Determinants of eyewitness performance
on a lineup*

ROBERT BUCKHOUT, ANDREA ALPER, SUSAN CHERN,
GLENN SILVERBERG, and MIRIAM SLOMOVITS
Center for Responsive Psychology, Brooklyn College, CUNY, Brooklyn, New York 11210

A simulated crime was staged before 64 witnesses in order to generate eyewitness reports to be analyzed for accuracy. Forty-eight of the witnesses attempted to pick out the suspect from two videotaped lineups—one with and one without the suspect. Successful witnesses showed significantly fewer errors of commission than those who picked the suspect but impeached their identification with another choice. There were 13.5% positive identifications, 13.5% impeached, 40.3% mistaken identifications, and 19.2% nonidentifications. An analysis was made of the differences between successful and unsuccessful eyewitnesses.

Staging a simulated "crime" in a classroom has become a standard teaching device in psychology and law school classes to demonstrate the fallibility of the eyewitness (Buckhout et al, 1973). The demonstration works so well that we rarely see hard data in print, a deficiency which has hampered the application of well established psychological theories of human perception in the handling of eyewitness testimony in the courts (Marshall, 1969). Observers of the legal setting, such as Wall (1966), effectively generalize the voluminous technical literature on memory and perception to the unique situation of crime, eyewitness reports, mug shot and lineup recognition, and court testimony, but stress the need for more specific research to demonstrate the role of classic research findings in a more relevant experimental design.

In previous research, Buckhout et al (1973) demonstrated the usual poor recall of 141 witnesses and showed that relatively few witnesses were able to pick out a suspect they had previously seen from a photo spread, except under conditions in which the photo spread was biased. We endeavored in the present study to create a simulated crime, test for recall, and administer lineups to witnesses under controlled conditions. It was predicted that the number of witnesses making a correct positive identification would be small and that successful witnesses would be those who demonstrated better recall immediately after the "crime."

METHOD

Subjects

The witnesses were 64 Brooklyn College undergraduate students enrolled in two separate introductory psychology courses; 52 witnesses were tested in the lineups.

Materials

A 40-item questionnaire and two videotaped lineups were

employed. The questionnaire measured the Ss' recall of the details of a purse-snatching incident, including a description of the suspect, the events, the amount of stress experienced, and the degree of confidence in their answers. From the list of correct answers, errors of omission, errors of commission, and estimates of incident time were computed.

The lineups were videotaped for the exepriment. Each lineup included five male college students selected on the basis of their resemblance to the suspect. The blank lineup (B) did not show the suspect, but rather a close look-a-like. The valid lineup (V) consisted of the suspect and four other male students. It did not include the look-a-like. Each lineup showed a long shot of the five males as a group at both the beginning and the end of the tape, and individual close-ups of approximately equal duration. Each lineup lasted 140 sec.

Procedure

The incident was staged in two Brooklyn College classes, midway through a lecture period. A male student appeared at the door and stated that he thought he had left a book in the room during the preceding class and that he would like to look for it. The instructor, who was aware that the incident was staged, agreed to the student's request. The male student (the assailant) then walked to a chair where a confederate was seated and bent over as if to look for a book. He grabbed the confederate's pocketbook and fled out the door. The confederate screamed as the assailant ran out of the class. The instructor cautioned the class to remain seated, and ran after the assailant. A few minutes later, the instructor and the Es returned and informed the students of the experimental nature of the incidents, and distributed questionnaires. Three weeks later, the Es again returned to the classes and distributed another set of questionnaires which asked the Ss to indicate their confidence in their ability to correctly identify the purse snatcher. The Ss were then shown the two videotaped lineups and asked to pick out the assailant. Order effects were avoided by counterbalancing: half of the Ss (randomly determined) viewed the blank lineup first, while the other half viewed the valid lineup first. The Ss were allowed to see each lineup as many times as they needed to make a judgment—a concession to standard police lineup procedure.

RESULTS

Data from the recall questionnaire (see Table 1) show the averaged comparison of those unsuccessful on the valid lineup against those who correctly identified the

Table 1
Summary of Recall Data

	Actual Data	Unsuccessful on Lineup (N = 34)	Picked Suspect on Lineup (V) (N = 14)
Height	70 in.	68.8 in.	69.9 in.
Weight	200 lb	167.1 lb	178 lb*
Age	17 yrs	20.2 yrs	20.7 yrs
Duration of Incident	21 sec	58.8 sec	33.7 sec*
Errors of Omission	–	8.34	7.50
Errors of Commission	–	7.34	6.93
Stress		5.43*	4.84
Confidence in Description		6.22	6.15

*$p < .05$

suspect. Successful witnesses had significantly higher (and closer to the actual) weight estimates of the suspect, significantly lower (more accurate) estimates of the duration of the incident, and reported significantly less stress. Successful witnesses tended to have fewer errors of omission and commission but the trends were not significant. Looking more closely, the "successful" witnesses were divided into seven who picked only the correct suspect on the valid lineup and those (seven) who picked the suspect but then impeached their identification by also picking someone (five also picked the look-a-like) on the blank lineup. The seven successful witnesses had significantly fewer errors of commission in recall (5.57 vs 8.14) than the impeached successful witnesses (p < .01, Mann-Whitney U test). Curiously, the impeached-successful witnesses showed significantly *higher* confidence in their descriptions than the clearly successful witnesses (p < .05).

