How Can Psychological Science Enhance the Effectiveness of Identification Procedures? An International Comparison

Tim Valentine  
*Prof. of Psychology, Goldsmiths University*

Stephen Darling  
*Research Fellow, University of Edinburgh*

Amina Memon  
*Prof. of Psychology, University of Aberdeen*

Follow this and additional works at: [http://lawecommons.luc.edu/pilr](http://lawecommons.luc.edu/pilr)  
Part of the Criminal Law Commons, and the International Law Commons

Recommended Citation
Recent experiments have studied and compared eyewitness identification procedures including the use of moving video versus photographs, culprit-absent lineups versus culprit-present lineups, and foil selection by culprit description versus selection by suspect resemblance.

How can psychological science enhance the effectiveness of identification procedures?
An international comparison.

Tim Valentine, Stephen Darling and Amina Memon

The sequential double-blind method protects the guilty, moving video images protect the innocent (a little), but foil selection strategy makes no difference.

The reliability of eyewitness identification has attracted concern from the legal profession in England for at least 100 years. In 1904 a committee of enquiry was established to investigate the trials of Adolf Beck. Incredibly, on two separate occasions Adolf Beck was wrongly convicted on the basis of mistaken eyewitness identification. In both trials, multiple eyewitnesses identified Beck as a confidence trickster who stole jewellery from them. The crimes were subsequently found to have been committed by William Wyatt. The 1904 Committee of enquiry led directly to the establishment of a Court of Appeal.2

Concern about further wrongful convictions based on mistaken identification led to a government enquiry into the reliability of eyewitness identification evidence, chaired by Lord Devlin, which reported in 1976.3 The Devlin report led directly to a landmark judgement in the English Court of Appeal, which established a requirement that in cases of disputed identification the trial judge must caution the jury about the dangers of eyewitness identification evidence. The judge should point out that confident eyewitnesses may be mistaken and instruct the jury to consider carefully the circumstances of the identification.4

From this historical perspective, it is unsurprising to learn that mistaken eyewitness identification is also a major problem for the United States courts. Nevertheless, the extent of the problem has proved to be greater than many may have anticipated. The work of the U.S. Innocence Project, which to date has led to 183 prisoners being exonerated by new DNA evidence, found that mistaken eyewitness identification was a factor contributing to three-quarters of the original wrongful convictions.5

Recent developments to eyewitness identification procedures

Eyewitness identification procedures used in the United States and the United Kingdom have some important differences. In the United States, live lineups and identification from arrays of photographs are both frequently used to collect formal eyewitness identification. Traditionally, in the United Kingdom all formal eyewitness identification evidence has been obtained from live lineups. Identification from arrays of photographs has never been permitted as a formal means of identification. Over the last few years video has replaced almost all live lineups. This innovation has been made possible by development of sophisticated computer systems used to compile video lineups from a standardised database of moving video clips.

Recently identification procedures in the United States have been the subject of consultation with eyewitness researchers. Identification from arrays of photographs is still widely used, but the U.S. National Institute of Justice set up a Technical Working Party for Eyewitness Evidence to review procedure and produced a guide to best practice.

(Valentine, continued on page 22)
Recent developments to identification procedures in the USA

The U. S. National Institute of Justice document *Eyewitness identification: A guide for law enforcement* ("the Guide")\(^6\) includes the following guidance:

- The foils in a lineup should be selected to generally match the witness’ description of the culprit.
- There should be a minimum of five foils.
- The witness should be advised that the culprit may or may not be in the lineup.
- The witness should state in their own words how confident they are of any identification.

Two methods of lineup presentation are endorsed by the Guide: 1) a *simultaneous* lineup, in which the witness is permitted to inspect all of the photographs or lineup members before making an identification and 2) a *sequential* lineup, in which the witness sees one photograph or person at a time and makes a decision prior to viewing any other photograph or person.\(^7\) The guide does not express any preference for one method over the other. The procedures mentioned here do not form an exhaustive list of the provisions in the Guide. It should be noted that the guidance is a recommendation of best practice and has no direct legal force.

In an earlier ‘white paper,’ written under the auspices of the American Psychology - Law Society ("AP-LS"),\(^8\) psychologists had advocated that the person who administers a lineup should not know which person in the lineup is the police suspect. That is to say that the administrator should be ‘blind’ to the identity of the suspect. This procedure is known as ‘double-blind’ as neither the administrator nor the witness has prior knowledge of who the suspect is in the lineup. This measure was strongly advocated by researchers because it removes all possibility of the witness being influenced by the lineup administrator. Such influence can be very subtle and may occur without any intention or awareness of either the administrator or the witness. The double-blind procedure is well established as an important aspect of scientific enquiry. For example, neither the patient nor the clinical staff should know which patients received a placebo in a drug trial. A recommendation of the double-blind method is conspicuously absent from recommended best practice in the Guide on eyewitness identification.

Research based on identification from photograph arrays suggests that mistaken identification can be reduced by sequential presentation of the photographs as outlined in the Guide.\(^9\) However, the Guide did not include the important stipulation of a ‘sequential double-blind method.’ Under sequential presentation instructions the witness should make a decision after viewing each photograph as to whether he or she is the culprit. If the witness rejects the photograph they are shown the next photograph. The procedure stops when the witness makes an identification. The method endorsed by researchers crucially stipulates that the witness should not know how many photographs are in the lineup, the witness is given unbiased instructions (e.g., that the person they saw may or may not be in the lineup) and, importantly, that the administrator is blind to the identity of the suspect.\(^10\)

**Video identification** has a number of important benefits [including]. . . dramatically reduce[ing] the delay before an identification can be organized, . . . usually produc[ing] a video lineup within two hours of request, . . . [has] a large database of video clips from which to select foils, . . . and [employs] a laptop which can be taken to a witness who is unable to attend the police station.

