Recent years have seen a convergence of scientific findings and legal policy regarding eyewitness evidence. Eyewitness scientists have provided a context for better understanding of the eyewitness error that permeates DNA-exoneration cases. Furthermore, memory principles established in the laboratory have produced empirically based guidelines for procedural changes that can enhance the accuracy of eyewitness evidence. These recommendations have focused on police lineups, the identification task that has the potential to secure eighty evidence about a suspect's criminal culpability. Among the recommendations for improved lineup practice are that each lineup — photo or live — be administered double-blind (often simply called “blind”) and that the lineup be presented sequentially (one at a time) rather than in the traditional simultaneous (all at one time) format.

A number of jurisdictions have implemented the double-blind sequential lineup as a means to gain greater confidence in eyewitness evidence within their investigations and for the courtroom. However, resistance to change in established practice can be strong, and the discord is perhaps nowhere better exhibited than in the brouhaha surrounding the Illinois Pilot Program on Double-Blind Sequential Lineup Procedure (the Program). In 2002, the Illinois Governor’s Commission on Capital Punishment recommended that police lineups be conducted by a blind administrator (a person who does not know who the suspect is) and also encouraged use of a sequential format for the display of lineup members. These recommendations were not adopted by the Illinois Legislature, however. Instead, a statute was enacted that called for a pilot study on the effectiveness of the sequential lineup. Results of the year-long Program were released in 2006 in a document commonly referred to as the Mecklenburg Report.

The Mecklenburg Report suggested that existing Illinois police lineup practice was superior to the procedures recommended by laboratory scientists. More specifically, in two of the three cities tested, the status quo method of presenting lineups (non-blind simultaneous) produced higher suspect identification rates and lower filler picks (known errors) than did double-blind sequential lineups. The third city tested, Joliet, produced outcomes in line with laboratory results. Scientists and critics immediately challenged the conclusions of the report on methodological grounds, and in the two years following, numerous commentaries were circulated. Debate emanated from two primary camps. Critics of the Mecklenburg Report decried the Illinois Program as unscientific and poorly executed project and rejected its conclusions. This position was bolstered by an independent, neutral, “Blue-Ribbon Panel” of seven highly credentialed scientists convened to examine the Program's
research design. The panel’s judgment: the study had a fundamental confound in its comparison of double-blind sequential and non-blind simultaneous lineups, with “devastating consequences for assessing the real-world implications … [and] guaranteed that most outcomes would be difficult or impossible to interpret.” The Program was deemed unreliable as a basis for determining effective eyewitness identification procedures.

On the other side, supporters of the Mecklenburg Report heralded the project as “a significant contribution to understanding real-world eyewitness identification issues.” For these advocates, the “real-life” aspects of the study seemingly trumped limitations of its design, as reflected in the comment by Joliet Police Chief Hayes: “But the studies that support the double-blind lineup are based on academic research, not field studies conducted by police on actual witnesses.”

Interpretation of the results matters; there has been and will continue to be an impact on law enforcement practice and more broadly on how laboratory memory research is viewed by law enforcement and policymakers, and in courtroom decisions. But, by 2008 the argument seemed to be stubbornly stalled, without resolution. A sticking point was the question of why the Illinois data look as they do. Methodological arguments aside, scientists could not definitively demonstrate that the design flaws of the study had influenced its outcome. And, very little was known about the exact police practice that produced the results, a non-trivial point given that Mecklenburg directly compared the Illinois results to lab studies and extant field data. Lab scientists evaluate sequential versus simultaneous lineups almost exclusively by examining the witness’s first attempt to identify a stranger-perpetrator. Yet, the Illinois results included instances of repeated identification tasks that involved the same witness viewing the same suspect. A comparison of first identification attempts in the lab to second or third attempts in the field is not an apples-to-apples contest. Practically speaking, jurisdictions seeking guidance about identification evidence also require knowledge of how closely their practices mimic Illinois. In short, we needed a better understanding of what we were arguing about. The Mecklenburg Report provided insufficient detail about police protocol and case details; the case files were needed. NACDL, represented by Professor Locke Bowman, Legal Director of the MacArthur Justice Center at Northwestern University School of Law, initiated an FOIA lawsuit in Cook County Circuit Court seeking the original files of the study from the three cities involved. I joined the effort as NACDL’s pro bono consultant.

