicating that the actual perpetrator was "Sam." Can the detective now put Sam in a lineup and show the witness? Yes but, unfortunately, the eyewitness is already burned as far as credible testimony later in court. In fact, if that happened, this is one of the very few situations in which judges would likely suppress any identification of Sam by this witness—the witness has impeached himself or herself. Filler identifications are unfortunate for everyone except the actual perpetrator.

Final Observations

Although we launched this work to find out if a sequential procedure used in many jurisdictions around the country reduces errors when compared with a simultaneous procedure, we found the performance of these witnesses to be quite poor regardless of the procedure used. If data like these had been obtained in a lab experiment, reviewers might have complained about the overall low performance and suggested that it was unrealistic. After all, four out of every 10 witnesses who made an identification from a simultaneous lineup identified a filler and three out of every 10 identified a filler from a sequential lineup. But, the result from this experiment is not an outlier in the data on actual eyewitnesses. In fact the rate of filler identifications among choosers for the nine published field and archival studies averages 33% and the sample of studies has a 95% confidence interval boundaries of 27% (lower bound) and 37% (upper bound). Moreover, these rates have to be underestimates of the true error rate to the extent that not all suspects who were identified were in fact the guilty parties.

The consistent result of finding three or more incorrect identifications out of every 10 identifications among actual witnesses seems to belie the claims of some critics of lab-based eyewitness research. Critics of lab-based eyewitness identification research findings argue that actual witnesses to serious crimes, unlike those in lab studies, rarely make errors because the witnessed events are "significant" and therefore greater caution is taken in making identifications when there are real-world stakes involved. Accordingly, some in the legal system have been dismissive of lab studies and argue that they create an exaggeratingly poor picture of eyewitness identification (e.g., Mecklenburg, Bailey, & Larson, 2008).

Another criticism of lab studies is that they use data from every witness whereas in actual cases it is only witnesses who say they

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

could identify the perpetrator who are shown lineups. But this criticism makes the current results appear even more alarming. Presumably, the witnesses in the current field experiment were those who said they thought they could identify the perpetrator and then viewed a lineup. And yet, three of every 10 of witnesses who identified a lineup member identified a known-innocent person.

There are several possibilities that alone and together might account for what we think is overall poor performance by eyewitnesses. First, of course, there is the base-rate issue. Lab experiments show that rates of filler identification increase when the actual culprit is not in the lineup, an observation that has been shown repeatedly for decades (e.g., Wells & Lindsay, 1980). A relatively low base rate (suspect not being the culprit) for the lineups in the current experiment would tend to account for both the relatively high rate of identifying fillers (as a proportion of all identifications) and the relatively low rate of identifying suspects. We consider this to be a credible possibility, but we have no way of knowing if this is the case.

Another possibility for the overall poor performance of the witnesses is that photos of the culprits failed to adequately capture the culprits' appearances, perhaps because they were out of date or of poor quality. But the photos used by all four of these agencies were of very high quality. And, although some might have been older photos that failed to capture how the culprit looked more recently, we see two problems with this explanation. First, the argument that many of the photos failed to capture the appearance of the suspect fails to explain why archival data on *live* lineups (not photos) also tends to produce similarly poor performance. For example, Behrman and Davey (2001) analyzed actual cases of live lineups (not photos) and found that 32% of witnesses who made identifications still picked known-innocent fillers. Second, even if the photos failed to capture the appearance of the suspect, it fails to explain why witnesses nevertheless made identifications of innocent fillers rather than identify no one.

Yet another possibility for the poor performance of the witnesses is that real-world witnessing and delay conditions are quite unfavorable for eyewitnesses. These conditions include poor lighting, long witnessing distances, the distracting presence of weapons, fear and stress, long delays to the time of the lineup, and so on. Lab studies rarely throw in these types of detrimental factors, even in isolation let alone in combination. This would account for the low rate of suspect identifications but, again, it begs the question of why witnesses would identify fillers three or four of every 10 times they make an identification rather than make no identification at all (e.g., use the "not sure" option).