Sequential presentation is believed to reduce mistaken identification by reducing the opportunity for the witness to make a relative judgement. In the traditional simultaneous presentation, a witness who believes that the culprit is in the lineup may identify the person who most looks like the person they saw, having had the opportunity to view all the photographs in an array. Sequential presentation aims to prevent relative judgements by forcing the witness to make
independent judgements to each lineup member. Sequential presentation has been adopted in some jurisdictions in the United States. However, in some cases the strict procedure advocated by researchers has not been followed in all of its aspects. It is worth noting that researchers did not include sequential presentation amongst the recommendations of the AP-LS white paper.11

Recent developments to identification procedures in England & Wales

The Police and Criminal Evidence Act of 1984 ("PACE") which applies in England and Wales (but not in Scotland or Northern Ireland), includes a code of practice for identification by eyewitnesses ("code D"). The code can be revised without the need for new primary legislation. In recent years the code has been revised on an annual basis. The current code of practice (2005)12 includes the following provisions:

- A lineup that includes one suspect must consist of at least eight foils.
- The foils must resemble the suspect in age, general appearance and position in life.
- The suspect has the right for their legal representative to be present during the identification procedure.
- The person who administers the lineup cannot be involved in the investigation of the case (but note that the administrator does know who the suspect is).
- Witnesses must be advised that the person they saw may or may not be present.
- Witnesses must be advised that if they cannot make a positive identification they should say so.
- Witnesses must view each member of the lineup twice before making any identification.
- Video identification should be used unless there is a reason why a live identification is more appropriate.

Although the code of practice does not have statutory force, trial judges have the discretion to exclude or allow eyewitness identification evidence. Therefore police forces have systems in place to demonstrate compliance with the code.

Two different IT systems are in widespread use in British police forces to provide video identification. VIPER™ (Video Identification Procedure Electronic Recording) and PROMAT™ (Profile Matching).13 The systems produce similar formats of video lineup, but each has its own database of images. Lineups consist of 15 second clips of each person shown one after another. The sequence starts with a head and shoulders shot of the person looking directly at the camera, who slowly turns their head to present a full right profile to the camera. The person then slowly rotates their head to present a full left profile to the camera. Finally the person returns to looking directly into the camera in a full-face pose.

Research on video identification

Research has demonstrated that VIPER video lineups from real criminal cases were fairer to the suspects than conventional 'live' lineups,14 and that VIPER video lineups were equally fair to white European and African–Caribbean suspects.15 In these studies, participants (known as 'mock witnesses') were shown a set of videos of VIPER lineups or a set of photographs of live lineups held as part of the investigation of the case. For each lineup they were given the first description of the offender made by the original witness. The mock witnesses were required to choose, on the basis of the witness' description, the lineup member who they think is most likely to be the police suspect. Therefore, a 'mock witness' simulates a witness who (a) has no memory of the culprit at the time of the identification procedure; (b) can remember the description they previously gave to the police and (c) nevertheless, makes an identification from the lineup. If the lineup is perfectly fair, and all members fit the description, the mock witness would have no basis on which to make their selection and would merely have to guess who is the suspect. Therefore, if a large number of the mock witnesses are asked to make a selection they would select the suspect on 11 percent of occasions (1 in 9) from each lineup, because the lineups all contained a suspect and eight foils.

(Valentine, continued on page 24)
(Valentine, continued from page 23)

Using this procedure 25 percent of mock witnesses (1 in 4) chose the suspect in the live lineups, more than expected by chance (25 percent vs. 11 percent).16 In comparison, 15 percent of mock witnesses selected the suspect from the videos of VIPER lineups.17 Statistical analysis showed that the VIPER lineups were significantly fairer than the live parades (15 percent vs. 25 percent), and the VIPER lineups were not significantly less fair than expected by chance (15 percent vs. 11 percent).18

Previous data from real cases suggested that live lineups may be less fair to ethnic minorities than to white Europeans. Therefore, the fairness of VIPER lineups of African–Caribbeans and of white Europeans was compared. The VIPER parades were found to be equally fair to suspects of both ethnic groups.19

Benefits of video identifications

Video identification has a number of important benefits compared to live lineups. First, use of video can dramatically reduce the delay before an identification can be organized. Live lineups have been subject to long delays to enable a selection of appropriate foils to be available to stand on a lineup (typically of one to three months).20 In contrast, VIPER can usually produce a video lineup within two hours of request. Second, approximately 50 percent of live lineups in England and Wales were cancelled, for example, due to failure of a bailed suspect to attend, failure of the witness to attend or lack of suitable volunteers.21 Cancellations contribute to a further increase in delay before the witness can view a lineup. Since the introduction of video identification, the proportion of procedures cancelled has fallen to around five percent.22 Third, availability of a large database of video clips from which to select foils (approximately 12,000) makes lineups fairer to the suspect. Fourth, use of video is less threatening to victims, who no longer have to attend an identification suite where their attacker may be physically present. A further advantage is that a laptop can be taken to a witness who is unable to attend the police station. In a recent high-profile case, Abigail Witchalls, a victim of an attack who was left paralysed, was able to view a video lineup from her hospital bed, and a suspect was eliminated from the enquiry as a result.

Can psychological science improve the effectiveness of video identification?

An empirical investigation was recently conducted in our laboratory to investigate whether the effectiveness of the British video identification procedure could be enhanced by adopting: (a) a sequential double-blind procedure and (b) selecting foils that match the witness description of the culprit rather than foils who resemble the suspect. The impact of using moving rather than still video images was also investigated. Substantial laboratory experiments designed to simulate a forensically relevant situation as closely as possible are described.