The Evanston Police Department complied with the FOIA lawsuit and forwarded case files from 100 field study identification tests. Detailed analysis of the Evanston data is reported in a recent peer-reviewed article in Law and Human Behavior. In this shorter article for The Champion, I wish to highlight four points that emerged from the Evanston data analysis, two that speak to critical methodological problems of the Illinois study and two that address aspects of identification evidence relevant to police practice.

‘Objective Scientific Research Methodology’

The legislative intent of the Illinois Pilot Program was to determine the efficacy of the double-blind sequential lineup compared to traditional Illinois practice, and the study was to be “designed to elicit information for comparative purposes, and … consistent with objective scientific research methodology.”

The power of a proper experimental design, in the lab or in the field, derives from the fact that the focal variable (the experimental treatment) is controlled by the experimenter but within the constraint of random assignment — the experimenter does not get to choose which subjects experience which treatment. For example, the specific lineup procedures to be tested (sequential and simultaneous) are defined by the researcher, but after each new lineup is prepared for a witness, it must be randomly assigned to one of the two lineup conditions. Simply put, each lineup/witness must have an equal chance of being assigned to each lineup condition (sequential or simultaneous). This basic methodological tool is used to make the compared groups roughly equivalent on extraneous factors — importantly in this case, factors that have an impact on eyewitness memory and identification decisions. Random assignment will evenly distribute — between the simultaneous and sequential lineup conditions — such factors as crime elements (e.g., type of crime, presence of a weapon); aspects of witness memory encoding (e.g., duration of the crime, cross-race event, viewing conditions, previous acquaintance with the culprit); and memory retrieval (e.g., delay between crime and lineup, quality of the lineup). The impact of lineup procedure (simultaneous or sequential) then can be directly assessed by comparing witness decisions between the two lineup conditions. Extraneous factors will have similar impact on the two experimental groups, thereby allowing the unique effect of lineup procedure to be revealed in witness decisions.

Method: Evanston’s Failure in Random Assignment of Lineups to The Two Tested Conditions

The Evanston lineups were assigned to lineup conditions using an even-odd case number method, a strategy that at first blush may seem reasonable; however, critics early on noted this violation of true random assignment. The criticisms were grounded in methodological theory — the presumption that non-random assignment to lineup format conditions introduced multiple confounds into the comparison between sequential and simultaneous lineups that would have an impact on the study results. What real difference, if any, did non-random assignment make? The FOIA lawsuit provided the data to explore this question.

A worst-case scenario from failure of random assignment is that the playing field will be so uneven that a fair test of the compared groups is impossible. In the case of Evanston, this unfortunate scenario is blatantly apparent in the preponderance of verification and confirmatory lineups in the non-blind simultaneous (NB-SIM) lineup condition. Compared to the double-blind sequential (DB-SEQ) lineup condition, the NB-SIM lineup witness was more likely to have known the offender prior to the crime (39 percent vs. 22 percent), or to be confirming a previous identification (17 percent vs. 2 percent). Alternately stated, witnesses with DB-SEQ lineups were significantly more likely to be attempting a first identification of a stranger-perpetrator (74 percent vs. 50 percent), a memory task that is clearly more difficult than to identify an acquaintance or to confirm a previous identification. (Confirmatory lineups presumably also screen in witnesses who were “successful” at the earlier identification task.) Verification and confirmatory lineups will likely boost suspect identification rates and lower filler picks. Indeed, when confirmatory and
verification lineups are removed from the analyses, we see a differential impact on the two lineup conditions. Suspect identifications in the NB-SIM condition are now 17.7 percent lower than reported by Mecklenburg, but only a small impact on suspect identifications is apparent in the DB-SEQ condition (an increase of 3 percent). In sum, the lack of true random assignment had a significant impact, so that a fair comparison between lineup conditions was not possible. And, the design created a specific circumstance that favored the non-blind simultaneous lineup over the double-blind sequential lineup.

