Notes on the Illinois Pilot Program on Sequential Double-Blind Identification Procedures

Roy S. Malpass

As a result of recommendations made by the Illinois Governor's Commission on Capital Punishment, the Illinois Legislature charged the Illinois State Police with conducting a pilot program to evaluate the effectiveness of the sequential, double blind identification procedure in the field. Sheri H. Mecklenburg was appointed Director of the Illinois Pilot Program and undertook to design the Illinois Pilot Program, seeking comments and approval from eyewitness researchers in the process. Reporting forms were developed, police personnel were given training on the new procedures and procedures were developed for deciding which lineups would be presented according to traditional or new procedures. These matters and much more are detailed in the Report to the Legislature of the State of Illinois: the Illinois Pilot Program on Sequential Double-Blind Identification Procedures ("the Report").

The author was approached by Mecklenburg, asking for our participation as analysts. I agreed to act in this capacity with the assistance of Laura A. Zimmerman, Stephen J. Ross, Lisa D. Topp, Vanessa Uribe, Dannette De Leon, Sarah Ramirez and Jessica Belisle, all members of the Eyewitness Identification Research Laboratory at the University of Texas at El Paso. Periodically we received sets of case reports from the three participating jurisdictions. We were given a free hand to structure our analysis in our own way. We constructed the code book and implemented an analysis. While we contributed our analysis of the data, we did not participate in writing the Report.

Professor Ebbesen of the University of California, San Diego also agreed to serve as an analyst for the Pilot Program. Professor Ebbesen and his research group received the same case reports and constructed their own way of coding and analyzing the data. Professor Ebbesen’s group and the Eyewitness Lab at University of Texas at El Paso reached the same conclusions, although our conventions for coding the raw field reports for analysis differed in some respects, leading to somewhat different numbers. Ebbesen and Malpass never discussed anything about their task - had no conversation whatever - until they met during the Symposium held at the Loyola University of Chicago Law School on April 21, 2006.

Design

The study was designed to determine whether or not a new eyewitness identification procedure (a particular variant of double-blind sequential lineup) is superior to the simultaneous lineup procedure in current use. The specifics of implementation of the design are discussed in the Report. This study was not the extension of an academic research program and was not undertaken to untangle theoretical issues.

Results

The major results are displayed in Table 1, for the total sample, aggregating the results across the three jurisdictions. There are three outcomes possible in this study: suspect identifications, filler identifications and non-identifications. It is important to note that suspect identifications cannot be interpreted as either correct or false identifications, and non-identifications cannot be interpreted as missing the offender or as rejecting a lineup that does or does not contain the actual offender. It is not known, for any lineup in this study, whether the suspect in the lineup is the actual offender. This can be known in laboratory studies, but not in the field without a considerable amount of additional research. We will return to this matter below.

The major results are these:

(Malpass, continued on page 6)
Witnesses who viewed a simultaneous lineup identified the suspect more often than those witnesses who viewed a sequential lineup (suspect identification rates of 59.9 percent and 45 percent respectively).

Witnesses who viewed a simultaneous lineup chose a filler less often than those who viewed a sequential lineup (filler identification rates of 2.8 percent and 9.2 percent respectively).

Witnesses who viewed a simultaneous lineup were less likely to choose no one than were those who viewed a sequential lineup (no identification rates of 37.6 percent and 47.2 percent respectively).

As noted above, these results cannot be interpreted directly as accurate or erroneous responses. Nonetheless, assuming that the increase in non-identifications for sequential lineups compared with simultaneous lineups reflects a proportionate increase in correct rejections in a culprit-absent lineup, and that the decrease in suspect identifications from simultaneous to sequential lineups is proportionate with a decrease in correct identifications in a culprit-present lineup, then the sequential advantage for culprit-absent lineups will be more than offset by the sequential disadvantage for culprit-present lineups. This comparison is worsened if one considers that culprit-present lineups are probably the more frequent. It seems implausible that on the average law enforcement does not do better than a .5 probability of getting the right person in the lineup.