Sequential double-blind presentation

Video identification naturally yields a sequential presentation. Research based on identification from photograph arrays suggests that sequential presentation can reduce mistaken identifications when the witness is required to make a decision after viewing each person as to whether he or she is the culprit. However, the current PACE code of practice does not allow any advantage of sequential presentation to be realised because it requires witnesses to view the entire lineup twice before making any decision.23 Thus, the question arises of whether video identification procedures could be improved by allowing the sequential double-blind instructions to be used.

We compared the outcomes of lineups when participant witnesses viewed a video lineup conducted under sequential double-blind instructions to the outcomes when following the procedures currently used by the police.24 Although the lineup administrator in police lineups is not blind to the identity of the suspect, for consistency all lineups in our experiments were conducted double-blind. All of the lineups were constructed under supervision of the police using the VIPER national database of foils. A video clip of the actors who played the role of a thief in our experiment were recorded at VIPER-equipped police stations. A video clip of the actors who played the role of a thief in our experiment were recorded at VIPER-equipped police stations. Approximately 200 students were recruited in small groups to take part in a study on mood and health. During the procedure the witnesses viewed an unexpected staged theft of a laptop. They gave a written description of
Figure 1: The outcome of culprit present video lineups run under the existing ‘view the lineup twice’ instruction and the sequential method. All lineup were administered double-blind. Statistically there were significantly more correct identifications of a guilty suspect made under the view twice instructions ($p<.05$).

Under sequential double-blind instructions there were significantly fewer correct identifications from culprit present video lineups compared to the existing procedure (36 percent vs. 65 percent of witnesses, see Figure 1). There were also fewer mistaken identifications of foils from culprit absent lineups (23 percent to 10 percent), but the latter effect was not statistically significant (Figure 2). The sequential instructions appear to reduce the rate of choosing, and therefore suppress correct identifications as well as incorrect identifications.

Sequential double-blind viewing instructions are believed to reduce the number of mistaken identifications by making it difficult for witnesses to make a relative judgement. In our experiment we asked the witnesses whether they had compared the faces of lineup members with each other or whether they had considered each person one at a time. 93 percent of witnesses who viewed a culprit present lineup answered ‘one at a time’, regardless of the viewing instructions they had been given. The naturally sequential presentation of a video lineup may make relative judgements very difficult even under the existing procedure. When these data for the culprit absent lineups are considered the proportion of witnesses answering ‘one at a time’ dropped to 80 percent under both lineup instructions. Thus, the presence of the culprit influenced the strategy witnesses used but the sequential double-blind instructions did not.

(Valentine, continued on page 26)
Recent research clearly shows that there is a reduction in the number of correct identifications of offenders under sequential double-blind instructions. In 2001, a combined analysis of 23 studies reported this effect. The Illinois Pilot Program, an evaluation of the sequential double-blind produce in real cases conducted by the Chicago Police, found the same effect. We have also found a reduction in correct identifications in a laboratory study under realistic conditions using video lineups constructed from the police national database of foils under police supervision. Although sequential double-blind presentation may provide some modest protection to innocent suspects, it did not show a reliable effect in our laboratory.

Moving images compared with stills
As part of the same experiment we have also investigated whether the moving images used in video identification contribute to its success compared to single full-face images, as frequently used in American photograph lineups. Intuition suggests that witnesses may be more likely to be able to identify a culprit from a moving video sequence that allows the face to be seen from a variety of angles. However, results from the live staged-incident experiment using video lineups showed that this was not the case. The rate of correct identification from culprit present video lineups was the same for 15-second moving video clips and for static full-face images presented on a monitor for 15 seconds (Figure 3). When the culprit was not in the lineup, there were significantly fewer mistaken identifications of foils from moving clips than from still images (Figure 4). Thus the use of moving video clips improves the fairness of lineups without affecting the sensitivity of the procedure. The same trend was found in a subsequent experiment, but the difference in mistaken identifications from culprit absent lineups between moving and still images was not statistically significant. When data was combined from an experimental condition which was common to both experiments, based on the existing identification procedure (i.e. viewing all lineup members twice), the advantage for moving images in culprit absent lineups was still significant. In conclusion, use of moving images may offer some protection to innocent suspects, but the size of any effect is small.

Research comparing selection of foils by culprit description and by suspect resemblance
The aim of a further experiment was to investigate whether video identifications could be made more reliable by using a culprit-description strategy, rather than a suspect-resemblance strategy to select the foils.
with higher rates of filler choices. This pattern remains constant across those lineups known to contain only a single suspect.

We recorded a total of 257 high confidence identifications (including all suspect structures and relationship categories). Of these 186 were suspect choices, 64 were no choices, and seven were filler choices. Thus, the rate of known errors for high confidence choices was 7/257 or 2.7 percent. When looking solely at those single-suspect lineups containing identifications of strangers, it is of some interest to note that there was only one known error (filler choice) out of a total of 65 high confidence identifications regardless of lineup type and procedure. Out of 81 high confidence stranger identifications (for both single and multiple-suspect lineups), there was also only one filler choice.

If witnesses were influenced by investigators in the simultaneous and not the sequential/blind procedure, then they should be more confident in their choices, on average, in the simultaneous than the sequential procedure. When we examined the percentage of highly confident witnesses (“That’s him. I’m certain.”; “100 percent sure.”; “100 percent absolutely positive.”; “I’m positive that’s the one that shot me.”; “Yep, that’s him. I’m sure, 200 percent”) for each lineup procedure, 78.3 percent and 77.1 percent were highly confident in their choices and 7.8 percent and 9.8 percent expressed low confidence (“I think that’s him, but I can’t be positive.”; “He looks like the guy, but I’m not positive.”; “#1 could have been the passenger.”; “Only 45 percent sure.”) in the simultaneous and sequential lineup procedure, respectively. Thus, contrary to the investigator influence explanation, witnesses were more likely to be confident in their choices under the simultaneous than sequential lineup procedure.