Can the comparison between NB-SIM and DB-SEQ lineups be valid after confirmatory and verification lineups are removed? In a word, no. The uneven placement of confirmatory and verification lineups in the two lineup conditions is a symptom of a larger problem. (This problem is perhaps aggravated by the small sample sizes in Evanston.) When random assignment has been violated, the list of confounding variables that may individually or in combination differentially affect eyewitness decisions in NB-SIM versus DB-SEQ lineups cannot be easily or exhaustively catalogued, or be controlled after-the-fact.

Some confounds in fact will go unrecognized or cannot be measured. Consider the potential impact of culprit-absent lineups on witness decisions (lineups in which the suspect is in fact innocent of the crime), a crucial factor that must be balanced between tested conditions in order for a fair comparison to occur, but one we cannot assess in the field. It is not difficult to recognize that a disproportionate collection of culprit-absent lineups in one lineup condition (either simultaneous or sequential) will affect study outcomes. Importantly, a culprit-absent lineup turns the tables on how we tally witness decision accuracy: a witness’s non-identification of the innocent suspect is an accurate response; a pick of the suspect is a dangerous error. Mecklenburg explains that a live lineup occurs only after probable cause has been established and therefore that Illinois physical lineups inherently have a greater rate of guilty suspects than do photo arrays.22 If Mecklenburg is correct, then the greater number of physical lineups in the NB-SIM condition (33 percent vs. 20 percent) means that the NB-SIM lineup condition involved fewer culprit-absent lineups than did the DB-SEQ.

Although many potential confounds exist, information regarding only a limited number can be gleaned from the available police reports. However, even with these few available facts, significant problems are apparent. As detailed in Law and Human Behavior,23 the Evanston data reveal differences between the non-blind simultaneous and the double-blind sequential lineup conditions on a number of factors that should have been equivalent between the two groups if random assignment had been effectively deployed. For example, NB-SIM lineups were significantly more likely to involve a short delay (a lineup within 24 hours of the crime) than were DB-SEQ (47 percent vs. 9 percent). Bystander eyewitnesses were more common in the NB-SIM lineup conditions (33 percent vs. 14 percent), as also were same-race identifications (52 percent vs. 35 percent), crimes with weapons (56 percent vs. 36 percent) and live lineups (33 percent vs. 20 percent).

The precise impact that this accrual of discrepancies between the two lineup conditions had on eyewitness decisions cannot be ascertained. However, the essential point is that these differences establish empirically that random assignment to the two tested lineup formats was not effectively employed in the Evanston study and that it produced significant a priori differences between the tested groups. This is a critical point: the numerous confounds in the design cannot be countered with attempts to explain them away one-by-one. No evaluative comparison between the two lineup conditions can be justified.

The remaining two cities, Chicago and Joliet, present similar problems of non-random assignment to lineup conditions. In each city, two sites were defined, by detective district and geographical region, respectively. Cases originating in one area were tested with NB-SIM lineups and cases from another were tested with DB-SEQ lineups. That is, assignment to condition was predetermined and fixed (not at the discretion of the detective), on a non-random basis.24