Regardless of the explanation for the meager performance of these actual eyewitnesses, it is obvious that the sequential procedure is not the silver bullet. Whereas the current set of reforms proposed in psychological science (e.g., prelineup admonition instructions, double-blind administration, proper filler selection, sequential presentation) can help reduce identification errors, there is a long way to go. One avenue is to develop additional system-variable improvements, perhaps by beginning to think of procedures outside the box of the traditional lineup procedure (see Brewer & Wells, 2011). A recent line of work using ecphoric confidence ratings rather than identification decisions is one example of a novel alternative to the traditional lineup (Brewer, Weber, Wooten, & Lindsay, 2012). But the current results also suggest that advances in system variables might not be enough and that efforts must continue to be devoted to an improved ability to postdict eyewitness identification accuracy using estimator variables derived from both situational factors (e.g., characteristics of the witnessing situation itself) as well as witness behaviors (e.g., decision latency, witnesses' verbal utterances during the identification decision).

In spite of the rather sobering results regarding identification decisions of real eyewitnesses, this project achieved a promising level of methodological sophistication, standardization, and control in a field lineup experiment with a randomly assigned manipulation, allowing reliable measurement of witness decisions within the complex environment of real crime investigations. We hope that the methodological ideas implemented here with actual eyewitnesses, especially the use of computer-controlled randomization, computer-driven presentation, and automatic recording of results will embolden other researchers to experimentally examine important lab findings in the field with actual eyewitnesses.

References

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

Behrman, B. W., & Richards, R. E. (2005). Suspect/foil identification in actual crimes and in the laboratory: A reality monitoring analysis. *Law and Human Behavior, 29,* 279–301. doi:10.1007/s10979-005-3617-y

Brady v. Maryland, 373 U.S. 83 (1963).

Brewer, N., Weber, N., Wooton, D., & Lindsay, D. S. (2012). Identifying the bad guy in a lineup using deadlined confidence judgments. *Psychological Science, 23,* 1208–1214. doi:10.1177/0956797612441217

Brewer, N., & Wells, G. L. (2011). Eyewitness identification. *Current Directions in Psychological Science, 20,* 24–27. doi:10.1177/0963721410389169

Charman, S. D., & Wells, G. L. (2007). Eyewitness lineups: Is the appearance-change instruction a good idea? *Law and Human Behavior, 31,* 3–22. doi:10.1007/s10979-006-9006-3

Clark, S. E. (2012). Costs and benefits of eyewitness identification reform: Psychological science and public policy. *Perspectives on Psychological Science, 7,* 238–259. doi:10.1177/1745691612439584

Clark, S. E., & Davey, S. L. (2005). The target-to-foils shift in simultaneous and sequential lineups. *Law and Human Behavior, 29,* 151–172. doi:10.1007/s10979-005-2418-7

Clark, S. E., & Godfrey, R. D. (2009). Eyewitness identification evidence and innocence risk. *Psychonomic Bulletin & Review, 16,* 22–42. doi:10.3758/PBR.16.1.22

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

Clark, S. E., & Wells, G. L. (2008). On the diagnosticity of multiple-witness identifications. *Law and Human Behavior, 32,* 406–422. doi:10.1007/s10979-007-9115-7

Cohen, Jacob (2008). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Erlbaum, Inc.

Garrett, B. L. (2011). *Convicting the innocent: Where criminal prosecutions go wrong*. Cambridge, MA: Harvard University Press. doi:10.4159/harvard.9780674060982

Greathouse, S. M., & Kovera, M. B. (2009). Instruction bias and lineup presentation moderate the effects of administrator knowledge on eyewitness identification. *Law and Human Behavior, 33,* 70–82. doi:10.1007/s10979-008-9136-x

Gronlund, S. D., Carlson, C. A., Dailey, S. B., & Goodsell, C. A. (2009). Robustness of the sequential lineup advantage. *Journal of Experimental Psychology: Applied, 15,* 140–152. doi:10.1037/a0015082

Gronlund, S. D., Carlson, C. A., Neuschatz, J. S., Goodsell, C. A., Wetmore, S. A., Wooten, A., & Graham, M. (2012). Showups versus