Reasonable people can begin with different assumptions, however. The proportions of suspect identifications contributed to correct and false identifications can be argued, and various probabilities that the culprit is actually in the lineup can be entertained.

Stability of the findings across jurisdictions is a matter of interest from the perspective of application. These findings are displayed in Table 2.

For simultaneous lineups, suspect identifications vary over a range of 10.7 points, from 57.0 to 67.7, and non-identifications vary over a range of 10.1 points, from 32.3 to 42.4.

For sequential lineups, suspect identifications vary over a range of 42.7 points, from 25.9 to 68.6, and non-identifications vary over a range of 34.4 points, from 28.6 to 63.0.

The difference between simultaneous and sequential lineups also varies considerably, from +41.8 to -7.3.

Sequential lineups appear to be more sensitive to differences in jurisdiction / location / context / background conditions, although it is not clear exactly what conditions these might be.
Discussion and Interpretation

The Illinois Pilot Program\(^7\) is a landmark eyewitness identification study, even among field studies:

- The Illinois Pilot Program makes a direct comparison between the traditional, intact simultaneous lineup procedures and a version of double-blind sequential lineups.
- It uses multiple jurisdictions.
- It contains more than 700 individual identifications.
- The criminal cases cover the entire range of crimes committed during the period of the study.

Field data are inherently noisy. Field studies are known for variability, and this is why laboratory studies are sometimes called “controlled” studies by way of contrast. There are many investigators, many contexts and many jurisdictions, and these lead to many variations in implementation. In some respects the noisy background may obscure relationships in the data that might be found under otherwise more controlled conditions. On the other hand, strong effects showing through the background variation would be robust. Additionally, confounding factors outside of the research design proposed as having an effect on study outcomes would also have to be strong (substantial empirical effect size), consistent and detectable to be taken seriously. Further, the noisy study environment is a valid reflection of the environment of application because it IS the environment of application.

It may take some time to frame new questions arising out of our attempts to interpret these results. The questions will lead to new and more informative research - certainly in the laboratory - and hopefully in field studies carried out in association with law enforcement. This is a very rich intellectual welfare program for researchers.

Clearly the problems with eyewitness identifications have not been solved, and as Barry Scheck pointed out in his remarks on the eve of the Loyola Conference, we should move forward to develop other areas of lineup reform while we clarify the contribution of sequential lineup presentation. Working relationships between law enforcement and academic researchers should be strengthened to study a range of identification questions.\(^8\)

Transposition from field categories to the categories of laboratory studies. An important thing to note in the interpretive process is that the three outcome categories of the field study cannot easily be disaggregated into the six (at least) categories of laboratory studies.

It would be possible to disaggregate the field study categories under two conditions: If we make assumptions about (1) the proportions of each field category to be distributed to each of the two cognate lab categories, and (2) the proportion of those figures to be considered, reflecting the \textit{apriori} probability of the suspect being the perpetrator, or not.

Filler Identifications. Filler identifications are the only responses that have an apparent clear interpretation. The only thing that can be said, really, is that sequential lineups attract a non-trivially greater frequency of filler identifications, overall. This is descriptively true for all three jurisdictions, but statistically reliable for only two. The absolute percentage of filler identifications is small. There does not appear to me to be a theoretically solid way to use this result to make inferences about the interpretation of the real interest of this study: the accuracy of suspect identifications and non-identifications.

Double-blind simultaneous lineups as a comparison. The purpose of the study, as stated above, was to determine whether or not a new eyewitness identification procedure (a particular variant of double-
blind sequential lineup) is superior to the simultaneous lineup procedure in current use. However, there are some nuances to the question.

First, to evaluate the effects of a change in practice against existing practice is a completely appropriate research strategy. To have changed the existing simultaneous lineup practice for purposes of comparison against a new practice would have given no guidance to law enforcement at the end of the study because it would have had little to do with their current practice. This was a simple and straightforward study design for a straightforward question.