Method: Faulty Pseudo-Experimental Comparisons

This second methodological issue speaks to Mecklenburg et al.’s claim that the “Illinois study contains valuable data unrelated to the confound” (referring to the confound highlighted by Schacter’s panel).24 For example, Mecklenburg et al. attempt to use their field data to raise doubt about established experimental laboratory findings for several variables — weapon focus, violence level, cross-race identification, and victim/bystander status — by stating that the Illinois study did not show the same effects seen in the lab.25 The Evanston data illustrate why Mecklenburg’s claim is misguided. Consider same-race crimes in comparison to crimes in which the witness and perpetrator were of different races. In Evanston, same-race crimes were more likely to involve weapons and multiple perpetrators, and less likely to involve Caucasian witnesses, stranger-perpetrators, and delays before the lineup — a morass of factors that are impossible to separate. Similarly, same-race versus cross-race crimes may confound other unrecorded variables such as quality of offender description and presence/absence of the culprit in the lineup. We simply cannot discern the specific impact of same or cross-race crimes on eyewitness decisions in these field data. The co-occurrence of multiple influences on eyewitness memory within a case makes the precise relationship between any one of these influences and the eyewitness’s decision indecipherable.

It is possible to examine a factor that was not experimentally controlled by the researcher in a field test, just as I have reported above with variables of delay, same-race, and weapons. For example, one may compare lineup decisions of witnesses who were exposed to a weapon versus those who were not. But, the witness’s experience of exposure to a weapon in these real-life cases was not controlled by the experimenter (and of course not randomly assigned), so it is prudent to remember that this pseudo-experimental comparison cannot provide valid information beyond a simple description of how the two groups of witnesses perform. However tempting, this circumstance does not allow conclusions about why any differences in performance occurred (cause-and-effect). The danger of pseudo-experimental comparison rests in the many confounding factors that are not equaled between the groups. For example, weapon presence may be confounded with type of crime (robbery versus shoplifting) and therefore also with factors such as differential witness attention, witness stress, duration of contact, quality of culprit description, and delay prior to lineup.

The reverse of this is true as well: a pseudo-experimental comparison that shows no difference between groups cannot guarantee that a cause-and-effect relationship is not present. For example, the absence of a weapon focus effect in the Illinois data cannot support a conclusion that a weapon does not (or does) affect eyewitness accuracy. Mecklenburg’s claims of “far-reaching implications of
this field data are based on pseudo-experimental comparisons that cannot challenge results from controlled laboratory experiments, either logically or empirically. The irony is that the “real-life” cases that may at face value seem to offer such compelling information in fact offer little concrete knowledge due to real life’s inherent confounds.

**Investigative Practice: More Precise Recording of Witness Responses in Identification Tasks**

Scientists have long argued that blind lineup procedures are more likely to yield specific and fine-grained lineup reports that are more faithful to actual witness identification responses. Specifically to the Illinois Program, one of the suspicions about the very low filler pick rate (zero) in the non-blind simultaneous lineups of Evanston and Chicago was that the lineup administrators perhaps did not effectively differentiate between filler picks and no-choice responses. In effect, this means that a filler pick may have been recorded as “failure to identify” by a detective who knew who the suspect was in the lineup. The Evanston data do not speak directly to this point. However, only the double-blind sequential lineup protocol required the detective to differentiate witness responses — a pick, no choice, or all ruled out — and without knowledge of the suspect’s identity or position in the lineup. Verbatim witness comments were significantly more often recorded in the DB-SEQ lineup reports (83 percent vs. 39 percent), a qualitatively different and arguably more precise type of statement (“that’s her, she’s got silver teeth”). The NB-SIM lineup files show a significantly greater prevalence of detective third-person paraphrasing of witness response (“the victim made a positive suspect ID”) than do the DB-SEQ lineups (“the witness stated ‘Number 3 is him’”). Lineup outcomes also were more often recorded with the DB-SEQ protocol (all DB-SEQ lineup outcomes were reported; 8 percent of NB-SIM outcomes were missing from the report). The design of the study prohibits a conclusion regarding a specific causal relationship between lineup procedures and recording quality (and in fact the double-blind sequential lineup protocol required this level of recording whereas the non-blind simultaneous lineup protocol did not). More to the point, however, is that a blind administrator is simply not in a position to interpret the witness’s pick or to slant the witness’s comments toward an investigative hypothesis because the blind administrator does not know which lineup member is the suspect. The Evanston files offer the first empirical data consistent with the notion that blind lineup procedures in the field can provide a qualitatively different (and arguably more precise) evidence record.