In Table 2, the overall performance data of the witnesses who showed up for the lineup reveals that only 13.5% made an identification which was veridical. The remaining witnesses impeached their identification, made mistakes, or failed to identify anyone. In the confidence measures taken just prior to the lineups (how certain are you that you can pick the man?), we found that the clearly successful witnesses expressed the same confidence as previously; the impeached witnesses had significantly *higher* confidence; and the unsuccessful witnesses had significantly *lower* confidence levels than their own previous rating during recall (p < .05).

DISCUSSION

Most witnesses thought that the simulated crime was realistic; reporting a stress level of 5.5 on a 0-10-point scale. In one class, a few witnesses took off in hot pursuit of the suspect, contrary

Table 2
Performance of Eyewitnesses on Two Lineups (N = 52)

	Percent
Positive Identification	13.5
Positive Identification Impeached	13.5
Picked Look-Alike Only	13.5
Other Mistaken Identification Only	36.5
Two Mistaken Identifications	3.8
No Identification Made	19.2

to any notions we had about apathetic bystanders. With this fact established, our eyewitnesses gave typically poor descriptions of the suspect, showing a general tendency to approximate the norm for a male college student as evidenced by the higher age estimate (matching the age of the witnesses) and the idealized height to weight proportions (5 ft 10 in., 168 lbs) which match the data on a life insurance company chart.[1] These convenient frames of reference may be turned to by witnesses who really don't recollect what they saw.

In presenting the witnesses with two lineups, we approximated the testing strategy of a psychophysicist who might employ "catch" trials or a signal detection paradigm to see whether identification was a product of perception or response set. In our experiment, as in real lineups, most witnesses (80.8%) made at least one identification. Wall (1966) and others report that the social pressure to name *someone* in a lineup can be very intense–a climate which can encourage false positives. Whether this witness behavior is a discernible personal style or is entirely situation-specific is an open question for research. Our feeling is that the lineup itself, with its formality, its authority figures, and its potential for reinforcing the witness and the police for "solving a case," encourages identifications, mistaken or not. In recently publicized crimes (*New York Times*, 1974), the mistaken identifcation of look-a-likes in a rape case seems to have resulted from excessive zeal on everyone's part.

In exploring the level of confidence of witnesses, our data clash with the widely held respect in the legal establishment for expressions of certainty by a witness, e.g., "I'll never forget his face." In the present study, as well as in our other research efforts, we have found no evidence to support this assertion. Correlations between confidence and recall accuracy have generally been low and nonsignificant. In the confidence rating expressed just prior to the lineup, we found a significant positive correlation between confidence and errors of commission (r = .35, p < .01) which goes along with the finding that the witnesses who impeached themselves had the highest amount of confidence. Errors of commission reflect the tendency for Os to perceptually fill in. Perhaps high confidence signals a witness who is only too quick to stereotype, to please the authorities and identify *someone*. Existing techniques for searching for suspects which involve the production of stylized drawings or photographs may exacerbate these tendencies and result in mistaken identifications of people who look like the idealized suspect.

What can we say of our unimpeached successful witnesses? It is clear that while they omitted as many details in their description as other witnesses, they had significantly fewer errors of commission. Additionally, the good witnesses were sensitive to the higher than typical weight of the suspect. The number of such witnesses was small, as it has been in most of our lineup research, a fact which limits generalizations. In the real world of the police lineup, where there is no veridical standard, many critics argue that mistaken identifications are very common. The present experiment is a beginning; testing such assertions under controlled conditions is possible and should be done.

REFERENCES

Buckhout, R., Figueroa, D., & Huff, E. Psychology and the eyewitness. American Journal of Criminal Law, May 1973, in press.

Marshall, J. Psychology and law in conflict. New York: Anchor, 1969.

New York Times. Victim in triple mistaken-identity case absolved after 15 months. New York Times, January 11, 1974.

Wall, P. Eyewitness identification in criminal cases. Springfield, Illinois: Thomas, 1965.

NOTE

1. Height and weight estimates were correlated (r = .43, p < .01), as are the actual data in real life.