

# Multiple Independent Identification Decisions: A Method of Calibrating Eyewitness Identifications

Sean Pryke, R. C. L. Lindsay, Jennifer E. Dysart, and Paul Dupuis  
Queen's University

Two experiments ( $N = 147$  and  $N = 90$ ) explored the use of multiple independent lineups to identify a target seen live. In Experiment 1, simultaneous face, body, and sequential voice lineups were used. In Experiment 2, sequential face, body, voice, and clothing lineups were used. Both studies demonstrated that multiple identifications (by the same witness) from independent lineups of different features are highly diagnostic of suspect guilt (G. L. Wells & R. C. L. Lindsay, 1980). The number of suspect and foil selections from multiple independent lineups provides a powerful method of calibrating the accuracy of eyewitness identification. Implications for use of current methods are discussed.

Many different systems of establishing human identity have been developed and used (Lindsay, 1999b). Some of these systems include forensic hair comparison (Bisbing, 1982), fingerprints (Durand, 1999), and most recently, DNA analysis (Connors, Lundregan, Miller, & McEwen, 1996). Although DNA and fingerprints are highly informative in many cases, other cases exist in which no physical samples are available for conducting these analyses. In these situations, a common method for establishing the identity of the criminal is eyewitness identification.

Eyewitness evidence, however, is not infallible and thus has been criticized since its widespread introduction to legal proceedings (e.g., Munsterberg, 1908; Whipple, 1909). Early research on eyewitness evidence focused on documenting eyewitness errors and was of relatively little use for the prevention of false identifications. Beginning in the late 1970s, psychologists began to focus on preventing eyewitness errors, an approach that specifically differentiated between estimator variables (outside the control of the justice system) and system variables (under the control of the justice system; Wells, 1978). System variables include the methods of obtaining identification evidence and provide an opportunity to improve the quality of identification evidence presented in court by developing superior procedures. Regardless of the identification method used, a central issue in the empirical investigations of identification methods has been the tendency of witnesses to use different judgment strategies (e.g., relative vs. absolute) when making an identification decision (Wells, 1984).

A relative judgment strategy is used when the witness selects the lineup member who, in comparison with the other lineup members,

is most similar to the person who committed the crime. An absolute judgment strategy is used when the witness directly compares each lineup member with his or her memory of the criminal. Research results indicate that relative judgments are inferior because they lead to inflated rates of false identification and do not increase the rate of correct identification dramatically (e.g., Lindsay & Wells, 1985; Wells, 1984). In other words, relative judgments lead to eyewitness identification decisions that are less reliable as an indication of guilt than are the identification decisions of witnesses who make their decisions on the basis of absolute judgments (Dunning & Stern, 1994; Lindsay & Bellinger, 1999; Lindsay & Wells, 1985). As a result, relative judgments reduce the probative value of identification evidence.

Probative value refers to the potential usefulness of any particular piece of evidence to a court of law for determining the issues before the court. In the case of eyewitness identification, probative value reflects the degree to which identification of a suspect is useful for discriminating whether the suspect is guilty or innocent. In experimental situations, researchers are able to estimate the probative value of an identification procedure by forming a ratio of correct to incorrect identification decisions for each procedure, known as a *diagnosticity ratio* (*DR*; Wells & Lindsay, 1980). Wells and Lindsay (1980) argued that when the diagnosticity ratio equals 1, the procedure does not provide any evidence about whether the accused is guilty or innocent. In other words, because the identification of a suspect is equally likely for either a guilty or innocent suspect, the identification procedure has no probative value.

In an attempt to reduce reliance on relative judgments and increase probative value, Lindsay and Wells (1985) developed a sequential method of lineup presentation. The sequential lineup involves more than just showing photographs to a witness one at a time. Witnesses are required to make a decision about the guilt or innocence of each person before the next photo is shown. Witnesses also are not aware of how many photos will be shown, and the procedure should be administered by a person who is unaware of which lineup member is the suspect (blind testing). A recent meta-analysis concluded that sequential lineups consistently produce lower false identification rates and higher diagnosticity ratios

---

Sean Pryke, R. C. L. Lindsay, Jennifer E. Dysart, and Paul Dupuis, Department of Psychology, Queen's University, Kingston, Ontario, Canada.

Jennifer E. Dysart is now at the Department of Psychology, Southern Connecticut State University, New Haven.

This research was supported by grants to R. C. L. Lindsay and Jennifer E. Dysart from the Social Sciences and Humanities Research Council of Canada.

Correspondence concerning this article should be addressed to R. C. L. Lindsay, Department of Psychology, Queen's University, Kingston, Ontario K7L 3N6, Canada. E-mail: lindsayr@psyc.queens.ca

when compared with simultaneous lineups (Stebly, Dysart, Fulero, & Lindsay, 2001).

Although sequential lineups have been shown to be an improvement over simultaneous lineups in some situations, there are limitations. For example, sequential lineups have not been successful at reducing false positive choices with child witnesses (Lindsay, Pozzulo, Craig, Lee, & Corber, 1997) in cases in which the appearance of the suspect has been changed (Pozzulo & Lindsay, 1995) or in cross-race situations (Lindsay & Smith, 2001). Furthermore, and most critical to the present article, sequential lineups do not *eliminate* eyewitness error with same-race adult witnesses who view suspects that have not changed their appearance. In fact, the average false identification rate with sequential lineups is 9% and the correct identification rate is approximately 35% (Stebly et al., 2001). Therefore, approximately one in five suspect identifications from sequential lineups may be wrong.<sup>1</sup> As a result, no existing eyewitness identification procedure can relieve the courts of the burden of deciding after the fact (or *postdicting*) which eyewitness identifications are accurate versus inaccurate.

Various factors have been used in attempts to postdict identification accuracy but with limited success. These factors include witness confidence (Bradfield, Wells, & Olson, 2002; Leippe, 1980; Luus & Wells, 1994; Wells & Bradfield, 1998; Wells, Ferguson, & Lindsay, 1981; Wells, Lindsay, & Ferguson, 1979), response latency (Smith, Lindsay, & Pryke, 2000; Sporer, 1993), and indication of type of decision made (i.e., relative vs. absolute; Dunning & Stern, 1994; Lindsay & Bellinger, 1999). Smith et al. (2000) found that by relying on multiple postdicting variables in combination, it was possible to correctly classify 76.3% of the witnesses who chose a suspect (i.e., choosers) from simultaneous lineups as accurate or inaccurate. However, almost one quarter of the cases using these variables would still produce mistaken decisions. Furthermore, the variables that provide some limited ability to postdict accuracy following identification from simultaneous lineups do not postdict accuracy of identification from sequential lineups (Smith et al., 2000).

The courts require methods of determining which particular identification decisions are correct versus incorrect. In the absence of a reliable method of discriminating correct from false identifications, the courts will convict many innocent suspects. A great advance in lineup procedures would be made if the probable accuracy of the identification decisions of individual witnesses could be estimated with greater precision than permitted by current procedures. With this goal in mind, Lindsay, Wallbridge, and Drennan (1987) attempted to produce stronger identification evidence by having witnesses identify the clothing worn by the target as well as, and independently from, the identification of the person. Their data indicated that one additional independent identification increased the diagnosticity of the identification evidence from 1.91 with one piece of evidence (choice from a lineup) to 13.60 with two (identification of the person and his clothing from independent lineups). Alternatively, 65.6% of witnesses who identified only the suspect were correct, whereas 93.2% of witnesses who identified both the suspect and clothing were correct. These data suggest that any individual feature that can yield an independent identification should improve diagnosticity using multiple independent lineup identifications.

There was, however, a serious limitation to Lindsay et al.'s (1987) study. Lindsay et al. (1987) used clothing, an easily change-

able cue, for the additional lineup. In real cases, it is not always possible or reasonable to conduct a clothing lineup (e.g., if similar clothing is not found with the suspect at the time of arrest, there may be no clothing to identify). In addition, the clothing used in the study was highly distinctive. If the clothing described by the witness is too common (e.g., jeans and a white T-shirt), then a clothing lineup may be less diagnostic because identifying such clothing may be very difficult. It would therefore seem prudent to use features that are part of the person. For example, independent lineups constructed for a person's voice and body would not raise the concerns related to clothing.