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

lineups: An evaluation using ROC analysis. *Journal of Applied Research in Memory & Cognition, 1,* 221–228. doi:10.1016/j.jarmac.2012.09.003

Horry, R., Halford, P., Brewer, N, Milne, R., & Bull, R. (2014). Archival analyses of eyewitness identification test outcomes: What can they tell us about eyewitness memory? *Law and Human Behavior, 38,* 94–108. doi:10.1037/lhb0000060

Horry, R., Memon, A., Wright, D. B., & Milne, R. (2012). Predictors of eyewitness identification decisions from video lineups in England: A field Study. *Law and Human Behavior, 36,* 257–265. doi:10.1037/h0093959

Horry, R., Palmer, M. A., & Brewer, N. (2012). Backloading in the sequential lineup prevents within-lineup criterion shifts that undermine eyewitness identification performance. *Journal of Experimental Psychology: Applied, 18,* 346–360. doi:10.1037/a0029779

Klobuchar, A., Steblay, N. M., & Caligiuri, H. L. (2006). Improving eyewitness identifications: Hennepin County's blind sequential lineup pilot project. *Cardozo Public Law, Policy, and Ethics Journal, 2,* 381–414.

Lindsay, R. C. L., Lea, J. A., & Fulford, J. A. (1991). Sequential lineup presentation: Technique matters. *Journal of Applied Psychology, 76,* 741–745. doi:10.1037/0021-9010.76.5.741

Lindsay, R. C. L., & Wells, G. L. (1985). Improving eyewitness identification from lineups: Simultaneous versus sequential lineup presentations. *Journal of Applied Psychology, 70,* 556–564. doi:10.1037/0021-9010.70.3.556

Malpass, R. S., & Devine, P. G. (1981). Eyewitness identification: Lineup instructions and the absence of the offender. *Journal of Applied Psychology, 66,* 482–489. doi:10.1037/0021-9010.66.4.482

Mecklenburg, S. H. (2006). *Report to the legislature of the state of Illinois: The Illinois pilot program on double-blind, sequential lineup procedures*. Springfield, IL: Illinois State Police.

Mecklenburg, S. H., Bailey, P. J., & Larson, M. R. (2008). The Illinois field study: A significant contribution to understanding real world eyewitnesses. *Law and Human Behavior, 32,* 22–27. doi:10.1007/s10979-007-9108-6

Memon, A., Havard, C., Clifford, B., Gabbert, F., & Watt, M. (2011). A field evaluation of the VIPER system: A new technique for eliciting eyewitness evidence. *Psychology, Crime & Law, 17,* 711–729. doi:10.1080/10683160903524333

Mickes, L., Flowe, H. D., & Wixted, J. T. (2012). Receiver operating characteristics analysis of eyewitness memory: Comparing the diagnostic accuracy of simultaneous and sequential lineups. *Journal of Experimental Psychology: Applied, 18,* 361–376. doi:10.1037/a0030609

Munsterberg, H. (1908). *On the witness stand*. New York, NY: Doubleday.

National Institute of Justice Eyewitness Technical Working Group on Eyewitness Evidence. (1999). *Eyewitness evidence: A guide for law enforcement*. Washington, DC: United States Department of Justice, Office of Justice Programs.

Oregon v. Lawson, SC S059306, 2012.

Palmer, M. A., & Brewer, N. (2012). Sequential lineup presentation promotes less-biased criterion setting but does not improve discriminability. *Law and Human Behavior, 36,* 247–255. doi:10.1037/h0093923

Perfect, T. J., & Weber, N. (2012). How should witnesses regulate the accuracy of their identification decisions: One step forward, two steps back? *Journal of Experimental Psychology: Learning, Memory, and Cognition, 38,* 1810–1818. doi:10.1037/a0028461

Police Executive Research Forum. (2013). *A national survey of eyewitness identification procedures in law enforcement agencies*. Retrieved from http://policeforum.org/library/eyewitness-identification/NIJEyewitnessReport.pdf

