Nearly 2,000 witnesses can be wrong

ROBERT BUCKHOUT

City University of New York, Brooklyn College, Brooklyn, New York 11210

A mock crime, a mugging and purse snatch, was staged as representative of the usually difficult observation conditions present in crime situations. The film was shown on a television newscast and to small audiences under more controlled conditions. A few moments after the crime, witnesses were shown a lineup and asked to identify the culprit. Of the television respondents, 74% attempted an identification; only 19% were correct (n.s.). The film audience witnesses were only slightly better (24.8%) than chance ($p < .05$). Analyses of recalled descriptions showed that most witnesses had seen and recalled very little of the culprit's features, but they were motivated by the demand of the lineup situation to try to identify anyway.

Most laboratory studies of visual perception involve the intensive analysis of data from hundreds of trials endured by a relatively small number of subjects. In contrast, research on the eyewitness has dictated a paradigm in which the event, a crime, is briefly presented once to a large number of subjects who are expected to recall the events, describe the "perpetrator," and then try to recognize the perpetrator at a later time. While laboratory perception research yields impressively high accuracy scores (over 90%) in testing for memory of briefly seen faces (Hochberg & Galper, 1967; Yin, 1969), eyewitness researchers commonly report recognition accuracy scores under 50% (Buckhout, Figueroa, & Hoff, 1975; Lipton, 1977; Marshall, 1966).

To an extent, the visual stimuli available for encoding in a real (or mock) crime are unique; this is a problem that hampers comparisons from one crime to another. The uniqueness (usually the target face) is overshadowed by the conditions for observation that can be distilled from reported crimes in which eyewitness identification is an issue. Wells (1978) refers to these elements in the crime situation as "estimator variables," whose effects on eyewitness testimony in a given case can only be estimated. Yet it is clear that the researcher who manipulates independent variables

must make that estimate in staging a mock crime that is a functional simulation of the real thing. Thus most mock crimes are staged or filmed to include (1) a less than ideal environment (e.g., poor lighting), (2) short time periods, (3) moving targets, (4) a complex set of events, (5) distracting noises and shouts, and (6) surprise. Such a "worst-case" set of conditions yields low recognition accuracy because those elements make it difficult for many witnesses to observe or encode details of the crime. Law enforcement officials acknowledge that these are some of the elements involved in many real crimes, and researchers have tended to use these elements as a means of achieving external validity in eyewitness research (Ellison & Buckhout, 1981). By contrast, a basic researcher concerned with memory for faces might reverse or ameliorate all six elements in order to make sure the subjects had an adequate opportunity to observe (and hence store) the information in long-term memory. This "best-case" approach yields higher recognition accuracy scores.

My point is that the results of both classes of research are so predictable and so well known that the phenomena have become dependable demonstrations for an instructor to stage in the classroom or on television talk shows. The worst-case demonstration generally yields poor accuracy in recall. The sometimes ludicrous descriptions given by eyewitnesses in such demonstrations may be due in part to an implicit demand that descriptions be complete. We have noted that witnesses rarely hesitate to give height, weight, and age estimates, even in free recall. Real-world witnesses and experimental subjects appear to respond to such demands whether they have stored such information or not (Buckhout, 1974). Part of the search in the present study will be devoted to exploring better methods of eliciting descriptions from eyewitnesses. As Loftus (1977) has noted, leading questions are commonly used by investigators in order to complete their visualizations of what took place. Loftus' many experiments have shown that eyewitnesses will alter their recall reports to include as "facts" items that were not pres-

The author wishes to thank Molly Sidi, WNBC-TV, New York, for her fine imaginative presentation of the documentary film of which this study became a part. I am also grateful to the fine team of undergraduate students from Brooklyn College who acted in, showed, analyzed, and helped to think through the many elements of our research program on the eyewitness. Support for this research was provided in part by Faculty Research Award Program of the City University of New York. A brief write-up about Study 1 appeared in *Social Action and the Law*, 1975, 2(3). The entire report, with supporting data, additional tables, forms, and instructions, will be available as *Center for Responsive Psychology Monograph No. CR-22*, 1980, Brooklyn College, Brooklyn, New York 11210. Reprints of this article and CR-22 can be obtained by writing to the author at the Center for Responsive Psychology, Brooklyn College, Brooklyn, New York, 11210.