The Lindsay et al. (1987) study also was limited in that it only used one additional lineup, and the rate of multiple identifications was relatively low (29%). That is, less than half of the participants who selected the actual criminal also selected his clothing. If voice and body cues also are used, it is unclear from the literature how often these cues will be successfully selected from a lineup. It is equally unclear as to whether these specific types of lineups will combine in a way that produces a practically useful number of multiple identifications. This leaves a number of questions unanswered. Will the addition of further lineups increase the number of witnesses who can successfully pick out the target from more than one lineup? Will the addition of further lineups substantially increase the number of witnesses who make multiple false identifications of the innocent suspect? That is, will a significant number of witnesses pick the innocent suspect out of more than one lineup as the number of lineups increases?

Previous studies have attempted to improve identification accuracy by providing multiple cues in a single live or videotaped lineup (e.g., Cutler & Fisher, 1990; Cutler & Penrod, 1988; Egan, Pitner, & Goldstein, 1977; Melara, Dewitt-Rickards, & O'Brien, 1989). Despite occasional signs of success, Cutler, Berman, Penrod, and Fisher (1994) found that when results were averaged across experiments, additional cues presented simultaneously in lineups produced a trivial effect on identification accuracy. One potential reason for the lack of improvement in false identification rates over various presentation methods may be that one cue (the face) overwhelmed the other features. In other words, witnesses may select lineup members based on the face and then convince themselves that inconsistencies of body or voice are not sufficient to alter their identification decision.

---

<sup>1</sup> Diagnosticity ratios are not commonly used in the field, and their interpretation is not obvious to many people. An alternative way of thinking about the results is as predictions of the proportion of witnesses that would be accurate if a procedure or decision criterion (such as multiple identifications) was applied to a large number of cases. To estimate such a value, it is necessary to know the "prior probability" of guilt, that is, the proportion of lineups that would be target-present. The base rate of target-present lineups in the real world has not been adequately assessed to date but G. L. Wells (personal communication, May 13, 1999) has been working on this issue. A prior probability of guilt near .5 is not unreasonable in real world cases. Estimated percentages of percentage guilty are provided based on an assumption of target-present and target-absent lineups being equally likely. Thus, the percentage guilty is determined by the ratio of correct identifications to all identifications. However, higher diagnosticity ratios lead to superior decisions in all situations other than prior probabilities of 0 or 1 (Wells & Lindsay, 1980; Wells & Turtle, 1986).

However, if the features were presented independently, then no feature could overwhelm another. Instead of merely giving a witness an opportunity to identify a person, witnesses in the current studies were given multiple opportunities to identify various parts of the person independently, including the face, the body (shoulders down), and the voice. The basic premise being tested is that multiple independent identifications of various features of the same person increase the likelihood of that person being the one who was previously encountered by the witness, thus increasing the probative value of the identification evidence. A multiple independent identification lineup is intended to “calibrate” the likely accuracy of eyewitnesses based solely on the results of their identification decisions. It was hypothesized that combining independent identifications of features, such as face, body, and voice, should produce evidence that can assist in estimating the probability that a suspect is guilty. The more often someone is selected from independent lineups, the more likely it is that the person is the target.

### Voice Recognition

A body of voice recognition literature has existed for some time (e.g., Bull 1978, 1981; MacLeod, Frowley, & Shepherd, 1994; McGehee, 1937; Pollack, Pickett, & Sumbly, 1954). Researchers have varied the length of speech samples (Bull & Clifford, 1984; Yarmey & Matthys, 1992), the number of foils in a test array (Bull & Clifford, 1984), the effects of disguising the voice (Reich, Moll, & Curtis, 1976; Orchard & Yarmey, 1995; Saslove & Yarmey, 1980; Yarmey, 1991), and delay between exposure and identification (Bull & Clifford, 1984; McGehee, 1937, 1944; Saslove & Yarmey, 1980). On the basis of these findings, relatively brief, undisguised voice samples and short delays were used in the studies reported to provide a reasonable expectation of accuracy of voice identification.

### Body Identification Literature

It is not possible to determine a priori if stationary bodies can be recognized by witnesses because there are no published studies examining this issue. Although studies investigating the ability of witnesses to identify body movement seem promising, lineups including motion would have to be conducted live or videotaped. The use of still photos is much less expensive and time consuming than these methods and thus is more likely to be adopted by police departments.

### Selection of Foils

Although it appears reasonable to use body, voice, and face identifications as independent pieces of evidence (creating different lineups for each feature), the issue remains whether all of the lineups need to be constructed by matching the foils to the suspect. In the current studies, all lineups were constructed based on the dual principles of match-to-description (Wells, Rydell, & Seelau, 1993) and default values (Lindsay, Martin, & Webber, 1994). According to the first principle, all lineup members are selected to match all aspects of the description of the culprit provided by witnesses. The second principle requires that any foil deemed so unusual that they would likely have been described differently by witnesses be replaced. Pilot work for the current studies, however, revealed that the targets’ voices were never described, and their

bodies were described only in very general terms (e.g., average build). As a result, it was unnecessary to separately select foils for each feature.

### Innocent Suspect

One complication in analyzing eyewitness identification data is the designation of an innocent suspect. In any target-absent lineup, variance in the rate of choice of the lineup members is common. Selecting a rarely versus frequently chosen person as the innocent suspect can dramatically influence the interpretation of the results. If the purpose of the research is to establish estimates of real-world error rates, this issue can be critical. However, when the purpose is to compare the value of different identification procedures, this is not necessarily a problem. Identification procedures will not matter a great deal if the innocent suspect, by chance, is less similar to the true criminal than other lineup members. On one hand, this will occur in real-world lineups with some frequency. On the other hand, because suspects are often arrested based on their fit to a witness’ description, suspects (even innocent suspects) will resemble the criminal more than other lineup members at least some of the time. Identification procedures that can protect such innocent suspects without sacrificing correct identifications are highly desirable. Because that is the aim of the present research, the data are analysed using a “worst case scenario” approach, in which the innocent suspect was designated as the person who was most often identified from the target-absent face lineup. In some contexts (e.g., combined body and voice identification), the data also are analyzed to reflect the worst possible result excluding face choices. To provide for such an analysis, it was necessary to use the faces, bodies, and voices of the same people in all three absent lineups to permit multiple identification rates of each member of the absent lineups. The lack of detailed description of voice and body made this relatively easy. In real police work, it would not be necessary to use the faces, voices, and bodies of the same people as foils in the three lineups.

### Experiment 1

Experiment 1 explored a method of using voice and body information about a criminal that is presumably stored in a witness’ memory of the event. This was done by constructing three separate lineups, one each for the face, the voice, and the body. Each lineup was presented to the witnesses for the purpose of identifying the target. The face lineup was always presented first, based on the reasoning that further identification evidence should only be collected if the suspect is selected from the face lineup. The emphasis on face recognition in the present study is consistent both with psychological literature and legal practice in which police and the courts may not be satisfied with the identification of other features if a witness was unable to identify the face. However, if multiple selections excluding the face are diagnostic of guilt, there is no reason to dismiss such data. Despite this possibility, face identification will likely remain an important factor, and the face lineup was presented first to avoid any possible interference of attempting voice or body identification prior to the face lineup. Another advantage of presenting the face lineup first is that the results provide an estimate of the identification rate that would have been obtained from standard identification procedures

(simultaneous lineup in Experiment 1 and sequential lineup in Experiment 2). No mention of the additional lineups was made to the participants until after their responses to the face lineup.

### Method

#### Participants

Participants ( $N = 147$ ) were drawn from a 2nd- and a 3rd-year undergraduate psychology course. Students volunteered to participate in a 15-min session following class.

