A Field Study on the Relationship Between Quality of Eyewitnesses' Descriptions and Identification Accuracy

Melissa A. Pigott, John C. Brigham, and Robert K. Bothwell.

Melissa A. Pigott is a research associate at Litigation Sciences, Inc., 999 Peachtree St. N.E., Suite 2070, Atlanta, GA 30309. She received a B.A. in psychology from Florida International University and her M.S. and Ph.D. in social psychology from Florida State University. Dr. Pigott was formerly employed as director of marketing research at Baptist Medical Center, Jacksonville, Florida, and adjunct professor of psychology at the University of North Florida. Her research interests include eyewitness identification, jury selection, and psychology and the law.

John C. Brigham is a professor of psychology in the Department of Psychology, Florida State Unviersity, Tallahassee, FL 32306. He has a B.A. in psychology from Duke University and an M.S. and Ph.D. in personality-social psychology from the University of Colorado. Dr. Brigham's interests include factors affecting the accuracy of eyewitness identification, racial attitudes and stereotypes, and general issues in psychology and law.

Robert K. Bothwell is an assistant professor of psychology in the Psychology Department, University of Southwestern Louisiana, P.O. Box 43131, Lafayette, LA 70504-3131. He received his B.A. in psychology from Pan American University and his M.S. and Ph.D. in social psychology from Florida State University. Dr. Bothwell has also held faculty positions at the University of Texas-El Paso and Pan American University. His research interests include factors involved in eyewitness memory such as arousal, hypnosis, race, and confidence.

How reliable is eyewitness memory? Given the conviction rates of juries in cases where the sole incriminating evidence is an identification of the accused (Devlin 1976), one would infer that most jurors place an overwhelming degree of faith in the veracity of eyewitness testimony. Most researchers and many legal scholars have argued that this faith stems primarily from a lack of understanding of basic processes involved in human memory.

In 1972, the U.S. Supreme Court ruled that at least five factors should be considered in determining the reliability of eyewitness testimony (*Neil* v. *Biggers* 1972): (1) opportunity of the witness to observe the criminal when the crime was committed, (2) amount of attention that the witness paid to the event, (3) the accuracy of the description of the criminal (and the crime itself) provided by the witness shortly after the crime, (4) the amount of time between the crime

and the identification, and (5) the witness's degree of confidence in the identification. The Court's criteria were not based on accumulated research findings but rather represented the justices' "educated guesses" about important features.

Since the Court's decision, numerous empirical studies have attempted to examine the relevance of these factors to eyewitness accuracy. Many studies have utilized research paradigms with low mundane realism, which are far distant from crime-like events. Some researchers (for example, Clifford 1978; McCloskey, Egeth, and McKenna 1986) and many courts (for example, *P.B. Johnson and State of Florida* 1983; *United States* v. *Fosher* 1979) have questioned the relevance of such findings.

The current study utilizes a forensically relevant setting to examine the utility of three of the *Neil* v. *Biggers* criteria in predicting the accuracy of eyewit

ness identifications. The variables of interest were opportunity to view, description accuracy, and witness confidence.

A key element in judging an eyewitness's opportunity to view the criminal is the witness's estimate of the duration of the event. Research indicates that people tend to overestimate the duration of an emotionally arousing event (for example, Alper, Buckhout, Chern, Harwood, and Slomovits 1974; Schiffman and Babko 1974). As a consequence, a longer time estimate by a witness may not necessarily represent better opportunity to observe, and hence better chance of identification accuracy, but may instead reflect an erroneous estimate of duration (Loftus, Schooler, Boone, and Kline 1987).

The accuracy of a witness's initial description is difficult to determine in actual cases because the police cannot be sure that the suspect is the perpetrator of the crime, especially when the only incriminating evidence is the identification. Laboratory research suggests that description completeness is a poor predictor of description accuracy (Wells 1985). In addition, although description congruence (the fit between the initial description and the characteristics of the identified suspect) can be empirically assessed, in laboratory studies congruence has not been predictive of identification accuracy (Pigott and Brigham 1985; Wells 1985).