ent in the original viewing. Reconstruction of memory then becomes a confabulation of the original stimulus details plus leads provided by the questions, evidence from other observers (whether true or not), and logical deductions—even though the witness may continue to feel that he has a single memory of a crime event. What is not so obvious is that even a simple request for a witness to estimate a suspect's height may constitute a leading question that encourages guessing rather than information retrieval in cases in which the witness did not encode the height.

How to question eyewitnesses for recall and how to test eyewitnesses for recognition are issues that fit into Wells' (1978) category of "system variables," variables that people in law enforcement can control. In testing for recognition, it has been generally noted that a high percentage of eyewitnesses to crime will attempt an identification when tested, whether the suspect is in the test (lineup or photospread) or is not (Buckhout & Greenwald, in press). Attempted identification rates by 80% of the total number of witnesses tested are reported by Buckhout, Alper, Chern, Silverberg, and Slomovits (1974), Buckhout et al. (1975), Loftus (1975), Fanselow and Buckhout (Note 1), and Malpass and Devine (Note 2), and the "Devlin Report" (Note 3) notes that over 50% of actual lineups in England result in some kind of attempted identification. This apparent willingness to say "yes" can, of course, be influenced by instructions, fear of making false accusations, fear of reprisal, demand effects in the test situation, coercion, or blatant bias (Buckhout, Note 4).

Asking a person how well he remembers elements of a crime situation is an inherently high-demand situation. If we accept the long history of poor quality eyewitness reports as indicative of the inadequate acquisition and storage of detail in memory, the usual approach to testing witnesses represents a mismatch of retrieval and recognition testing to the nature of the acquisition. The first study reported here represented such a high-demand setting in the form of a challenge to audiences of witnesses to a mock crime to view a lineup shortly after the crime and make an identification. We began by staging the crime before a television audience estimated to be in the millions. The second study compared the attempted identification rates and recognition accuracy scores of witnesses to the television broadcast with the data obtained from audiences to a filmed version of the same "crime." The film-audience condition allowed us to collect recall performance data, in the interest of helping to explain what we expected would be marginal recognition performance.

METHOD

Subjects

The subjects in the "television-viewer" condition were 2,145 (60% male, 40% female) viewers of a news broadcast on WNBC-TV in New York on December 19, 1974. Participation

was voluntary, as viewers called in to make a lineup identification. Subjects in the "film-audience" condition were members of audiences ranging from 20 to 100 in size at legal seminars conducted by the author shortly after the broadcast. Participation by film-audience subjects was voluntary. The median age of the audiences was 26 years; the composition was 73% male, 27% female, and 90% white.

Stimulus Films

A scenario was developed for a stimulus crime film to create a fast-moving crime, with a limited view of the face of the perpetrator of the crime, some distractions, and a high level of mundane realism. Based on statistics on street muggings in New York City, the whole incident, involving a victim being knocked to the ground and a purse being stolen, took only 13 sec, including 3.5 sec in which the face of the "perpetrator" was clearly visible. The film was shot by a professional camera crew from WNBC News, using 16-mm color/sound film and studio lighting.

The scenario has a young woman (the victim) walking down a hallway, with the perpetrator lurking in a doorway. Suddenly, the man runs up behind the woman, knocks her down, grabs her purse, and runs face forward toward the camera, while the woman screams. The perpetrator wore a hat, a leather jacket, jeans, sneakers, and a small mustache. To test recognition, a separate filmed lineup was shot by the same film crew. Five stand-ins were recruited who matched the general physical features and appearance of the perpetrator and stood with him in the lineup. The perpetrator had removed his mustache and changed to ordinary sports clothes similar to that worn by the stand-ins. The lineup film pictured an overall view of the six men standing against a wall. Each showed a number from 1 to 6 on his chest; the perpetrator was in the number 2 position.