#### Materials

The materials consisted of photos of faces (neck up) and bodies (shoulders down), and 60-s samples of the undisguised voice of the target and foils. Note that no clothing cues were present in the face photographs, and all clothing in the body photographs was different than that worn by the target during the initial exposure. The photos and voices were of students between the ages of 19 and 25. The lineups consisted of six photographs that were presented simultaneously or six voices presented sequentially. During the identification phase of the study, the photo-arrays were projected on a large screen (approximately  $4 \times 6$  m) and were clearly visible from the back of the classroom. The voices were played sequentially from a computer program and were clearly heard in the classroom, which was equipped with multiple, high quality speakers.

*Recruitment of lineup members.* Potential lineup members were recruited on campus based on their physical appearance. Each potential member came to the laboratory where face and body photos were taken and digital voice recordings made. The six foils used were selected on the basis of fit to the general description of the target, but none were selected on the basis of extreme similarity to him (the target). Little effort was made to select on the basis of similarity to features below the shoulders (all were of average build), and no effort was made to select on the basis of voice similarity. There were two reasons for this approach. First, it is unlikely that police will be willing to expend much effort finding foils that match the suspect on the basis of more than general similarity. Second, prior experience with eyewitness studies suggests that witnesses rarely provide detailed information about the culprit's voice or body, unless it is unusual in some way, which was not the case for the target used in this study. As a result, the body and voice lineups conformed to the standard that all lineup members should fit the witness' description of the criminal (Technical Working Group for Eyewitness Evidence, 1999).

*Independence.* If witnesses were able to identify the target's voice or body from seeing a picture of the target's face, then the independence of a multiple identification of the face and either the voice or body would be compromised. To ensure that the assumption of independence was reasonable, we conducted a pilot study. University students ( $N = 12$ ) were brought into the laboratory individually and asked to match the voice, body, and face of the lineup members. Participants were informed (both verbally and with written instructions) that lineup order was varied from lineup to lineup such that face number one did not necessarily correspond to voice or body number one and so on. This was done to preserve independence both in fact and in the perception of participants. Overall, the proportion of correct matches between features was slightly lower than chance expectation. Note that failure to achieve independence would only work against our hypothesis because it would permit easy, multiple identifications of both guilty and innocent people.

#### Procedure

The target was introduced to students by the course instructor at the end of class. He was described as a student seeking volunteers to participate in an experiment. The target spoke for approximately 1 min. Following this,

the target left the room and the experimenter entered. Students were informed that the study pertained to their memory of the target. Those wishing to leave were excused (almost all stayed). Identification forms were distributed to all remaining students. One class responded to lineups including the target (target-present) and the other class responded to lineups with a replacement for the target (target-absent). Participants first responded to a simultaneous lineup of faces. Next, the instructions indicated that participants would be asked to attempt to identify the voice of the target by listening to recordings of men's voices one at a time. Each recording was approximately 1 min in duration and could be replayed if necessary (it was not). Participants were asked to indicate whether each voice was of the target and were warned that once they had indicated a decision, they could not listen to the voice again.

Finally, participants were shown a simultaneous lineup of bodies from the shoulder down. For both target-present and target-absent conditions, the faces, voices, and bodies were of the same six individuals. Thus, it was possible to make multiple identifications of any of the lineup members.

The identification forms provided instructions and places to respond for each of the three lineup tasks. Participants were informed that, as in a real police procedure, the person they were trying to identify may or may not be present in the lineups. When the voice and body lineups were presented, participants were reminded that the target may not be present. They were also told that he would either be in all or none of the lineups they saw. For the face and body identifications, participants circled a number from 1 to 6 arranged in two rows of three to correspond to the array of photos projected at the front of the class, or they circled "not present" to indicate that the target was not in the lineup (standard simultaneous lineup form). The voice identification was recorded as a yes or no decision following each voice until all six voices had been played.

### Results and Discussion

Before addressing the data for multiple identifications, the results for each of the three lineups are presented independently. Correct facial identifications of the target were obtained from 73.5% of participants in the present condition, whereas the most selected innocent face was chosen by 20.2% of participants in the absent condition ( $DR = 3.63$ , 78% guilty). For the body lineup, 52.9% of participants correctly identified the target and 81.0% falsely identified the most selected innocent person ( $DR = 0.65$ , 39% guilty). Comparable percentages exist for the voice lineup, with 39.7% of participants correctly identifying the target and 35.4% falsely identifying the most selected innocent person ( $DR = 1.12$ , 53% guilty). However, because it was presented sequentially, it was possible to make multiple choices from the voice lineup. One approach to analyzing the data is to consider multiple choices as impeaching the witness' ability to make a reasonable identification decision either for this feature (suggesting pairwise deletion by treating multiple voice choices as no voice choice) or for any identification (suggesting listwise deletion by dropping all data for the witness). Following the pairwise deletion strategy, the voice lineup produces a somewhat higher diagnosticity ( $0.34/0.27 = 1.46$ , 59% guilty). Thus, the face lineup was reasonably diagnostic, whereas performance on the voice lineup was poor, particularly if the witnesses making multiple selections were included. The innocent suspect was more likely to be selected from the absent, body lineup than the target was from the present body lineup, thus producing a diagnosticity ratio of less than 1. If these lineups were used by themselves, there would be a very large ratio of false to accurate identifications for the most frequently selected voice or body compared with the proportions for the most frequently selected face.

At first glance, the low diagnosticity ratios of the voice and body lineups may appear to threaten the value of using them as independent sources of evidence. We could consider ourselves unlucky that the person from the absent lineup with the most selected face also had the most selected voice. In fact, we argue that this misfortune illustrates just how powerful multiple independent identifications actually are as a source of evidence.

There are many ways that the data can be described and discussed. One way is to describe and interpret the diagnosticity ratios of the various combinations of selections (see Table 1). The second column of Table 1 shows the diagnosticity ratio for the combination indicated for the row (e.g., face only, face and voice, etc.) and is calculated by using as the innocent suspect the person whose face was most frequently selected from the target-absent lineup. The third column estimates the percentage of suspects that would be guilty given this evidence and equal likelihood of an innocent versus guilty suspect in the lineup. The fourth column, labelled *Worst case*, provides the least positive diagnosticity ratio regardless of which lineup member is designated the innocent suspect. Thus, for any situation in which the outcome would be worse if the suspect was not the person whose face had been selected most often, this column will show a lower ratio than the first column; otherwise, the two values will be the same. The final two columns provide the diagnosticity ratios using as the innocent suspect the person whose face was most frequently selected from the target-absent lineup but with pairwise deletion of multiple voice choices (multiple voice choices are treated as no voice choices but other selections for the witness are counted) or listwise deletion of multiple voice choices (all data for the witness is dropped if a multiple voice selection was made).

Contrary to expectation, failure to make any selections was not diagnostic of innocence (occurring at a slightly higher rate in the present than the absent condition,  $DR = 0.98$ ). However, listwise deletion of multiple voice choices resulted in a very high diagnosticity ratio for making no choices (12.55, 93% innocent). This pattern reflects a high rate of multiple false voice choices in the absent lineup among witnesses who otherwise (face and body) made no choices.

Identification of the face alone produced a diagnosticity ratio of 1.37 (58% guilty), which is substantially lower than the 3.63 (78%) value obtained from examining the data for facial recognition,

ignoring other decisions. This value was not dramatically altered by deleting multiple voice choices. Many more witnesses who made correct than false facial identifications also made additional correct choices. Indeed, less than 20% of witnesses in the target-present condition identified only the face of the target, and less than 6% made no selections.

The diagnosticity of body only choices was 4.68 (82% guilty) with the designated suspect but was 0.24 (19%) in the worst case scenario. This pattern illustrates the primary strength of multiple independent identification lineups. One body in the absent lineup drew more choices than the target's body in the present lineup. However, the innocent suspect is at risk only if he has the misfortune to resemble the criminal in two or more independent features. Even then, the same witnesses that see a facial resemblance would have to see the body resemblance. The voice lineup produced this unfortunate coincidence for our designated innocent suspect; thus, the diagnosticity of the voice selection alone was 0.20 (17% guilty), a very dangerous level. As stated earlier, we recommend that face selection be a criterion for considering the information value of other identification decisions. The data just described support this view. Using either body or voice data alone, there is a one in six chance that the innocent suspect will be the most selected lineup member and be as likely or more likely to be identified than the target would have been if present. Eliminating multiple voice choices does not eliminate this problem.