Second, perhaps there is an interest in the ultimate question of whether some version of a double-blind sequential presentation is superior to some version of a double-blind simultaneous presentation. That is a different question, and it is not the one asked in the Illinois Pilot Program. To answer that question would require different research design strategies. But more than that, it would place law enforcement agencies in Illinois in the position of doing research to answer our (academic researchers) questions rather than their own. I would like it very much if I could establish the Field Study Unit of the Eyewitness Identification Research Laboratory within the law enforcement community of the State of Illinois, but this project could not serve in that capacity.

The comparison between blind sequential and non-blind simultaneous lineups is confounded. A confounded comparison exists when there is at least one difference between two research conditions in addition to the difference introduced for study. In the Illinois Pilot Program, sequential lineups are uniquely blind while simultaneous lineups are uniquely not blind. So there are two factors (at least) differentiating them: mode of presentation and blind vs. not blind. The question is to what degree the blind vs. not blind difference hinders interpretation of the presentation mode difference. On the one hand, confounded comparisons are the rule for field studies, and when the very powerful research strategies possible with true experiments are not present, investigators are left with trying to reason their way through the interpretation of observed differences. Un-confounding requires that the important factors are known and either controlled or measured for post-hoc analysis. There may be many confounding factors in this study with effects that cannot be anticipated or estimated. It was not possible to monitor the administration of even a...
The interpretation of witness confidence is as much a matter for appropriate advice to jury members and lawyers as for identification procedures. However, such advice does not always cause jurors to be suitably sceptical of tenuous eyewitness evidence. In a well-known case in England, Barry George was convicted of the assassination of Jill Dando, a famous TV presenter. The main prosecution evidence against him was that he was identified during the identification procedure by only one of 16 witnesses who attended a lineup. The witness who identified Barry George had not witnessed the murder but saw a man in the street about four hours before the murder. By her own testimony she saw his face for five to six seconds. The identification procedure was held approximately 18 months after the murder. Nevertheless she was a highly convincing witness in court. The judge gave the appropriate ‘Turnbull’ warning to the jury about the problems of eyewitness identification evidence. The jury convicted by a 10-1 majority. The conviction was upheld on appeal.

Tim Valentine is a professor of psychology at Goldsmiths, University of London. Stephen Darling is a research fellow at the University of Edinburgh. Amina Memon is a professor of psychology at the University of Aberdeen. The research on video identification was funded by the Nuffield Foundation (Grant no. OPD/00227/G). Correspondence may be addressed to Professor Tim Valentine Department of Psychology, Goldsmiths, University of London, New Cross, London SE14 6NW; email t.valentine@gold.ac.uk.

1 Tim Valentine is a professor of psychology at Goldsmiths, University of London. Stephen Darling is a research fellow at the University of Edinburgh. Amina Memon is a professor of psychology at the University of Aberdeen. The research on video identification was funded by the Nuffield Foundation (Grant no. OPD/00227/G). Correspondence may be addressed to Professor Tim Valentine Department of Psychology, Goldsmiths, University of London, New Cross, London SE14 6NW; email t.valentine@gold.ac.uk.

On the other hand, interpretation of the presentation mode differences are still possible even though interpretation requires careful evaluation. The observed differences between study conditions are substantial. To be taken seriously a “confound” theory would have to be convincing based on known empirical effect sizes. Regarding the blind vs. non-blind confound, there are two matters (at least) for discussion.

First, it would be of interest to estimate the relative magnitude of the effects of blind administration on simultaneous vs. sequential lineups. My discussions with the other scientists on the National Institute of Justice Technical Working Group reflected an assumption that packaging blind administration with sequential was more important because sequential lineups are inherently more susceptible to inadvertent influence than simultaneous lineups— that packaging “blind” with sequential was a requirement whereas packaging “blind” with simultaneous provided little added value. If there is an empirical base for this supposition then we will have a way to evaluate the importance of the “blind” confound. It does not appear, however, that we can support this assumption at present.