We can test the investigator influence explanation even more precisely by noting that if the administrator was leaking cues to pick the suspect (and not the fillers) during the simultaneous lineups, only those witnesses that picked the suspect would have the consensual validation of their choices. Those who picked the fillers would actually be disagreeing with the administrator’s influence attempt. This reasoning predicts that the witnesses viewing the simultaneous lineup (and who chose the suspect) should be more confident in those choices than witnesses who chose the suspect from a sequential lineup. In contrast, those who chose the fillers from a simultaneous lineup should be less confident than those who chose fillers from a sequential lineup. We analyzed the percent of witnesses who expressed high confidence for suspect choices and witnesses who expressed high confidence for filler choices. For simultaneous lineups, 69 out of 87 (or 79.3 percent of the) witnesses who chose the suspect did so with high confidence. For sequential lineups, 118 out of 140 (or 84.3 percent of the) witnesses who chose the suspect did so with high confidence. Thus, if anything, contrary to the investigator bias explanation, witnesses were more likely to be confident in their suspect choices in sequential/blind lineups than in simultaneous lineups.

When the filler choices were examined, 66.7 percent of the filler choices made to simultaneous lineups and 21.5 percent made to sequential lineups were done so with high confidence. While the Ns are small, the trend is nonetheless opposite to the investigator influence explanation for the results. Thus, those who chose a filler from a simultaneous lineup were more confident even though their choices should have disagreed with...

Table 3. Number and Percent of Suspect and Filler Choices as a Function of Racial Similarity of Witness/Victim and Culprit for All Lineups Containing Race Information

<table>
<thead>
<tr>
<th>Racial Similarity</th>
<th>Number of Choices</th>
<th>Percent of Choices</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Suspect</td>
<td>Filler</td>
</tr>
<tr>
<td>Other Race</td>
<td>126</td>
<td>13</td>
</tr>
<tr>
<td>Same Race</td>
<td>303</td>
<td>18</td>
</tr>
<tr>
<td>Total</td>
<td>429</td>
<td>31</td>
</tr>
</tbody>
</table>
the influence attempts of the administrator (assuming they existed).

iii) Investigator bias and cross-racial identifications

General consensus among experts is that an own-race bias exists such that witnesses are more accurate in their selections from lineups of their own race than in their selections from lineups containing individuals of a different race.\(^7\) Consistent with this idea, results from the Illinois Pilot Program\(^8\) (Table 3) indicate the probability that witness/victims chose a suspect increased when the suspect and witness were of the same race compared to when they were of different races. However, the filler choice rates were unaffected. Thus, when witness/victims attempted to identify suspects who were in a different racial group than their own, they were less likely to identify the suspect as the culprit and were no more likely to make a known error by identifying a filler. Another way to describe this result is that when confronted with an other-race lineup, suspects were less likely to choose someone as the perpetrator.

Examining the cross-racial data yields findings inconsistent with the notion of stronger investigator bias in simultaneous lineups. Investigator bias should be predictably stronger when memory for the culprit is weak. Research on the own-race bias would suggest that witnesses have weaker memory for culprits of another race than for culprits of their shared race. It would be hypothesized then, that effects of investigator bias would be more sizeable between simultaneous and sequential lineups given cross-racial identifications than would be seen in same-race simultaneous and sequential identifications. Suspect choice rates in same- and other-race simultaneous lineups were 188/267 (70.41 percent) and 74/140 (52.86 percent) respectively, while filler choice rates were 3/267 (1.12 percent) and 4/140 (2.86 percent). These rates can be compared to suspect choices in same- and other-race sequential lineups at 115/192 (59.9 percent) and 51/125 (40.8 percent) respectively, while filler choice rates were 15/192 (7.81 percent) and 9/125 (7.2 percent). Data is included from both single and multiple-suspect lineups.

As can be seen, the increase in suspect choice rates from sequential to simultaneous lineups was not larger when witnesses had presumed weaker memories for other-race suspects (40.8 percent \(v.\) 52.9 percent) than stronger memories for same-race suspects (59.9 percent \(v.\) 70.4 percent). In addition, the decrease in filler choices was not larger for other-race fillers (7.2 percent \(v.\) 2.9 percent) than same-race fillers (7.9 percent \(v.\) 1.1 percent). In short, these results are inconsistent with the investigator influence explanation.

Addressing critics of the Illinois Pilot Program

Some critics may argue that the results of the Illinois study are compromised by a confound between lineup procedure and blind testing. That is, the sequential lineups were conducted by investigators blind to the identity of the suspect in the lineup, whereas the investigators administering simultaneous lineups were not blind to the suspect’s identity. These conditions, however, were proposed as a package deal, meaning that the double-blind sequential lineup represents the policy as it would and should be instantiated over the traditional lineup. Thus, in evaluating the proposed policy change, we compared the components of the proposed change with the standing policy. Had the effects of lineup procedure and blind testing been evaluated independently, researchers would not be able to draw conclusions about the suggested policy change as a whole. The Illinois Pilot Program\(^9\) was not designed to test the varying options of simultaneous/blind, simultaneous/non-blind, sequential/blind and sequential/non-blind. These four procedures would need to be evaluated separately for the most comprehensive assessment. The primary conclusion researchers can make is that the sequential double-blind procedure, as tested in Illinois, is not superior to traditional simultaneous lineups.

Previous field studies that promote the success of sequential double-blind lineups, such as that in Hennepin County, MN\(^2\) have not included traditional simultaneous control groups to fully examine the proposed policy change. Regardless, results of the Hennepin County program closely mirror those of the Illinois Pilot Program\(^3\) with a 54 percent suspect identification rate and eight percent filler choice rate.
for all known single-suspect, double-blind sequential lineups. Had a control group been included, is it not possible that they would have also seen a higher rate of suspect identifications and lower rate of filler choices in traditional single-suspect, simultaneous lineups? The error on which researchers have clearly focused is that of reducing false identifications, while omitting the consequences of a witness failing to select the suspect.