Examining the combined choices presents a much more positive picture. The diagnosticity of combined choices of face and body is undefined because this pattern never occurred for the designated innocent suspect, despite occurring for 23.5% of witnesses in the target-present condition (thus 100% guilty). Even the worst case scenario produces a diagnosticity ratio of 6.35 (86% guilty) for this combination, which is substantially higher than the value using facial lineups only. Face and voice selections combined to produce a diagnosticity ratio of 3.52 (78% guilty). This value is comparable with facial identification alone but increases to over 5 when multiple voice choices are deleted. Body and voice were never both identified for the designated innocent suspect (again, 100% guilty), but the worst case scenario again produces a dangerously low value (0.17, 15%), suggesting that face selection is crucial. Finally, the diagnosticity ratio for all three selections in combina-

Table 1  
*Diagnosticity Ratios (DRs) of Lineup Selections in Combinations for Designated Suspect, Worst Case Scenario, and Designated Suspect With Pairwise and Listwise Deletion of Multiple Voice Selections (Experiment 1)*

Suspect choice	Estimated				
	DR	% guilty	Worst case	Pairwise	Listwise
Not selected	0.98	51	0.98	0.98	12.55
Face only	1.37	58	1.37	1.34	1.51
Body only	4.68	82	0.24	4.68	3.70
Voice only	0.20	17	0.20	0.24	0.22
Face and body	—	100	6.35	19.58	—
Face and voice	3.52	78	3.52	5.88	5.27
Body and voice	—	100	0.17	—	—
Face, body, and voice	11.00	92	11.00	—	—

*Note.* Dashes indicate undefined values because the frequency in the absent condition only was zero.

tion was 11.00 (92% guilty), including witnesses making multiple voice choices and undefined without them (100% guilty).

One shortcoming of the analysis thus far is the absence of inferential statistics. Perhaps these results look good but are not significantly different from chance expectations. There are no direct tests to compare diagnosticity ratios with their chance level. A slightly different presentation will permit another useful view of the results and clear indications that the pattern obtained is statistically significant. Effect size is indicated as meta-analytic  $r$  (Rosenthal, 1991). In Table 1, each combination of identification results is treated as if it was uniquely of interest. Our hypothesis was that multiple suspect identifications in which the face was one of the identified features would be highly diagnostic of guilt. Table 2 presents the proportion of witnesses who did not identify the suspect's face, identified just the suspect's face, identified the suspect's face and one additional feature (body or voice), or identified all three features of the suspect.

Nonidentification of the suspect is diagnostic of innocence ( $0.80/0.26 = 3.01$ , 75% innocent). Identification of only the face is somewhat diagnostic of guilt ( $0.19/0.14 = 1.37$ , 58% guilty). However, identification of two suspect features is highly diagnostic of guilt ( $0.41/0.05 = 8.08$ , 89% guilty), and identification of all three features is more diagnostic still ( $0.13/0.01 = 10.15$ , 91% guilty). The pattern of frequencies generating these diagnosticity ratios differs from chance expectation,  $\chi^2(3, N = 147) = 49.02$ ,  $p < .001$ ,  $r = .58$ . Comparing the frequencies just for face, face plus one, and face plus two suspect feature selections across the target-present and target-absent lineups indicates that the results deviate from chance expectation, even excluding nonchoosers,  $\chi^2(2, N = 66) = 9.60$ ,  $p < .01$ ,  $r = .38$ .

Another factor that can be considered is the diagnostic value of foil choices. Foil choices are known errors and indicate that the suspect may be innocent (Wells, 1984; Wells & Lindsay, 1980). Evidence supporting this position exists in the literature (Wells & Olson, 2002). If multiple as compared with single suspect choices from independent lineups are more diagnostic of guilt, then multiple as compared with single foil choices may be more diagnostic of innocence, that is, the more foil choices the witness makes the lower the probability may be that the suspect is guilty. Foil choices were more common in the target-absent ( $M = 1.71$ ) than in the target-present ( $M = 0.81$ ) condition,  $F(1, 145) = 42.58$ ,  $p < .001$ ,  $r = .48$ . Participants who made one or more choices of someone other than the suspect, were significantly more likely to be viewing

the target-absent than the target-present lineup,  $\chi^2(1, N = 147) = 57.20$ ,  $p < .001$ ,  $r = .63$ .

Using a single lineup, it is rare to obtain both a suspect and a foil selection from the same eyewitness. However, using multiple independent lineups, a new measure can be created by subtracting the number of foil choices from the number of suspect choices. This produces a variable (in the current results) with a range from  $-4$  to  $3$ . The diagnosticity ratios (percentage guilty) for this variable were 0.00 (0%), 0.17 (14%), 0.09 (8%), and 0.47 (32%) for negative net choices (four, three, two, and one more foil than suspect choices, respectively). When an equal number of suspect and foil choices were made, the diagnosticity ratio was 0.82 (45%). When suspect choices exceeded foil choices, the decisions were diagnostic of guilt: 5.36 (84%), 13.54 (93%), and undefined (100%) for one, two, and three more suspect than foil choices, respectively. Again, the pattern of frequencies on which these ratios are based is unlikely to be a chance result,  $\chi^2(4, N = 147) = 61.56$ ,  $p < .001$ ,  $r = .65$ . (There are four degrees of freedom because it was necessary to collapse over the combinations of four, three, and two more foil than suspect choices to avoid expected values less than 5 in the target-present condition.)

Finally, the data were further examined to determine the percentage of witnesses who made a multiple suspect selection with and without making foil choices. In the target-present condition, .25 of the witnesses selected the target from the face lineup and at least one other lineup while also making a foil selection. In the target-absent condition, .06 of the witnesses picked the innocent suspect out of a face lineup and at least one other lineup while also making a foil selection. This led to a diagnosticity ratio of 3.95 (80% guilty). In the target-present condition, .29 of the witnesses selected the target from the face lineup and at least one other lineup while not making any foil selections. However, in the target-absent condition, only .01 of the witnesses selected the innocent suspect from the face lineup and at least one other lineup while not making a foil selection. This led to a diagnosticity ratio of 23.34 (96% guilty). Clearly, identification of a suspect from multiple independent lineups is highly diagnostic of guilt, especially if no foil choices were made by the witness.

Exposing witnesses to additional lineups comprised of parts of the suspect clearly increased our ability to assess the reliability of an eyewitness' identification of the suspect. The results indicate a dramatic increase in probability of guilt as the number of times that the suspect is selected by a single witness from independent lineups increases and as the number of foil choices decreases.

Despite highly encouraging results, Experiment 1 has many limitations. First, 75% of witnesses successfully chose the target from the facial lineup, but only 55% of witnesses selected the target from the facial lineup and either or both of the voice and body lineups. If a correct multiple identification is suggested as an alternative definition of correct identification, police and courts may be reluctant to use extra lineups. If the relative rates of correct identification for this experiment (75% for face only and 55% for multiple identifications) generalized to the real world, this would lead to a dramatic reduction in the number of witnesses who would be considered to have made positive suspect identifications. An important factor in the acceptance of sequential lineups as a substitute for simultaneous lineups by police was the generally similar correct identification rates for the two procedures (Lindsay, 1999a). However, this reduction need not be a problem because

Table 2  
*Suspect Choices as a Function of Target Presence or Absence (Experiment 1)*

Condition	Frequencies				Proportions—diagnosticity			
	None	1	2	3	None	1	2	3
Present	18	13	28	9	.26	.19	.41	.13
Absent	63	11	4	1	.80	.14	.05	.01
<i>DR</i>					3.01 <sup>a</sup>	1.37	8.06	10.15
% Guilty					25	58	89	91

Note. *DR* = diagnosticity ratio.