Second, while the research on experimenter bias and the communication of hypotheses cannot be ignored, recent research comparing double blind vs. not blind conditions has not been encouraging. One study finds an increase in mistaken identifications resulting from administrator knowledge of the identity of the suspect. Further, Haw and Fisher found contact between the administrator and the witness to increase mistaken identifications, especially in simultaneous lineups. But that result does not directly address the double blind vs. not blind comparison. Two additional studies found no effects in double blind vs. not blind comparisons. The effect is found under some conditions, and not others. It is not obvious how strong a “blind” effect could be claimed, and how the “blind” confound would be used to interpret the results of the field study.
The laboratory research underlying the sequential lineup itself is no model of virtue. It should be noted that in the foundational studies that were included in the Steblay et al. meta-analysis, none of the researchers, including advocates of the sequential technique, studied the effects of blind administration on the sequential vs. simultaneous lineup comparison. The literature underlying the meta-analysis is deficient in a number of ways, and this is merely one of them.

The research underlying the sequential lineup itself contains important variables that confound the simultaneous vs. sequential comparison. For decades researchers advocating the superiority of “sequential” lineups have been content to use a package of variables under the label “sequential lineup” without examining the effects of the other members of the sequential package that confound the simultaneous vs. sequential comparison. Recent research indicates that when these confounding factors are exchanged between simultaneous and sequential procedures the superiority of sequential lineups in laboratory studies disappears. It may be that sequential presentation of persons/faces is actually not the factor responsible for the “sequential superiority” effect. If this is true, and it certainly deserves more study, then it would be useful to know just which attributes(s) of any lineup procedure produce its advantages so that those can be the focus of application, while the non-effective components can be left behind.

What, then do we learn from the Illinois Lineup Project?

There are some preliminary truths that we knew beforehand: Field studies as this one produce noisy data. Everyone who has ventured into an evaluation of a field project knows that. Moreover, when the evaluators do not have the ability to make primary observations/measurements on site, the data are noisier. Questions that should be considered include: (1) Is there more noise in the data resulting from previous well practiced procedures or from new procedures in which a small amount of training was received? and (2) Which data set would be expected to have a larger within set variance? The data from the Illinois Pilot Program show substantially greater variation across jurisdictions for the novel procedure.

The important differences observed were statistically reliable. A non-pejorative theory of how one data set’s outcomes would be facilitated or the other suppressed, apart from the effects of the procedures being compared, is difficult to generate. Plenty of pejorative interpretations can be developed, however, including imputing insufficient devotion to serious academic analysis, or conspiracy to shape the findings by the officers running the identifications. People have their own beliefs about these matters, and about the trustworthiness of law enforcement personnel in such contexts. It would help if there had been manipulation checks on these points, but there were not. It may be extremely difficult to implement such oversight - partly because of personnel costs. In the end, it would introduce additional ambiguities because the context of ongoing implementation of the procedures would not be done with an eyewitness researcher looking over the investigators shoulder. As a consequence the results, could not be generalized from the study environment to the future environment of application - one without researcher monitoring. In the absence of information stronger than suspicion and distrust, it is best to take the findings as reflecting good faith efforts by the participants.

It has to be said that not much good news will be found in the Illinois Pilot Program for those who advocate sequential lineups to law enforcement and governmental organizations. Simultaneous lineups lead to substantially more suspect identifications than do sequential lineups, while sequential lineups lead to more non-identifications than simultaneous. But these are not completely offsetting: the advantage of simultaneous lineups over sequential lineups for suspect identifications is greater than the advantage of sequential lineups over simultaneous lineups for non-identifications.

There are factors which might affect the importance of the non-offsetting result. If these two outcomes have different values in the sense of desirability or utility, then one outcome may become more important than the other and weigh more in the offsetting equation. But in order to do those calculations one must disaggregate the field categories into their cognate laboratory categories and make assumptions about
SYMPOSIUM ISSUE

(Malpass, continued from page 40)

the relative frequencies of the perpetrator actually being present in the lineup, or not.

Is there a discontinuity between the corpus of laboratory research on simultaneous and sequential lineups and the field studies?

There is no discontinuity. First, it must be understood that the field data cannot be resolved into the categories of laboratory studies (figure 1): All that can be observed in this study are identification rates. The field study findings indicate that sequential lineups are associated with a lower rate of lineup choices (suspect identifications) than are simultaneous lineups (45 percent < 62.7 percent). The Steblay et al. meta-analysis of laboratory studies shows a similar result.