With respect to any policy, the details must be clearly articulated. Simply indicating that a sequential lineup procedure is to be instituted is insufficient. There are several procedural variations of a sequential lineup, which if altered and implemented in different combinations, may dramatically affect any resulting eyewitness decision. One such variation is the inclusion of a stop rule, in which witnesses are told they will not be allowed to view the remaining alternatives in the lineup once they have made a positive identification. If no stop rule is included (following the sequential lineup procedure of the present study), an interesting psychological phenomenon arises. Suppose a witness positively identifies a foil in the lineup before the suspect is shown. What do witnesses do when they then see the suspect? Do they raise their decision criterion at that point? Do they dismiss the suspect’s photo because now that they believe they have completed the task of making a selection, are no longer paying careful attention to the lineup? In other words, we need to determine where witnesses set their decision criterion for making a selection from the lineup and whether this decision criterion is set differentially for simultaneous and sequential lineup procedures.

These issues beg the question regarding the decision strategy witnesses use in their selections. One key variable in this decision process is the witness’ strength of memory for the culprit. Also, a witness’ ability to discriminate between the actual culprit and innocent foils may depend on: how similar innocent foils appear to the culprit, how similar the culprit’s looks in the lineup are to his looks at the time of the crime, and how similar an innocent suspect’s looks in the lineup are to the culprit’s looks at the time of the crime. It is also important to know what witnesses use as their standard for determining this degree of match. When witnesses are presented with items in sequence, rather than all at once, we raise the distinct possibility that witnesses will use different decision criterion for different items as they progress through the sequence.

One of the key problems with the nature of a sequential protocol containing a stop rule is that it prevents the witness from being able to choose the best lineup member when there is more than one that is above the witness’s decision criterion for a good match. Thus, in low similarity lineups where the witness’ ability to discriminate (between the suspect and foils) is high, and the witness’ criterion for choosing is high, one would see a small difference in selection choices using different decision strategies. Conversely, in high similarity lineups, we might expect that one of the foils presented before the suspect might be a “good enough” match for the witness to pick him. However, were this foil and the suspect presented side by side, the witness might choose the suspect because the suspect is an even better match to the witness’ memory than is the foil. One consequence of this is that more foils will be chosen when the suspect is placed later in the lineup.

When witnesses attempted to identify suspects who were in a different racial group than their own, they were less likely to identify the suspects as the culprit and were no more likely to make a known error by identifying a filler. Another way to describe this result is that when confronted with an other-race lineup, suspects were less likely to choose someone as the perpetrator.

All foils that meet or exceed the witness’ criterion for making a positive identification will be chosen before the witness even gets to see the suspect.

From the Illinois Pilot Program data, we were able to examine differences in the rate of sequential lineup fillers selected before and after the suspect appeared in the lineup. Of the 21 sequential lineup foil choices (compared to a total of six simultaneous lineup foils chosen), 13 were selected by witnesses before the witness had the opportunity to view the suspect in the lineup, while eight were chosen after the suspect appeared.
was viewed. Eliminating the pre-suspect foil choices in sequential lineups, the difference in filler choice rates between the two lineup procedures is virtually eliminated. Interestingly, these results suggest that the higher rate of foil choices seen in sequential lineups may be driven by pre-suspect fillers that meet or exceed a witness’s decision criterion for making a positive identification.

Future Research

Evident from the current discussion, more laboratory and field research is needed to examine the efficacy of varying methods of conducting eyewitness identification lineups. Sequential double-blind lineups do not appear to yield fewer known errors than traditional simultaneous lineups. No single study can yield definitive results or subsequent recommendations that are to be widely applied.25 These findings should encourage more law enforcement agencies to conduct further research to scientifically investigate the costs and benefits that would be associated with instituting a given policy change. As seen from the different pattern of findings in Chicago and Evanston compared to those in Joliet, any policy changes must be evaluated against jurisdictional differences in the outcome and perhaps adjusted accordingly.

Furthermore, procedural variants of sequential lineups should be clearly defined and examined before any policy recommendations should be enacted, and the same holds for details of the double-blind procedure. In addition to the aforementioned option of including a stop rule, sequential lineups can differ in protocol based on several factors: what witnesses believe about the size of the lineup, what happens to lineup items after they are viewed, how many passes through the lineup a witness is given, and where in the lineup the suspect is positioned. The specifics of the sequential procedure may play an important role in eyewitness accuracy. For example, providing witnesses with information on the number of alternatives in the lineup could create pressure for the eyewitness to select someone before the end of the lineup is reached, or it may build expectations toward the end of the lineup that the culprit is not present. Conversely, if witnesses are not told how many faces are in the lineup, they may withhold making a selection, believing there are always more alternatives to be seen. If each alternative remains in view after it is presented, witnesses may utilize these as comparisons to the current item upon which they are deciding. Another possibility is to remove each item from view after a yes/no decision has been made, forcing an absolute decision for each item. Also, allowing multiple passes through a lineup may lead witnesses to withhold a selection until all lineup members have been viewed and compared to one another.

Yes, the specifics of a particular protocol can easily be defined as to which procedure will be utilized. However, research has not yet been conducted to determine how all procedural variations of sequential lineups interact to produce different results, nor has it been determined how foil choice rates and suspect (guilty suspect vs. innocent suspect) choice rates are affected by these combinations. Further research is needed regarding the double-blind procedure and its effects on accuracy as well. Options to be evaluated include witnesses making selections in private, in front of a blind administrator, or on a computer screen. The main argument for a double-blind procedure stems from the presumed possibility of investigators inadvertently sending signals to witnesses, thus influencing their selections.26 Perhaps, though, the mere presence of an investigator (blind or not) may differentially affect witness selections. For this reason, the recommended blinding procedure should be evaluated with witnesses making their final selections without any investigator contact. Alternatively, simply telling the witnesses that they should not assume that the investigator knows who the suspect is might be more than sufficient to produce whatever benefits researchers believe might be achieved with the more complex double-blind procedure.