<sup>a</sup> Diagnosticity of nonidentifications is the ratio of lineup rejections in the absent condition to lineup rejections in the present condition.

additional selections could be viewed as a way of calibrating the strength of an identification, not as necessary in order to deem the face choice as acceptable as evidence.

Second, the results contain a large number of foil choices, with more than four out of every five witnesses in the target-present condition selecting a foil. To the extent that known errors (i.e., foil choices) could be used to discredit the witness, multiple lineups may weaken criminal cases by providing more opportunities for witnesses to make such errors. On the other hand, as the results also revealed, suspect identifications in the absence of foil choices are more diagnostic of guilt. The foil selection issue can be addressed in the same manner as the suspect selection issue. The more suspect choices and the fewer foil choices made by the witness, the higher the probability that the suspect is guilty. Thus, allowing for multiple foil selections may simply be another aspect of calibrating eyewitness accuracy.

Third, the study is limited by several methodological features. Only a single target and set of lineups was used, and thus the results could be specific to this unique set of materials. Further research is needed with other materials to convincingly demonstrate that the described effects do not lack target generalizability (Wells & Windschitl, 1999). In addition, the face and body lineups were only presented simultaneously. As discussed earlier, sequential lineups have been shown to reduce false positive choices and produce higher diagnosticity ratios (Lindsay & Wells, 1985; Lindsay et al., 1991; Steblay et al., 2001). We have not yet demonstrated that multiple independent identification decisions are superior to sequential lineups or that they are useful when used with sequential lineups. Indeed, presenting all of the lineups sequentially might dramatically reduce the high rate of foil choices and lead to even more diagnostic decisions. Classroom presentation of the event risks inflation of the error rate because of some students not attending to the target (unlikely given the high rate of face identification) and the potential for students to influence each others' decisions (even though they were instructed to respond independently).

Classroom presentation of the target does not simulate a crime scenario and thus may limit the results in some ways as well. However, the use of a noncrime as the stimulus does not limit generalizability to the real world entirely. Many real world situations generate eyewitnesses who are not aware that a crime is occurring when the criminal is visible to them. Victims of fraud discover only later that the crime occurred. Bank robberies often are obvious only to the teller handed a note. Other employees and customers may only be informed that a crime has occurred after the robber is no longer visible. Bystanders may see a person run out of a store or down the street with no idea that a crime has been committed but later be asked by police if they can identify the person they saw. Despite these possibilities, many crimes occur that are obvious to witnesses. Ideally, the research should provide a variety of exposures to a variety of targets to increase generalizability to real-world contexts where crimes occur in a great many variations. To the extent that results are replicable across different targets presented in different ways, generalizability is more likely.

Another issue was that not all of the diagnosticity ratios for multiple identifications were as high as the diagnosticity ratio found for multiple identifications obtained by Lindsay et al. (1987) in the first study to attempt this approach (13.60, 93% guilty). Only when face and body were identified did the diagnosticity approx-

imate the Lindsay et al. (1987) result. The cause for the lower diagnosticity ratios in Experiment 1 is unknown, but it may be that the high rate of choosing overall led to coincidental multiple identifications in the absent condition. Alternatively, the diagnosticity ratio for multiple identifications that involve voice might be lower because voice lineups have lower individual diagnosticity than clothing lineups. A final possibility is that clothing identifications (used as the second lineup by Lindsay et al., 1987) are more diagnostic than voice or body identifications.

## Experiment 2

Each of the limitations described above is addressed to some degree in Experiment 2. A fourth lineup (clothing) was added in hopes of increasing the proportion of multiple identification in the target-present condition, all lineups were presented sequentially to reduce foil choices, a strong admonishment to avoid mistaken identifications was added to the lineups subsequent to the face lineup in an attempt to further reduce foil choices (the face lineup did not include the strong warning about false positives to preserve similarity to standard sequential lineup procedures), a different confederate and experimenter conducted the research using the traditional laboratory approach of staging a live crime (purse theft) for individual witnesses.

### *Clothing Identification Literature*

Lindsay et al. (1987) conducted a series of experiments that involved asking small groups of participants to witness a staged crime. These studies were designed to test the impact of clothing bias on identification accuracy from lineups. Clothing bias exists when only the suspect in the lineup is attired in clothing similar to that described as worn by the criminal during the crime. Collectively, the experiments demonstrated that clothing bias increased identifications of innocent suspects but not the target.

In the last study of the series, Lindsay et al. (1987) suggested that clothing information could be useful if tested differently. The authors recommended seeking identification of the clothing separately from identification of the face. The participants were asked to identify the face of the targets in one lineup and the clothing of that target in a separate, clothing only lineup. Lindsay et al. (1987) reasoned that the principles of person identification should be equally applicable to clothing identification. If the suspect's clothing is embedded in a lineup with other clothes of similar style and color, then it should be possible to use an independent identification of clothing as an indication of guilt or innocence.

Comparisons were made between the data of those witnesses who did and did not identify the suspect's clothing. Of those participants who did not identify the suspect's clothing, 21% successfully selected the guilty suspect's face, whereas 11% selected the innocent suspect's face, a diagnosticity ratio of 1.91. Of those eyewitnesses who did identify the suspect's clothes, 68% identified the guilty suspect's face but only 5% identified the innocent suspect, a diagnosticity ratio of 13.60 (93% guilty).

Experiment 2 conceptually replicates Experiment 1 (use of multiple independent lineups) but under different conditions. First, there were four features used for individual lineups rather than three, including face, voice, body, and clothing. Second, the lineups were presented sequentially rather than a mixture of sequential

and simultaneous lineups. Third, the materials were presented by computer rather than manually. Fourth, a live staged crime (more typical of eyewitness research) was used.

### Method

#### Participants

Participants ( $N = 90$ ) were recruited from a 1st year introductory psychology course. Course credit was awarded to the participants in exchange for their participation. To avoid issues related to cross-race identification (Ng & Lindsay, 1994), participants who were obviously not of European ancestry in the opinion of the researcher were diverted to an alternate study.

#### Materials

The materials consisted of photos of faces, bodies, and clothing, and taped voice samples (as described for Experiment 1) of the target, the target's replacement, and foils. The target-present and -absent clothing lineups presented photos of grey sweatshirts, each with the word *Queen's* written on the front. The sweatshirts were displayed on hangers, and the backgrounds were edited electronically to be identical.

The recruitment process for the foils for Experiment 2 was similar to that for Experiment 1. In this case, a male graduate student in psychology (PD) staged thefts for unsuspecting participants. Once this student was enlisted, it became necessary to find others who roughly matched his appearance. The lineup members were selected mostly on the basis of facial similarity to the confederate, as in Experiment 1.

#### Procedure

After arriving at the laboratory, participants were left alone in a room while the experimenter retrieved some forms. During this time, the target entered the room and interacted with the participant until eye contact had been established twice (no more than 1 min), stole a purse that had been left by the experimenter, and fled. The experimenter returned immediately and informed participants of the purpose of the experiment and asked them to sign a consent form. Participants then wrote a description of the target and completed the identification procedures.

All identification instructions and stimuli were presented via computer. The face identification task was completed first. Witnesses were given standard sequential lineup instructions, including the warning that the criminal's photograph may or may not be shown. Participants were asked to use the mouse to indicate whether each photograph was of the thief and were warned that once they had responded, they would not be permitted to see the picture again.

Once participants completed the facial lineup task, the instructions for the body identification task appeared on the screen. The instructions told participants that forms of identification, other than with faces, are possible and that they would be exposed to additional lineups to attempt to identify the thief's body, voice, and clothing, in that order. The body lineup instructions were the same as those for the face, with the addition of the following warning: "SPECIAL NOTE: Choosing the wrong body would decrease the credibility of any other evidence you have provided. REMEMBER, CHOOSING NO ONE IS BETTER THAN AN INCORRECT CHOICE!"

Once the participant completed the body lineup task, the instructions for the voice identification task appeared on the screen. Again, the instructions indicated that participants would be asked to attempt to identify the voice of the thief by listening to recordings of men's voices one at a time. Each recording was approximately 1 min in duration and could be replayed if necessary. Participants were to indicate whether each voice was of the thief and were warned that once they had indicated a decision, they could not

listen to the voice again. Again, participants were warned of the dangers of making inaccurate choices (as above).