**Investigative Practice: Repeated Lineups Present Problems for Veracity of Eyewitness Evidence**

One strong impetus for the FOIA request was the absence of information in the Mecklenburg Report about each eyewitness’s identification history. As reported above, Illinois police sometimes conduct a confirmatory physical lineup after the witness has identified a suspect from a showup and/or photo array, the same witness producing both the first identification during the investigation and the live identification evidence that may be required by the court. Knowing this, questions arise as to whether and how confirmatory lineups were counted in the Illinois data analysis. As is now apparent, confirmatory lineups were included in Evanston, as were earlier lineups by the same witness (double-counting of these witnesses).

In a best-case scenario, a live lineup is an opportunity to recheck the witness’s earlier photo identification decision. A photo will not fully capture all aspects of a suspect’s appearance, such as stance, gait, height, and weight. The witness may take a look at the live suspect and confirm or retract an earlier ID (“no, that’s not him”). In the latter circumstance, there is a logical fail-safe mechanism in the requirement that a witness identify a live suspect. However, that fail-safe mechanism does not work so cleanly in the case of a confirmation.

Consider the Georgia case of John Jerome White, who served 22 years of a life sentence for rape before he was exonerated by DNA evidence in 2007. The victim was “almost positive” when she selected White from an initial photo array; she later positively identified him from a live lineup. Her identification was remarkable in that the police had unknowingly placed her real attacker in the physical lineup. The victim did not recognize her true attacker and instead identified White.

Thus, the broader issue for the Evanston lineups (and lineup practice in general) is whether a repeated identification task is in fact a true test of eyewitness memory for the culprit observed at the crime scene. People are much better able to recognize faces than they are able to remember where they saw them, and a familiar face at the second lineup may stem from the witness’s exposure to a photo or showup rather than the crime scene. Also, the fact that all fillers have changed around the same suspect from first to second lineup is highly suggestive; the common denominator across lineups may be detected by the witness. At minimum, a witness may feel a commitment to identify the same suspect that was identified earlier. Recent laboratory examination of the effects of repeated lineups found that risk to an innocent suspect increases sharply at the second lineup. In addition, a second lineup inherently involves some time delay between identification tasks, thereby increasing the likelihood of witness exposure to confirmatory feedback about the earlier identification or additional case-related information. Both laboratory and field research have established the powerful impact on eyewitness confidence and memory when post-identification feedback bolsters the witness’s decision.

Methodologically, confirmatory identification tasks should be separated from assessments of eyewitness memory for a first identification of a stranger. And, the impact of undetected confirmatory (and verification) lineups must be evenly distributed between lineup conditions via random assignment. The broader issue for police practice and legal policy is the potential dangers that repeated identifications bring to investigations and to the veracity of eyewitness evidence.

**Conclusion**

The Illinois Pilot Program attempted to evaluate the performance of double-blind sequential lineups in comparison to traditional simultaneous lineup practice. The methodological shortcomings of the Program were apparent in 2006 when the Mecklenburg Report was released, and have been catalogued by scientists since that time. In particular, Schacter et al. judged the central design confound (comparing double-blind sequential lineups to non-blind simultaneous lineups) to be fatal for the study’s effectiveness in answering its primary research question. Lack of random assignment is likewise a devastating design error. The Evanston data now allow us to see just how this
design deficit played out in its effects on the study outcomes. Importantly, the Evanston data provide empirical evidence that the non-random assignment of lineups to experimental conditions does matter, that this design problem biased the Evanston outcomes and makes them uninterpretable and essentially useless. Furthermore, all three cities involved in the Illinois study operated with these two insurmountable design errors.