When research is proposed to compare polices, all recommendations for change should include methods and measures that will allow one to monitor which procedure is more successful; before commencing research, the measures of success should be defined and agreed upon before any policy change is instituted. In the present domain, work that uses DNA, for example, to establish the accuracy of suspect choices would avoid criticisms about the use of filler choices.
enforcement purposes independent of reliability issues. Chicago, for example, has recently paid $15,000,000 to plaintiffs who proved that police “tipped” lineup witnesses. In all likelihood that award will generate a legion of similar suits—some justified, most unjustified, but all expensive to defend. “Double-blind” administration of lineups and arrays can allow the police effectively to defend themselves against such claims: a “double-blind” administrator can’t tip anyone. Double-blind techniques will also allow prosecutors to insulate the investigators from similar defense attacks in criminal trials.

There is a new zone of cooperation developing, where mutual education occurs between the street, the lab and the courtroom. This zone has been increasingly utilized in the aftermath of the DNA exoneration cases. Where practitioners have ventured into this zone, in Boston, New Jersey and Minnesota, for example, they have found the results to be rewarding—to mark a place in which to generate genuine improvements in practice. It presents novel challenges, but it also promises significant rewards. It was with those rewards (and those challenges) in mind that we joined with our colleagues to form the John Jay Center For Modern Forensic Practice, an effort to build a model of a neutral, academic venue where practitioners and scientists can meet to protect the innocent and catch the guilty—to get, in the words of Hennepin County District Attorney Amy Klobuchar, “Stronger cases and more justice.”

Still, there isn’t much point in involving scientists in these discussions unless we will allow them to “do science”—to approach the empirical questions we have identified with their proven, rigorous procedures. The Illinois Legislature posed an empirical question, and it ought have, but did not, receive an empirical answer derived by accepted scientific practice. We think that there are ways to apply the “objective scientific methodologies” that the Illinois Legislature wanted in the field. All of us have a lot to learn both from rigorous field studies and from the complementary future laboratory studies that must answer questions that the field studies are structurally unable to approach. As a next step along the path we have sketched a proposed protocol for future eyewitness field studies, which is reproduced below. Like everything else in this new zone of cooperation our suggestion is up for discussion.

We believe one particular aspect of the situation deserves priority treatment—the advent of efficient, reliable, convenient digital technology. Inevitably this technology because of its convenience and flexibility will soon begin to dominate eyewitness investigative procedures. That looming fact provides a spur to immediate further field studies, but it can also be a very significant aid to conducting those field studies. The capacity of new technologies to capture data and facilitate a simple, witness-driven identification encounter with a minimum of disruption in police operations can help to unravel many of the practical difficulties that led to the fatally confounded study recounted in the Mecklenburg Report. The capacity to simply standing on the far side of a laptop while the witness makes choices and records his or her level of confidence can allow the police to accomplish a great deal in the way of preserving both scientific rigor and investigative continuity.

One hundred years after Hugo Munsterberg first upbraided the legal system about its misuse of eyewitness memory, we still have a lot to learn from each other about the resolution of cases that turn on eyewitness memory. The learning should be around the conference table with everyone included, not in the courtroom, or through the press release. Our hope is that conference tables at academic institutions around the country may provide the neutral ground where all parties can engage the science to inform local practice. One of the things we’d bring to the conference table is the protocol that follows:

---

1 James Doyle was a panelist for New Policies, New Practices: Fresh Perspectives on Eyewitness Identification on April 21, 2006 at Loyola University Chicago School of Law. He and his co-authors are from the Center for Modern Forensic Practice, John Jay College of Criminal Justice at the City University of New York.


(D Doyle, continued on page 46)
But experiments showed that some burned things gained weight, like certain rusting metals and the theory died when Lavoisier identified oxygen as the key to combustion.19

Well into the development of modern physics there was theory of luminous aether, the medium by which light traveled through the universe. Great scientists believed in the aether until the beginning of the 20th century when the special theory of relativity eliminated the need for it. Experiments, like those of Michelson-Morley20 measuring the speed of light-had early demonstrated its short comings. There are still a few who believe in the aether.

In the 1960s the Soviets thought they had discovered polywater, a substance which had the consistency of syrup, did not freeze at 0 centigrade or boil at 100 centigrade. Now we know there is no polywater, the phenomenon was the product of creating polywater in contaminated glassware.21

It is crucial to understand that the process of scientific discovery begins, except for serendipity, with something unscientific-a hunch, an act of faith. A scientist believes or intuits that X is true and sets out to prove it. Because of this built in bias, science requires that experimental proofs be replicated by scientists not committed to the result. Even when a scientist has no vested interest in proving a particular proposition, the scientist always wants to design an experiment that proves or disproves something. Flawed or ineffective experiments are a waste of time and a scientist who does too many of them loses reputation and fund ing. At the very least, experimenters are emotionally committed to the validity of the experiment they design and to the results it produces. Again, we demand replication. The scientist's faith (often based on promising preliminary data) is often spectacularly disappointed. Consider hormone replacement therapy,22 the cancer drug Iressa,23 the pulmonary fibrosis treatment using Interferon gamma-lb24 and the use of St. John's Wort to treat depression.25

A simple rule is to treat all initial studies and experiments in any scientific field with caution, to require scrutiny of the experiment’s design and implementation and to require that its results be replicated. There is another, even more profound difficulty with tests of eyewitness identification procedure. Experiments in the physical sciences tend to yield invariable results like the exact speed of light. Davison-Germer proved the wave nature of the electron.26 Stern-Gerlach proved electron spin into two orientations.27 Shockley, Bardeen and Brattain detected the invariable characteristics of semiconductor surfaces.28 Experiments in human psychology are not so decisive. Even human physiology is not that clear. A pancreatic cancer diagnosis is widely, and correctly, perceived to be a death sentence, but 5-10% of patients survive it.29 Heroin and opium addiction destroys or damages its users but a few are not harmed.30 Returning to human psychology, I offer the account of the following two experiments in psychology.