The questionnaire used in the film-audience condition consisted of demographic items on the witness, recall questions on features of the perpetrator remembered by the witness, measures of the adequacy of the opportunity to observe, confidence ratings, and time estimates.

Procedure

The television show came first, providing an opportunity for a massive replication of the classic lecture hall demonstration of the difficulties involved in visual observation. The crime film was shown on the air during a special 5-min segment of the local news program devoted to the problems of eyewitness testimony and recognition testing. Following a discussion of the problems, the crime film was shown, and members of the viewing audience were informed that they were all witnesses to a crime. Approximately 2 min later, the announcer presented the lineup film, asking the viewers to look carefully at each man and decide if they recognized anyone. They were shown an overview of the lineup and a close-up of each man. A special phone number was flashed on the screen, and the viewers were urged to call and report their lineup choice number to one of our student volunteers. If they were convinced that the perpetrator was not in the lineup, viewers were to state the number "0." Volunteers recorded the choice number and the sex of the respondent. Viewers were urged to watch the program the next evening to check on their accuracy. On the next broadcast, viewers saw a replay of the crime, revelation of the correct answer, and a discussion of the significance of the study by the author and the announcer—a mass debriefing. The response was overwhelming, with jammed phone lines, over 2,000 recorded responses, plus numerous comments.

Subsequently, data were collected from audiences at legal conferences by showing a 16-mm print of the material from the television broadcast. The crime film was shown without warning, followed by the announcement that the members of the audience were to be considered eyewitnesses and were to fill out the questionnaire according to instructions. The film was projected on a large standard-size screen in a darkened room, providing

more controlled and presumably more optimal viewing conditions for the subjects than could have been obtained in a television broadcast. Approximately 5 min later, the members of the film audience were shown the lineup film twice and asked "if they recognized anyone." If they did, subjects were asked to check a number corresponding to the person in the lineup; they were asked to circle "0" if they felt that they could not recognize anyone. All subjects circled a confidence estimate for their decision from 0% to 100%.

The Perpetrator

The physical characteristics of the perpetrator were as follows: male, age 23 years, 5 ft 11 in. (71 in.), 145 lbs., white, tan-brown skin color, with a slight mustache during the crime but not in the lineup film. Thus we had a perpetrator who could be called tall, thin, and dark rather than average in appearance. The darker skin color is characteristic of a Sephardic Jew, even though the correct technical racial designation of the perpetrator was white. We anticipated that the fact that the perpetrator's physical appearance did not conform to either white or non-white racial stereotypes would result in some confusion among the witnesses.

RESULTS AND DISCUSSION

Recognition Performance

The distribution of lineup choices by the witnesses in both conditions is presented in Table 1. The results show a high degree of consistency in preferences among the seven individual choices by both groups; the chi square was nonsignificant. If we could assume that each of the seven possible answers (including 0) was equally likely to be made on the basis of chance, it is clear that the television viewers, in choosing the correct suspect in the lineup 14.1% of the time, were performing at about the chance level (14.29%). The film-audience viewers performed slightly better than chance, with 19.8% choosing the correct suspect. Since an assumption that each lineup choice had an equal likelihood of being chosen by chance is probably not warranted, a comparison was made between the response patterns of both groups of witnesses (see Table 2), showing that the television viewers made fewer correct choices (true positives) and more frequent false negative responses than did the film-audience members. A further analysis of the response patterns of only those witnesses who attempted identifications (see Table 3) showed that the film-audience identifiers had significantly more correct identifications than a hypothetical group of chance-performing witnesses [$\chi^2(1) = 4.30$, $p < .05$]. Thus the improved viewing conditions offered the film-audience witnesses significantly improved recognition performance, if only by a slight margin.