Two additional aspects of lineup practice relevant to the quality of eyewitness evidence have been forcefully brought to our attention through the Evanston data: (1) revised lineup procedures can produce stronger documentation of verbatim eyewitness responses (facilitated by blind administration); and (2) repeated identification tasks may pose substantial challenges for the fidelity of eyewitness evidence. These topics deserve more research and legal attention.

Finally, there is a positive note. Rigorous field testing of lineup practice can be done. Importantly, field experiments overseen by the American Judicature Society and currently underway in Tucson, San Diego, and Austin, Texas, bring exacting experimental design to the test of sequential versus simultaneous lineups. At the same time, laboratory research on lineup procedure has expanded dramatically over the last decade, and now provides a more comprehensive knowledge of lineup format effects.” These developments signal an increasingly useful intersection of eyewitness science and legal policy.

Notes


15. Complaint, National Association of Criminal Defense Lawyers v. Superintendent of the Chicago Police Department, Chief of the Evanston Police Department, Chief of the Joliet Police Department, and Director of the Illinois State Police, No. 07 CH 3622 [hereinafter NACDL v. Chicago P.D.]. The pleadings, appellate briefs, and appellate court decision (NACDL v. Chicago P.D., III. App., No. 07 CH 3622) are available at http://www.nacdl.org/sldocs/nsf/Issues/eyewitnessid?OpenDocument; see also Norman L. Reimer, Obtaining Underlying Data From the Illinois Report, The Champion, March 2007 at 43; Norman L. Reimer, Removing a Roadblock to Reform, The Champion, April 2010 at 7; N. Steblay, Affidavit of Nancy Steblay (2007a), September 24, NACDL v. Chicago P.D. Note: Chicago and Joliet Police Departments have fought the lawsuit, although a recent ruling favors NACDL and the redacted data from these two cities apparently will be forthcoming. In this ruling, the appellate court stated that “the public interest at stake in obtaining these documents is significant, both for the people of Illinois and for people across the country who are considering the results of the Illinois study. … If an examination of the data used in the study … discloses flaws in the study methodology or design, reform based on the sequential, double-blind identification procedures is still a possibility in Illinois, as well as in other states that have been conducting such reform.”


18. See the Meeckelburg Report.

19. Doyle, et al., supra note 7; Wells, supra note 8.


22. Steblay, supra note 16.

23. See the Meeckelburg Report.

note 9.
25. Id.
26. Id.
28. See Wells, supra note 2, Wells, supra note 8, O’Toole, supra note 7, and Mecklenburg, Addendum, supra note 21.

About the Author
Nancy Steblay is a professor of psychology at Augsburg College. Her research interests are psychology and law, with an emphasis on eyewitness accuracy, pretrial publicity, and inadmissible evidence and judicial instructions to disregard. She served as NACDL’s consultant in its efforts to obtain the data underlying the Illinois study described in this article.

Nancy Steblay, Ph.D.
Augsburg College
2211 Riverside Avenue South
Memorial Hall 331A
Minneapolis, MN 55454
612-330-1201
E-MAIL steblay@augsburg.edu

Street Crimes, Stress, and Suggestion
Continued from page 29
11. State v. Young, 35 So.3d 1042 (La. 2010).
12. Id. at 1050.
13. Id.
17. See Brades v. State, 614 S.E.2d 766 (Ga. 2005); Commonwealth v. Santoli, 680 N.E.2d 1116 (Mass. 1997). Counsel should also consider filing a motion in limine pre-trial to preclude the government from eliciting the witness’s certainty about the identification as both irrelevant and more prejudicial than probative.
18. See Kyles v. Whitley, 514 U.S. 419, 446 & n.15 (1995) (approving of tactic of discrediting sloppy police methods and noting: “When … probative force of evidence depends on circumstances in which it was obtained … indications of conscientious police work will enhance probative force and slovenly work will diminish it.”).
19. Some of these shortcomings are addressed in Clopten, supra note 10.
20. The model cross-examinations in this section come from a presentation by Timothy P.O’Toole. He also provided guidance in helping me think through the content and structure of this article. ■