In the 1970s, a psychologist-lawyer prepared eight

volunteers and himself to present themselves, unwashed, unshaved and disheveled at psychiatric emergency rooms and say, “I am hearing a voice. It is saying thud.” In all other respects they were to answer all questions truthfully except for their names and occupations. The mental institutions ran the gamut from well-funded to minimally funded. All who did this were admitted and kept anywhere from a week to a little over seven weeks. When David Rosehan published his results31 there was a storm of protest from psychiatrists. The initial admission was, perhaps, understandable because of the heard voice, the appearance of the person and they fact they picked a psychiatric emergency room to present themselves.32 It was the diagnoses made days or weeks later when the heard voices had disappeared and personal hygiene was restored and these sane people answered questions honestly that caused so much concern. The protests

(Zagel, continued on page 20)

SYMPOSIUM ISSUE

The best solution is to prevent [eyewitness] suggestion in the first place and to reduce the risk of error, even without suggestion-to do it right . . . before the lawyers get their hands on it. After a mistaken identification is made because of a suggestive procedure, it is very hard to correct.

(Zagel, continued from page 20)
SYMPOSIUM ISSUE

(Zagel, continued from page 32)

from the psychiatrists about the unfairness of the experiment led Rosenhan to proclaim that in certain coming months he would send in more pseudo-patients to one hospital. The idea was that, with fair warning, admitting psychiatrists would be able to better recognize the sane volunteers, at the very least, admitting psychiatrists would be able to better recognize the sane volunteers, at the very least, discounting the fact that the person self-presented at a mental institution. After the appointed period the hospital staff reported detecting 41 fake patients. Rosenhan then revealed he had sent none. This latter phase of his experiment may have been the most telling.

But the question is, what does Rosenhan’s data prove? It clearly does not prove that most psychiatric admissions were erroneous. In Rosenhan’s own account of his incarceration he encountered many deeply ill patients. It is also true that eventually, all of the institutions reached the correct result.

Rosenhan did not have to go to court to get one of his volunteers released from long-term commitment. It did prove that there was a significant risk of error but it provided no particular way to insure that the errors did not occur. Indeed, the only reason we know there were errors in nine particular cases is that, before the fact, Rosenhan, the psychologist was able to establish that his volunteers were clinically sane.

Elizabeth Loftus started with a belief that false memories can be implanted. In her experiment she used twenty-four persons, each of whom was given a packet of papers that, in one short paragraph each, described three true incidents that happened in the person’s childhood. The source of these stories was the family of the person being tested. To this a false paragraph was added (with the consent of the family). This false paragraph described the person being lost in the mall as a young child. After the packet was read by the subject, he or she was asked to write some more details of each incident or to simply say they didn’t remember the incident. Three-quarters of the recruits said, correctly, that they didn’t remember being lost is the mall. Twenty-five percent did remember and then added detail to the false lost in the mall story e.g., being scared, mother scolding, seeing toys, etc. 33

What Loftus showed was that some people, one-quarter of her subjects, could have a false memory implanted in their consciousness by a relative they trusted and that, once implanted, the memory could be auto-enhanced by their own minds. There are open questions as to whether implantation can be achieved by someone who is not trusted, but in criminal cases the allegation usually is that false memory was implanted by someone trusted, a sympathetic therapist or a nice-guy detective investigating a crime.

But the question here too is what does it prove? When you deal with a witness, how do you know whether the witness is one of the 25 percent or one of the 70 percent who resist implantation? So too with someone who confesses: how do you know whether the witness is one of those who can be convinced he or she committed a crime when they did not? There are clues, indicia to help one decide, but they are rarely conclusive. Even when expert testimony is offered, it is never decisive and, given our current state of knowledge, cannot be decisive. All that can be said is that this person or that is more likely to remember falsely than someone else. Even this is iffy because no has offered a properly tested diagnostic tool that can predict in advance precisely who will be implantable. Prediction (and some forms of what is called postdiction) is a very important element in the validation of scientific premises.

The same is true of testimony about whether an eyewitness is mistaken. In my own experience, some clearly mistaken identifications were made under perfect conditions and some utterly reliable ones were made under indefensible conditions.

And this, I think, is the point of the enterprise of pilot programs on identification procedures. After suggestion has worked on the minds of those who are susceptible, it is too late effectively to detect whether this has, in fact, occurred through cross-examination of the witness or by opinion evidence about the reliability of the witness. Some false memory and bad identification cannot be demonstrated to a court or jury. At best, you have the educated guess of an expert who does not, as yet, have enough science to say that a particular identification is unreliable with the requisite degree of scientific certainty. Indeed, the courts have not found support for the proposition that there is a generally accepted degree of scientific certainty against which to measure expert testimony. This is the reason that, where expert evidence of this sort is ad-

33 | Public Interest Law Reporter
mitted, it is remitted to the discretion of the trial judge, which is not the common standard for hearing the testimony about mental disease or defect, medical malpractice or any of the dozens of kinds of expert evidence permitted automatically in our trials.

The best solution is to prevent the suggestion in the first place and to reduce the risk of error, even without suggestion-to do it right, as I said, before the lawyers get their hands on it. After a mistaken identification is made because of a suggestive procedure, it is very hard to correct. So we must find the best way to achieve valid identification before it is made.

Where do we go from here?