Finally, the instructions for the clothing lineup task appeared on the screen. The instructions were comparable to the body and voice instructions, including the warning about false positive choices. Once the participants completed the clothing lineup task, they were debriefed by the experimenter.

### Results and Discussion

Designation of innocent suspects for the face, body, and voice lineups was dealt with in a similar manner as for Experiment 1. Thus, the person whose face was selected most often from the target-absent face lineup was considered as the innocent suspect. However, for some purposes, a worst case scenario was followed such that another member of the target-absent lineup was designated as the innocent suspect if the results were more negative using that approach. Designating a shirt as belonging to the innocent suspect presented a new problem. The voices and bodies are clearly associated with the faces of the lineup members, but this was the case for clothing only for the target. Therefore, the six sweatshirts in the absent lineup were randomly assigned to lineup members using the reasoning that, in a real case, the shirt worn by an innocent suspect would be randomly similar to the one actually seen by the witness.

Before addressing the data for multiple identifications, the individual diagnosticity for each of the four lineups was examined. Correct facial identifications were obtained from 70.0% of participants in the present condition, whereas the most selected innocent face was chosen by 15.0% of participants in the absent condition, resulting in an overall diagnosticity ratio of 4.67 (82% guilty) for the face lineup. Percentages for the voice lineup are comparable, with 13.3% correct identifications, 35.0% false identifications, and a diagnosticity ratio of 0.38 (28% guilty). For the body lineup, the percentages are 23.3% correct identifications, 11.6% false identifications, and a diagnosticity ratio of 2.01 (67% guilty), and for the clothing lineup, the percentages are 50.0% and 10.0%, respectively, and a diagnosticity ratio of 5.00 (83% guilty), respectively.

Once again, examining the decisions for multiple suspect identifications produces a fairly positive message (Table 3). Identifying only the face (1.72, 63% guilty), body (0.33, 25% guilty), or clothing (1.52, 70% guilty) resulted in relatively low diagnosticity ratios and expectations of guilt. No participant identified only the voice of either the target or innocent suspect, nor did any participant select either suspect from any combination of lineups that did not include the face. Thus, all combinations of body, voice, and clothing identifications in the absence of face choices were absent from the data.

Combinations including the face occurred frequently and generally were highly diagnostic. Of seven possible combinations, one never occurred for either condition (face, body, and voice). The diagnosticity of the face and clothing decisions was 7.06 (88% guilty). All other combinations generated undefined diagnosticity ratios and 100% estimates of likelihood of guilt using the multiple independent lineup procedure and obtaining the observed pattern of choices.

Again, the data were examined to determine the diagnosticity of selecting no one, the suspect's face only, or the face plus one, two, or three additional features (see Table 4). Not choosing the suspect's face was diagnostic of innocence (2.83, 74% innocent),

Table 3  
*Diagnosticity Ratios (DRs) of Lineup Selections in Combinations for Designated Suspect, Worst Case Scenario, and Multiple Voice Selections (Experiment 2)*

Suspect choice	Estimated				
	DR	% Guilty	Worst case	Pairwise	Listwise
Not selected	5.22	16	0.98	3.11	
Face only	1.72	63	1.50	2.02	
Body only	0.33	25	0.33	1.52	
Voice only					
Face and body	—	100	4.12		
Face and voice	—	100	2.06		
Face and clothes	7.06	88	7.06	8.31	
Face, body, and voice					
Face, body, and clothes	—	100	—		
Face, voice, and clothes	—	100	—	—	
Body and voice			0.00		
Body and clothes					
Voice and clothes			0.00		
Body, voice, and clothes					
Face, body, voice, clothes	—	100	—	—	

*Note.* Dashes indicate undefined values because the frequency in the absent condition only was zero. Blank cells indicate that the frequency in both the present and absent conditions was zero.

selecting only the face was somewhat diagnostic (1.72, 63% guilty), and selecting the suspect’s face plus another feature was highly diagnostic (10.09, 91% guilty). If the suspect’s face and two or three other features were selected, the diagnosticity ratio was undefined (100% guilty). Again, the pattern on which these descriptive results are based is unlikely to have occurred by chance,  $\chi^2(2, N = 90) = 33.10, p < .001, r = .61$ . Deleting nonchoosers does not eliminate the effect,  $\chi^2(1, N = 30) = 6.21, p < .05, r = .45$ . The primary reason that these results are as strong as they are is that the sequential face lineup led to only 9 of 60 witnesses selecting the most frequently identified face in the target-absent condition.

In the absent condition of Experiment 1, 58% of witnesses made two or more foil choices compared with 31% of witnesses in the absent condition of Experiment 2. However, only 18% of witnesses in the present condition of Experiment 1 made two or more foil choices, compared with 43% of witnesses in the present condition of Experiment 2. It is unclear why the warning against

excessive choosing and the use of sequential lineups suppressed foil choice rates for the absent condition of Experiment 2 but not for the present condition. Regardless of the reason for this finding, the effect eliminated any relationship between foil choices and accuracy and rendered listwise deletion of witnesses equivalent to having no identification evidence.

The second experiment demonstrated that it was possible to increase the diagnosticity of identifications when sequential lineups were used. However, there still was a dramatic increase in diagnosticity of identification decisions as the number of times that the suspect was selected from different sequential lineups by a single witness increased.

### General Discussion

Lindsay et al. (1987) demonstrated that presenting more than one lineup to a witness could improve the diagnosticity of the witness’s identification decisions under one condition: when a

Table 4  
*Frequency, Proportion, Diagnosticity, and Expected Percentage Guilty on the Basis of Witnesses Identifying the Suspect From Zero to Four Times*

Condition	No. Frequencies					Proportion–diagnosticity				
	None	1	2	3	4	None	1	2	3	4
Present	9	6	10	3	2	.30	.20	.33	.10	.07
Absent	51	7	2	0	0	.85	.12	.03	.00	.00
DR						2.83 <sup>a</sup>	1.71	10.09	—	—
% Guilty						26	63	91	100	100

*Note.* Dashes indicate undefined values because the frequency in the absent condition only was zero. DR = diagnosticity ratio.

<sup>a</sup> Diagnosticity of nonidentifications is the ratio of lineup rejections in the absent condition to lineup rejections in the present condition.

lineup that was diagnostic of accuracy on its own was used in addition to a face lineup. It was unclear from their study whether adding more than one additional lineup would lead to a number of coincidental identifications of the most frequently selected person in the target-absent facial lineup. It was equally unclear if lineups that were not individually diagnostic could be useful when presented in combination with other lineups. Further, Lindsay et al. (1987) used a lineup for a feature that was only weakly related to the suspect so that it remained unclear whether the procedure would be effective with a lineup that used different parts of the suspect. In this situation, witnesses might have been able to “spoil” the independent nature of these individual pieces of evidence by matching up the voice, body, and face of the lineup members.

All of these issues have been addressed to some degree in the current studies. The results support the contention that multiple independent identification decisions by the same witness can be highly diagnostic of guilt. The use of a multiple independent identification procedure provides evidence that calibrates witness accuracy by providing a variety of patterns of response that are associated with varying expectations of witness accuracy. The more suspect identifications and the fewer foil selections, the more likely the suspect is to be guilty.

Numerous caveats need to be kept in mind. Despite a clear replication of the success of the multiple independent lineup approach in Experiment 2, only two confederates and lineups and two variations of presenting the targets to witnesses were used (three if the Lindsay et al., 1987, study is included), leaving generalizability an issue. The differences in the pattern of results between the two current experiments suggest the need for more research to determine the best procedure to use. In addition, simultaneous lineups generally preclude multiple choices. If simultaneous lineups were used in combination with the strong admonition to avoid false positives, high levels of both diagnosticity and calibration may be obtained without excessive threats to witness credibility. Obviously, multiple false identifications could be obtained if the lineups were biased. We dismiss this issue as generally uninteresting because any identification procedure will produce evidence of questionable validity if police conduct the procedure dishonestly or incompetently (Lindsay, 1994; Phillips, McAuliff, Kovera, & Cutler, 1999).