Rodriguez, D. N., & Berry, M. A. (2014). The influence of administrator blindness on the recording of eyewitness identification decisions. *Legal and Criminological Psychology, 19,* 69–79. doi:10.1111/j.2044-8333.2012.02058.x

Steblay, N. (1997). Social influence in eyewitness recall: A meta-analytic review of lineup instruction effects. *Law and Human Behavior, 21,* 283–297. doi:10.1023/A:1024890732059

Steblay, N. (2008). Commentary on "Studying eyewitness investigations in the field:" A look forward. *Law and Human Behavior, 32,* 11–15. doi:10.1007/s10979-007-9105-9

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

Steblay, N., Dysart, J., Fulero, S., & Lindsay, R. C. L. (2003). Eyewitness accuracy rates in police showup and lineup presentations: A meta-analytic comparison. *Law and Human Behavior, 27,* 523–540. doi:10.1023/A:1025438223608

Steblay, N., 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,* 99–139. doi:10.1037/a0021650

Steblay, N., & Phillips, J. D. (2011). The not-sure response option in sequential lineup practice. *Applied Cognitive Psychology, 25,* 768–774. doi:10.1002/acp.1755

Thompson-Canino, J., Cotton, R., & Torneo, E. (2010). *Picking cotton*. New York, NY: St. Martin's Press.

Tollestrup, P. A., Turtle, J. W., & Yuille, J. C. (1994). Actual witnesses to robbery and fraud: An archival analysis. In D. F. Ross, J. D. Read, & M. P. Toglia (Eds.), *Adult eyewitness testimony: Current trends and developments* (pp. 144–162). New York, NY: Cambridge University Press. doi:10.1017/CBO9780511759192.008

Valentine, T., Pickering, A., & Darling, S. (2003). Characteristics of eyewitness identification that predict the outcome of real lineups. *Applied Cognitive Psychology, 17,* 969–993. doi:10.1002/acp.939

Warnick, D. H., & Sanders, G. S. (1980). Why do eyewitnesses make so many mistakes? *Journal of Applied Social Psychology, 10,* 362–366. doi:10.1111/j.1559-1816.1980.tb00716.x

Weber, N., & Perfect, T. J. (2012). Improving eyewitness identification accuracy by screening out those who say they don't know. *Law and Human Behavior, 36,* 28–36. doi:10.1037/h0093976

Wells, G. L. (1978). Applied eyewitness testimony research: System variables and estimator variables. *Journal of Personality and Social Psychology, 36,* 1546–1557. doi:10.1037/0022-3514.36.12.1546

Wells, G. L. (1980). Eyewitness behavior: Special Issue. *Law and Human Behavior, 4,* 237–242. doi:10.1007/BF01040616

Wells, G. L. (1984). The psychology of lineup identifications. *Journal of Applied Social Psychology, 14,* 89–103. doi:10.1111/j.1559-1816.1984.tb02223.x

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

Wells, G. L. (2008). Field experiments on eyewitness identification: Towards a better understanding of pitfalls and prospects. *Law and Human Behavior, 32,* 6–10. doi:10.1007/s10979-007-9098-4

Wells, G. L., & Lindsay, R. C. L. (1980). On estimating the diagnosticity of eyewitness nonidentifications. *Psychological Bulletin, 88,* 776–784. doi:10.1037/0033-2909.88.3.776

Wells, G. L., Steblay, N., & Dysart, J. E. (2012). Eyewitness identification reforms: Are suggestiveness-induced hits and guesses true hits? *Perspectives on Psychological Science, 7,* 264–271. doi:10.1177/1745691612443368

Wright, D. B., & McDaid, A. T. (1996). Comparing system and estimator variables using data from real lineups. *Applied Cognitive Psychology, 10,* 75–84. doi:10.1002/(SICI)1099-0720(199602)10:1<75::AID-ACP364>3.0.CO;2-E

Wright, D. B., & Skagerberg, E. M. (2007). Post-identification feedback affects real eyewitnesses. *Psychological Science, 18,* 172–178. doi:10.1111/j.1467-9280.2007.01868.x

Received June 29, 2012
Revision received March 17, 2014
Accepted March 17, 2014 ■

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.