With recognition performance hovering around chance for nearly 2,000 television-viewer witnesses, it seemed astonishing that over 1,800 witnesses made mistaken identifications. The subjective comments of the call-in witnesses were in several categories. Some callers reported disagreement in their own households. A police sergeant called in to say that all the men in his precinct had voted on Number 4 as the perpetrator. Numerous callers reported that one or more of the noncriminal participants in our lineup had mugged the caller! Many

Table 1
Distribution of Lineup Choices by Television (T) and Audience (A) Viewers (Witnesses) of a Simulated Crime

Group	Lineup Positions						Not in Lineup
	1	2	3	4	5	6	
T	14.7	14.1	10.1	13.0	14.3	6.6	25.9
A	21.8	19.8	8.5	9.2	16.7	3.8	20.1

Note—Distribution is expressed as percentages. For television viewers, $N = 2,145$; for audience viewers, $N = 293$. The actual attacker was in Position 2. $\chi^2 = 4.799$, *n.s.*

Table 2
Distribution of Eyewitness Identification Responses to a Lineup Following a Simulated Crime by Television Viewers (T) and Film-Audience Members (A)

Group	N	TP	FP	FN
T	2145	14.1	60.0	25.9
F		19.8	60.1	20.1

Note—TP = true positives (correct identification); FP = false positives (mistaken identification); FN = false negatives (missed the perpetrator). Distribution is expressed in percentages. $\chi^2(2) = 9.127$, $p < .02$.

Table 3
Study 1: Responses of Identifiers Only, Dropping False Negative Responders

Group	N	TP	FP
Television Viewers	1589	19.0	81.0
Film-Audience Members	234	24.8	75.2
Hypothetical Chance Performance		16.7	83.3

Note—TP = true positives (correct identification); FP = false negatives (mistaken identification). Values given are percentages.

callers complained that the crime was too fast, some remarking that it would have been fairer to rerun the crime before we asked for an identification. Additionally, we found that witnesses complained that the suspects were wearing clothing different from what they had seen on the perpetrator during the crime—Then again, so do real criminals.

There are obvious limitations to the study, since (1) the televised picture quality is not very high, (2) the witnesses may have been viewing in color or in black and white (or on different-sized picture tubes), and (3) the nature and extent of sampling errors among television audience witnesses was unknown. We therefore do not make too much of the overall poor recognition performance of the witnesses, but we look rather to the attentional and social factors involved. With a short event, we are dealing with "one-shot perception" and the witness's knowledge (given by the lineup test administrator) that he or she legally could be expected to make an identification. But, of the many witnesses who attempted an identification, we can be certain that many looked but did not in fact see very much (low sensitivity), perhaps spending more viewing time looking at the attractive victim and seeing very little of the attacker's features. Still, they tried to identify, conform-

ing perhaps to the social pressure to try, in spite of their inadequate perception of the original event. As Goldstein, Johnson, and Chance (Note 5) note, "The two skills, recognition memory and reconstruction ability, had little in common" (p.12). Goldstein et al. cite an impressive body of data that suggests that the accuracy of verbal description has little or nothing to do with lineup recognition memory performance.

Recall Performance by Film-Audience Witnesses

Considering that it was a tough assignment for any observer to correctly encode and store the amount of information required to make a positive identification, the generally poor recognition performance is not terribly surprising. On average, the film-audience witnesses recalled only 3.26 features correctly (out of a possible 7), and the true positive identifiers were indistinguishable from the false positive and false negative respondents. In general, witnesses tended to underestimate the perpetrator's height (which was 71 in.), reporting a mean height of 69 in., within a range of 60-74 in. The median weight estimate of 155 lbs., within a range of 120-190 lbs., reflects a general tendency to overestimate the perpetrator's actual weight of 145 lbs. Age estimates varied from 17 to 35 years, with a modal age of 25 years, against the perpetrator's actual age of 23 years. If we developed a composite description of the suspect based on the modal descriptions reported by film-audience witnesses, we would be looking for a man of average height and weight, instead of the taller, thinner man who committed the crime.