For known perpetrator present lineups, sequential lineups lead to fewer lineup choices (54 percent < 74 percent) and for known perpetrator absent lineups (with designated suspects) sequential lineups also lead to fewer lineup choices (28 percent < 51 percent). It is therefore not the case that one can invest oneself in either lab or field studies as the definitive source of truth and reject the other. Their results are very similar for the level of data aggregation at which they can be compared. This consistency is in fact a vindication of the laboratory-based experimental research strategy. We did not know beforehand that we could extrapolate from the corpus of laboratory studies to answer the question posed in the Illinois Pilot Program, but we are on firmer ground given the Illinois results. Most important, however, is the assurance that questions raised in the context of application can be examined in laboratory environments with a high likelihood of relevance for the application environment.

1 Roy Malpass is a professor in the Department of Psychology and the Criminal Justice Program at the University of Texas at El Paso. He was a panelist for New Policies, New Practices: Fresh Perspectives on Eyewitness Identification at Loyola University Chicago School of Law on April 21, 2006. Correspondence may be addressed to Department of Psychology University of Texas at El Paso, El Paso, TX 79968; email rmalpass@utep.edu; phone (915) 539-0510.

(Norton, continued from page 2)

not be present at the procedure. While the Supreme Court of Connecticut claimed broad inherent supervisory authority over the administration of justice, it did not claim that it could order law enforcement agencies to follow particular identification procedures. Rather, it found its authority in providing guidance concerning jury instructions. It directed that, if the administrator of an identification procedure fails to warn the witness, the jury must be instructed that such a failure “may increase the likelihood that the witness will select one of the individuals in the procedure even when the perpetrator is not present. Thus, such action on the part of the procedure administrator may increase the probability of a misidentification.”

The Supreme Court of Wisconsin reached a similar conclusion in State v. Dubose. The court there held that showups would violate state due process if “unnecessarily” suggestive. Necessity is to be determined by the “totality of the circumstances,” but the court specifically held that “a showup will not be necessary . . . unless the police lacked probable cause to make an arrest or, as a result of other exigent circumstances, could not have conducted a lineup or photo array.” Beyond dictating a specific preference for lineups and photo arrays over showups, the Wisconsin Court also urged that for showups police should follow procedures similar to those proposed by the Wisconsin Innocence Project, including warning the witness that the perpetrator may or may not be there. Clearly the court was leaving open the possibility that, under the rubric of assessing whether the procedure as “unnecessarily” suggestive, it might review the specific identification procedures used.

III. Expert Witness Testimony

Until recently, eyewitness expert testimony was rarely allowed. As one Tennessee court put it, “[W]e are of the opinion that the subject of the reliability of eyewitness identification is within the common understanding of reasonable persons. Therefore, such expert testimony is unnecessary.” However, a recent writer has observed “eyewitness expert testimony has become more frequent in recent years. Since 2002, a number of published federal court opinions have allowed or upheld the admission of eyewitness expert
SYMPOSIUM ISSUE
(Malpass, continued from page 41)

3 Id.
4 Report to the Legislature, supra note 2.
5 W.A. Geller, Suppose we were really serious about police departments becoming “learning organizations”? 234 NAT’L INST. OF JUST. J. 2-8 (1997).
11 S.M. Greathouse & M.B. Kovera, Single vs. double blind lineup administration’s effects on identification accuracy, witness confidence, and administrator behavior (unpublished manuscript, on file with the John Jay College, City University of New York).
14 Id.
17 Id.
18 Steblay, supra note 15.

The Center for Modern Forensic Practice of the John Jay College of Criminal Justice of the City University of New York will hold a conference Off The Witness Stand: Using Psychology In The Service of Justice marks the 100th anniversary of the publication of Hugo Munsterberg’s On The Witness Stand, the pioneering effort to introduce the lessons of witness psychology to the legal system. The conference will be held in New York on March 1-3, 2007. Speakers will include Hon. Janet Reno, Elizabeth Loftus, Gary Wells, Saul Kassin, Elizabeth Klobuchar, Steven Penrod and James Doyle.