So what did the pilot program study show? Mainly it showed that the particular sequential double-blind method of identification was not better (and possibly worse) than a simultaneous method of presentation. As a collateral matter, there were lower rates of false identifications than some predicted, nothing to confirm the presence of police suggestion deliberate or otherwise, or to support age and racial factors in identification. Identification of strangers was less reliable than non-strangers, an expected result, and the weapon focus theory was not supported.

Does this end the debate? No. One study, maybe even a dozen studies are not enough. There is more than one alternative to the ones tested here. It may well be that even this study replicated exactly might yield different results. Science moves in fits and starts. The scientists who predicted different results may, with dispassion, criticize the methods and designs of the study, which is what they ought to be doing.

What I urge here is dispassion, a lot less passion. It does society, whose stake in this is very large, no good if any expert in this field becomes another Nikolai Fedyakin, the discoverer of polywater who fought tooth and nail to discredit every experiment that trashed the existence of polywater. This just served to confirm the view that the West had of Soviet science from the days of Stalin and Lysenko. Moreover, I think, in the end, it will have to be the work of people new to the field who experiment and confirm whatever results we find. The current experts in the field may be viewed as having too great an interest in proving they were right. It is not an insult to Einstein to say that his theories were not fully accepted until someone other than he made the observations that validated his predictions. This may or may not be unfair, but it is the way of our world now.

I urge less passion for the lawyers as well. I have heard that prosecutors are happy with the results of this study and defenders are not. Why? Because they are in the adversarial mind set which is not the way to view these results and future results. What we ought to be after is accuracy of identification and what should please us is improvement in accuracy not an improvement in the conviction or acquittal rate.

I have not heard how the police perceive these results. I am sure there are some detectives who regard this as pointless academic meddling, but I believe that most will not, in the end, think so. It is useful to remember this about our current system: the role of the judge or jury is to decide the case on the basis of the evidence presented, the prosecutor who believes in good faith that he has good evidence is to pursue an accused and the defender is to make the best possible case for the accused. It is true too that once the police have concluded their investigation they are perceived, and sometimes rightly so, as adjuncts of the prosecution. But before the lawyers get their hands on the case, the police properly view their role as detecting what crime has been committed and who, in fact, committed it and who, in fact, did not. The investigative stage may be controlled by legal rules but it is not part of the adversary system and the allegiance of the police is to the truth. They have no mandated role to play for the prosecution or the defense. Now I do not dispute that pure adherence to appointed social roles is not invariable. Scientists take sides without adequate proof. Prosecutors will turn their backs on good evidence when they have a hunch that the defendant is innocent, and I have seen defense counsel, whom has covertly aided the prosecution by their tactical decisions because they were repelled by the client and his crime. There are police who will go too far in order to close a case. Of the unreliable people I have seen in my life, I have not found that police investigator/detective ranks have a higher proportion of them than does the bar or the cadre of expert witnesses. At least the investigator’s income is not influenced by success in litigation, and
the investigator’s duty is not to present the best case, it is to present the truth. We are all people and influenced by ambition and personal morals deemed more binding than society’s rules. But what struck me most when I went from being a prosecutor to being a police chief was the openness of most investigators about where the evidence, not their preconceptions, led them. They speak a different language. Prosecutors mostly talked over whether they had enough evidence to convict the accused and defenders engage in the same conversation from the opposite perspective. When the evidence was not strong, police tended to discuss whether this or that suspect or none of them was the right one for the crime. It is the reason why Jim Doyle discovered, and noted in True Witness, that it was the police who were most interested in the experts’ efforts to improve identification.

Most important here is not whether the pilot program showed that one method is no better than another. That it was done at all is the news of the day and if this method did not make a difference perhaps another will. If we keep our eye on a revived sense of the importance of truth, we will find better ways. You can laugh at the early days when some poor graduate student ran into a classroom carrying a fake gun and pretending to steal the professor’s wallet just so the professor could require all the undergraduates to report what they had seen and view a lineup.

But the efforts improve over time and some of it is due to the feedback that comes from seeing predictions fail and reconsidering the evidence for the predictions. And because we see there may be better ways to send more of the right people to prison and none of the wrong ones, the increased attention to truth in criminal cases will bring the support we need for this new attempt to climb a very old mountain.


The U.S. National Institute of Justice Technical Working Group for Eyewitness Evidence recommended use of a culprit description strategy.28 In contrast, the English code of practice requires a suspect resemblance strategy.29

When the police in England construct a video lineup they choose foils on the basis of their resemblance to the suspect. The problem with this strategy is that high similarity between the suspect and foils makes identification of an offender present in the lineup very difficult even for a witness who has a good memory of the offender. The logical extreme is a lineup of clones, which obviously would render the process ineffective.

The logic behind a culprit description strategy is that the witness may remember the description they gave to the police and may look for somebody who matches that description. If all foils have been selected to match that description, there will no bias against the suspect. It does not make a lineup unfair if the members differ in characteristics not mentioned in the description. In fact, this is a useful characteristic of a lineup. A witness seeing an offender on a lineup may recognize a feature they did not describe. However, this is no more likely to occur for an innocent suspect than for any of the foils.30

The effectiveness of a culprit description strategy and a suspect resemblance strategy was compared in an experiment using a staged live incident similar to that described above. All witnesses provided a written description of the offender immediately after an unexpected staged incident. The participants individually attended a video identification approximately a week after the incident. All video identifications were conducted in accordance with the current code of practice, except that the administrator was always blind to the identity of the suspect and did not know if the culprit was in the lineup. The lineups were conducted from the VIPER national database of foils. Half of the lineups were constructed using the suspect resemblance strategy required by the code of practice. The remaining lineups were constructed by selecting foils who matched the description given by the individual witness. Within that constraint, foils were selected who differed as much as possible, although foils did not dif-