The use of clothing lineups poses particular problems because clothing can not be unambiguously associated with only one person. For example, it is possible that a very high rate of target-absent, suspect clothing identification could occur in a real-world case because the innocent suspect, by chance, was wearing the same style or even the exact same item of clothing worn by the criminal during the crime. Thus, a criminal wearing a university sweatshirt is almost certain not to be the only person who wears that exact style of clothing. It is not reasonable to expect that a witness would be able to distinguish between the target's shirt and another identical shirt in most cases (wear, staining, or damage aside). In addition, we have no data regarding clothing identification using less distinctive forms of clothing. Would witnesses even attempt to select which of several plain white shirts a criminal had been wearing? If they did, would they be able to do so with any useful level of accuracy? For these reasons, it may be better to treat clothing identification as a source of evidence but not as part of suspect identification.

The reaction of police and the courts to single versus multiple identifications by the same witness is unknown at this time. If failure to obtain multiple identifications was treated as failure to identify at all, there would be a dramatic reduction in identification evidence available to prosecute cases, and police and prosecutors would almost certainly oppose the procedure. However, this need not be the case. If multiple identifications are looked on as stronger evidence than face or whole person identifications alone, then face identifications alone could be presented but should have less weight because they are less diagnostic and thus, in legal terms, have lower probative value. Other patterns also may present problems. Foil choices could be used successfully by the defense to discredit witnesses. From a scientific perspective, it should be possible with more research to establish some guidelines to the probability that a pattern of identification is associated with guilt versus innocence and to provide this calibrating information to the courts. It is not clear whether judges and jurors are capable of applying such information usefully to the binary decisions of guilt and innocence that they must make. However, the information would be more diagnostic of guilt than the current binary decisions to choose or not choose the suspect. It is difficult to see how the system can be hurt by better information. The use and interpretation of single versus multiple identifications by police and the courts is beyond the scope of this article but is an important factor if using multiple independent identification decisions is to be proposed in the future. Mock jury studies of such issues may be a first step in determining the impact of such evidence.

However, there are some clear advantages to presenting multiple independent identification evidence in court. The false identification rate for multiple identifications (3% in our two studies) is much lower than the overall false identification rate found in studies of sequential (9%) or simultaneous (27%) lineups (Stebly et al., 2001). Using multiple independent identification procedures would obviously reduce the possibility of wrongful conviction. These results rely on any multiple identification. As the number of suspect choices increases, the false identification rate decreases even further (1 of 150 witnesses in our two studies). As a result, the probative value of the identification evidence would be higher the more suspect identifications the witness made. If the criminal justice system is committed to the goal of exonerating the innocent, multiple independent lineups potentially provide the best method yet studied to achieve this goal without unacceptably high loss of conviction of the guilty.

A final issue is the admissibility of such evidence. Because there is no precedent for such procedures, clearly we can not claim that the courts would accept eyewitness decisions from multiple independent lineups as evidence. However, facial and voice identification already are admitted as independent evidence of identification. Body identification may be the most controversial—we know of no case in which such evidence has been admitted. The notion that multiple identification is better evidence than a single identification should not be conceptually difficult to sell to the courts. It is based on a principle the courts already accept: corroboration.

More than a quarter of a century ago, the Devlin commission concluded that a single eyewitness identification of a suspect was dangerous evidence on which to base a criminal conviction (Devlin, 1976). Devlin urged the courts to demand some form of corroboration (i.e., independent evidence of guilt other than the single identification). Multiple independent lineups provide the

possibility of “corroborated” and highly probative evidence even in cases in which the sole source of information is the memory for the criminal provided by a single eyewitness.

## References

- Bisbing, B. E. (1982). The forensic identification and association of human hair. In R. Saferstein (Ed.), *Forensic science handbook* (pp. 184–221). New Jersey: Prentice Hall.
- Bradfield, A. L., Wells, G. L., & Olson, E. A. (2002). The damaging effect of confirming feedback on the relation between eyewitness certainty and identification accuracy. *Journal of Applied Psychology, 87*, 112–120.
- Bull, R. (1978). Eyewitnesses also have ears. In M. M. Gruneberg, P. E. Morris, & R. N. Sykes (Eds.), *Practical aspects of memory*. London: Academic Press.
- Bull, R. (1981). Voice identification by man and machine: A review of research. In S. Lloyd-Bostock (Ed.), *Psychology in legal contexts*. London: Macmillan.
- Bull, R., & Clifford, B. R. (1984). Ear witness voice recognition accuracy. In G. L. Wells & E. F. Loftus (Eds.), *Eyewitness testimony: Psychological perspective* (pp. 92–123). New York: Cambridge University Press.
- Connors, E., Lundregan, T., Miller, N., & McEwen, T. (1996). *Convicted by juries, exonerated by science: Case studies in the use of DNA evidence to establish innocence after trial*. Washington, DC: National Institute of Justice.
- Cutler, B. L., Berman, G. L., Penrod, S., & Fisher, R. P. (1994). Conceptual, practical, and empirical issues associated with eyewitness identification test media. In D. F. Ross, J. D. Read, & M. P. Toglia (Eds.), *Adult eyewitness testimony: Current trends and developments* (pp. 163–181). New York: Cambridge University Press.
- Cutler, B. L., & Fisher, R. P. (1990). Live lineups, videotaped lineups, and photoarrays. *Forensic Reports, 3*, 439–448.
- Cutler, B. L., & Penrod, S. D. (1988). Improving the reliability of eyewitness identification: Lineup construction and presentation. *Journal of Applied Psychology, 73*, 281–290.
- Devlin, P. (1976). *Report to the secretary of state for the home department of the committee on evidence of identification in criminal cases*. London: Her Majesty’s Stationery Office.
- Dunning, D., & Stern, L. B. (1994). Distinguishing accurate from inaccurate eyewitness identifications via inquiries about decision processes. *Journal of Personality and Social Psychology, 67*, 818–835.
- Durand, H. (1999). Fingerprints. In G. M. Chayko & E. D. Gulliver (Eds.), *Forensic evidence in Canada* (pp. 455–474). Aurora, Canada: Canada Law Book.
- Egan, D., Pitner, M., & Goldstein, A. G. (1977). Eyewitness identification: Photographs versus live models. *Law and Human Behavior, 1*, 199–206.
- Leippe, M. R. (1980). Effects of integrative memorial and cognitive processes on the correspondence of eyewitness accuracy and confidence. *Law and Human Behavior, 4*, 261–274.
- Lindsay, R. C. L. (1994). Biased lineups: Where do they come from? In D. F. Ross, M. Toglia, & J. D. Read (Eds.), *Adult eyewitness testimony* (pp. 182–200). New York: Cambridge University Press.
- Lindsay, R. C. L. (1999a). Applying applied research: Selling the sequential lineup. *Applied Cognitive Psychology, 13*, 219–225.
- Lindsay, R. C. L. (1999b). Eyewitness evidence. In G. M. Chayko & E. D. Gulliver (Eds.), *Forensic evidence in Canada* (pp. 157–211). Aurora, Canada: Canada Law Book.
- Lindsay, R. C. L., & Bellinger, K. (1999). Alternatives to the sequential lineup: The importance of controlling the pictures. *Journal of Applied Psychology, 84*, 315–321.
- Lindsay, R. C. L., Lea, J. A., Nosworthy, G. J., Fulford, J. A., Hector, J., LeVan, V., et al. (1991). Biased lineups: Sequential presentation reduces the problem. *Journal of Applied Psychology, 76*, 796–802.
- Lindsay, R. C. L., Martin, R., & Webber, L. (1994). Default values in eyewitness descriptions: A problem for the match-to-description lineup foil selection strategy. *Law and Human Behavior, 18*, 527–541.
- Lindsay, R. C. L., Pozzulo, J., Craig, W., Lee, K., & Corber, S. (1997). Simultaneous lineups, sequential lineups, and showups: Eyewitness identification decisions of adults and children. *Law and Human Behavior, 21*, 391–404.
- Lindsay, R. C. L., & Smith, S. M. (2001). *Race and identification procedure: Does the sequential lineup effect generalize?* Paper presented at the Fourth Biennial Conference of the Society for Applied Research in Memory and Cognition, Kingston, Ontario, Canada.
- Lindsay, R. C. L., Wallbridge, H., & Drennan, D. (1987). Do the clothes make the man? An exploration of the effect of lineup attire on eyewitness identification accuracy. *Canadian Journal of Behavioural Science, 19*, 464–478.
- Lindsay, R. C. L., & Wells, G. L. (1985). Improving eyewitness identifications from lineups: Simultaneous versus sequential lineup presentations. *Journal of Applied Psychology, 70*, 556–564.
- Luus, C. A. E., & Wells, G. L. (1994). The malleability of eyewitness confidence: Co-witness and perseverance effects. *Journal of Applied Psychology, 79*, 714–724.
- MacLeod, M. D., Frowley, J. N., & Shepherd, J. W. (1994). Whole body information: Its relevance to eyewitnesses. In D. F. Ross, M. Toglia, & J. D. Read (Eds.), *Adult eyewitness testimony* (pp. 125–143). New York: Cambridge University Press.
- McGehee, F. (1937). The reliability of the identification of the human voice. *Journal of General Psychology, 17*, 249–271.
- McGehee, F. (1944). An experimental study of voice recognition. *Journal of General Psychology, 31*, 53–65.
- Melara, R. D., DeWitt-Rickards, T. S., & O’Brien, T. P. (1989). Enhancing lineup identification accuracy: Two codes are better than one. *Journal of Applied Psychology, 74*, 706–713.
- Munsterberg, H. (1908). *On the witness stand*. New York: Doubleday.
- Ng, W. J., & Lindsay, R. C. L. (1994). Cross-race facial recognition: Failure of the contact hypothesis. *Journal of Cross-Cultural Psychology, 25*, 217–232.
- Orchard, T. L., & Yarmey, A. D. (1995). The effects of whispers, voice-sample duration, and voice distinctiveness on criminal speaker identification. *Applied Cognitive Psychology, 9*, 249–260.
- Phillips, M. R., McAuliff, B. D., Kovera, M. B., & Cutler, B. L. (1999). Double-blind photoarray administration as a safeguard against investigator bias. *Journal of Applied Psychology, 84*, 940–951.
- Pollack, I., Pickett, J., & Sumbly, W. (1954). On the identification of speakers by voice. *Journal of the Acoustical Society of America, 64*, 403–406.
- Pozzulo, J. D., & Lindsay, R. C. L. (1995). *Eyewitness identification procedures when appearance has been changed*. Paper presented at the First Biennial Conference of the Society for Applied Research in Memory and Cognition, Vancouver, British Columbia, Canada.
- Reich, A., Moll, K., & Curtis, J. (1976). Effects of selected vocal disguise upon spectrographic speaker identification. *Journal of the Acoustical Society of America, 60*, 919–925.
- Rosenthal, R. (1991). *Meta-analytic procedures for social research*. London: Sage.
- Saslove, H., & Yarmey, A. D. (1980). Long term auditory memory: Speaker identificatory memory. *Speaker identificatory, 65*, 111–116.
- Smith, S. M., Lindsay, R. C. L., & Pryke, S. P. (2000). Postdictors of eyewitness accuracy: Can false ID’s be diagnosed? *Journal of Applied Psychology, 14*, 81–96.
- Sporer, S. L. (1993). Eyewitness identification accuracy, confidence, and decision times in simultaneous and sequential lineups. *Journal of Applied Psychology, 78*, 22–33.
- Stebly, N., Dysart, J. E., Fulero, S., & Lindsay, R. C. L. (2001). Eyewitness accuracy rates in sequential and simultaneous lineup presentations: A meta-analytic comparison. *Law and Human Behavior, 25*, 459–473.