The test witnesses, like all real eyewitnesses, were asked the race of the perpetrator. Only 66.4% correctly stated "white"; 21.6% said "hispanic," and 11% said "black." Since most witnesses who settled for the white stereotype also picked a lighter skin color, the implication is that, like the concept of central traits in person perception, a stereotyped label can polarize judgments of other physical characteristics. We should note that race is a social conclusion rather than an objective physical description. We found some evidence of different regional stereotypes by comparing the racial mislabeling patterns of New York film audiences and a Washington, D.C., audience. New Yorkers who made a mistaken racial designation called the perpetrator "hispanic," whereas Washingtonians mislabeled him "black." These "mistakes" reflected the then current stereotyped norms for criminals common to New York and Washington, respectively.

It was thus clear that the high rate of attempted identifications in both conditions was probably due more to the inherent demands of being tested for recognition with a lineup than to any correspondence between the witnesses' memory of the culprit's features and the various alternatives in the lineup. These high

rates of attempted identification have also been reported in studies employing a "blank lineup," in which all stand-ins are innocent and any identification is a mistake (Buckhout, Note 4). The results are a sobering reminder that present methods employed by the police to test eyewitnesses cannot be trusted to reliably sort out those witnesses who really remember the culprit from those who are merely guessing.

REFERENCE NOTES

1. Fanselow, M., & Buckhout, R. *Effects of non-verbal biasing in testing eyewitnesses on a photospread test* (Monograph CR-26). Brooklyn: Brooklyn College, Center for Responsive Psychology, 1980.
2. Malpass, R. S., & Devine, P. G. *Eyewitness identification: Lineup instructions and the absence of the offender*. Unpublished manuscript, 1979.
3. *Devlin Report* (Report to the Secretary of State for the Home Department on evidence of identification in criminal cases, House of Commons, April 26, 1976; Chairman: Hon. Lord Devlin). London: Her Majesty's Stationery Office, 1976.
4. Buckhout, R. *Trying too hard: Recognition in spite of poor recall*. Unpublished manuscript, 1980.
5. Goldstein, A. G., Johnson, K. S., & Chance, J. *Face recognition and verbal description of faces from memory*. Paper presented to the Annual Meeting of the Psychonomic Society, Washington, D.C., November 1977.

REFERENCES

- BUCKHOUT, R. Eyewitness testimony. *Scientific American*, 1974, **231** (6), 23-31.
- BUCKHOUT, R., ALPER, A., CHERN, S., SILVERBERG, G., & SLOMOVITS, M. Determinants of eyewitness performance on a lineup. *Bulletin of the Psychonomic Society*, 1974, **4**, 191-192.
- BUCKHOUT, R., FIGUEROA, D., & HOFF, E. Eyewitness identification: Effects of suggestion and bias in identification from photographs. *Bulletin of the Psychonomic Society*, 1975, **6**, 71-74.
- BUCKHOUT, R., & GREENWALD, M. Witness psychology. In R. Imwinkelreid (Ed.), *Scientific and expert evidence in criminal advocacy*. New York: Practising Law Institute, in press.
- ELLISON, K. W., & BUCKHOUT, R. *Psychology and criminal justice: Common grounds*. New York: Harper & Row, 1980.
- HOCHBERG, J., & GALPER, R. Recognition of faces: I. An exploratory study. *Psychonomic Science*, 1967, **9**, 619-620.
- LOFTUS, E. Leading questions and the eyewitness report. *Cognitive Psychology*, 1975, **7**, 560-572.
- LIPTON, J. P. On the psychology of eyewitness testimony. *Journal of Applied Psychology*, 1977, **62**, 90-95.
- MARSHALL, J. *Law and psychology in conflict*. Indianapolis: Bobbs-Merrill, 1966.
- WELLS, G. L. Applied eyewitness testimony research: System variables and estimator variables. *Journal of Personality and Social Psychology*, 1978, **36**, 1546-1557.
- YIN, R. K. Looking at upside-down faces. *Journal of Experimental Psychology*, 1969, **81**, 141-145.