- Technical Working Group for Eyewitness Evidence. (1999). *Eyewitness evidence: A guide for law enforcement*. Washington, DC: U.S. Department of Justice, Office of Justice Programs, National Institute of Justice.
- Wells, G. L. (1978). Applied eyewitness testimony research: System variables and estimator variables. *Journal of Personality and Social Psychology*, *36*, 1546–1557.
- Wells, G. L. (1984). How adequate is human intuition for judging eyewitness testimony? In G. L. Wells & E. F. Loftus (Eds.), *Eyewitness testimony: Psychological perspectives*. New York: Cambridge University Press.
- Wells, G. L., & Bradfield, A. L. (1998). "Good, you identified the suspect": Feedback to eyewitnesses distorts their reports of the witnessing experience. *Journal of Applied Psychology*, *83*, 360–376.
- Wells, G. L., Ferguson, T. J., & Lindsay, R. C. L. (1981). The tractability of eyewitness confidence and its implication for triers of fact. *Journal of Applied Psychology*, *66*, 688–696.
- Wells, G. L., & Lindsay, R. C. L. (1980). On estimating the diagnosticity of eyewitness nonidentifications. *Psychological Bulletin*, *88*, 776–784.
- Wells, G. L., Lindsay, R. C. L., & Ferguson, T. J. (1979). Accuracy, confidence, and juror perceptions in eyewitness identification. *Journal of Applied Psychology*, *70*, 413–419.
- Wells, G. L., & Olson, E. A. (2002). Eyewitness identification: Information gain from incriminating and exonerating behaviors. *Journal of Experimental Psychology: Applied*, *8*, 155–167.
- Wells, G. L., Rydell, S. M., & Seelau, E. P. (1993). The selection of distractors for eyewitness lineups. *Journal of Applied Psychology*, *78*, 835–844.
- Wells, G. L., & Turtle, J. (1986). Eyewitness identification: The importance of lineup models. *Psychological Bulletin*, *99*, 320–329.
- Wells, G. L., & Windschitl, P. D. (1999). Stimulus sampling and social psychological experimentation. *Personality and Social Psychology Bulletin*, *25*, 1115–1125.
- Whipple, G. M. (1909). The observer as reporter: A survey of the psychology of testimony. *Psychological Bulletin*, *6*, 153–170.
- Yarmey, A. D. (1991). Voice identification over the telephone. *Journal of Applied Social Psychology*, *21*, 1868–1876.
- Yarmey, A. D., & Matthys, E. (1992). Voice identification of an abductor. *Applied Cognitive Psychology*, *6*, 367–377.

Received August 29, 2002

Revision received February 28, 2003

Accepted April 16, 2003 ■

### **Members of Underrepresented Groups: Reviewers for Journal Manuscripts Wanted**

If you are interested in reviewing manuscripts for APA journals, the APA Publications and Communications Board would like to invite your participation. Manuscript reviewers are vital to the publications process. As a reviewer, you would gain valuable experience in publishing. The P&C Board is particularly interested in encouraging members of underrepresented groups to participate more in this process.

If you are interested in reviewing manuscripts, please write to Demarie Jackson at the address below. Please note the following important points:

- To be selected as a reviewer, you must have published articles in peer-reviewed journals. The experience of publishing provides a reviewer with the basis for preparing a thorough, objective review.
- To be selected, it is critical to be a regular reader of the five to six empirical journals that are most central to the area or journal for which you would like to review. Current knowledge of recently published research provides a reviewer with the knowledge base to evaluate a new submission within the context of existing research.
- To select the appropriate reviewers for each manuscript, the editor needs detailed information. Please include with your letter your vita. In your letter, please identify which APA journal(s) you are interested in, and describe your area of expertise. Be as specific as possible. For example, "social psychology" is not sufficient—you would need to specify "social cognition" or "attitude change" as well.
- Reviewing a manuscript takes time (1–4 hours per manuscript reviewed). If you are selected to review a manuscript, be prepared to invest the necessary time to evaluate the manuscript thoroughly.

Write to Demarie Jackson, Journals Office, American Psychological Association, 750 First Street, NE, Washington, DC 20002-4242.