In general, eyewitness confidence is a poor predictor of identification accuracy, especially in "low-optimality" situations where the identification task is difficult (Bothwell, Deffenbacher, and Brigham 1987). Several recent studies have suggested that choosing may be an important moderating variable of the confidence-accuracy relationship (Brigham 1988; Fleet, Brigham, and Bothwell 1987; Krafka and Penrod 1984; Pigott and Brigham 1985). These studies have shown that the relationship between accuracy and confidence in one's decision tends to be stronger among choosers (those who make an identification) than among nonchoosers (those who indicate that the target is not in the lineup).

The present study examined the following questions:

1. Are several measures of the quality of an eyewitness's description related to accuracy of identification?
2. Will witnesses overestimate the duration of a staged event?
3. Is accuracy of time estimates related to identification accuracy?
4. Is the correlation between confidence and identification accuracy stronger among choosers than among nonchoosers?

METHOD

Subjects

Forty-seven female bank tellers were chosen at random as witness/subjects, with the restriction that bank officers had not indicated that the teller was pregnant, highly emotional, or a past victim of a traumatic bank robbery. Twenty-three tellers saw Target A and 24 saw Target B. Most of the tellers (77 percent) indicated that they had attended some type of training on eyewitness techniques.

Instruments

We used a suspect identity chart developed by the Florida attorney general's office. The chart is currently used by law enforcement agencies to obtain suspect descriptions from bank employees after a bank robbery. It lists 16 features concerning a suspect's appearance; a male figure with descriptive nouns corresponding to body parts is pictured.

One six-photograph lineup was developed for each target person. Foil photos were chosen by a member of the Florida State University Police Detective Unit from a large group of head-and-shoulder Polaroid color photos (7.25 x 9.5 cm) of college-age white males. All foils were similar in general appearance to the target persons. Analysis of 50 college-student "mock witnesses'" responses yielded a functional size of 6.00 and effective size of 4.15 for Target A's lineup and a functional size of 1.92 and an effective size of 3.75 for Target B's lineup. (See Wells, Leippe, and Ostrom 1979, and Malpass 1981 for details on the calculation of these lineup measures.)

Staged Event

The target person entered each bank through its main entrance, went to the center island, and pretended to fill out a deposit slip. The target person approached a teller and asked if she would cash a check. He then presented her with a crudely altered United States Postal Service money order ("$10.00" was changed to "$110.00" in blue ink) and attempted to cash it.

When the teller refused to cash the money order (all tellers refused), the confederate stressed that the post office had made a mistake when issuing the money order and had initialed the mistake. After the teller's repeated refusal, the confederate became irate, took the money order, and made a hurried exit out of the bank. The entire confederate-teller interaction lasted approximately 1½ minutes.

Four to five hours after the confederate's visit, a 26-year-old female, dressed in civilian clothes and

posing as a law officer, entered the bank and asked to speak to the person in charge (who was aware the study was taking place). The law officer then asked to question the teller in private. The law officer interrogated the teller about the staged event, recorded descriptions and time estimates, and then showed her a target-present or a target-absent lineup. After the teller had chosen someone or rejected the lineup, confidence in her decision was assessed on a 5-point scale.

All tellers were then informed that they had been participants in a field study on eyewitness evidence. At this time the teller completed a post-experimental questionnaire and was thoroughly debriefed about all aspects of the study.

Scoring of Dependent Measures

Time estimation. Tellers made two estimates of the duration of the customer's visit: a free-response estimate and a mental reconstruction of the interaction while being timed by a stopwatch. This provided four possible duration scores: the length of the estimate, the accuracy of the estimate (the absolute difference between the estimate and the actual 1½-minute duration), the elapsed time for the mental reconstruction of the incident, and the accuracy of the elasped-time measure (the absolute difference between the elapsed time and the actual duration).

Description quality. Twenty-five college student pilot subjects viewed each target person (live) for several minutes and described each target's hair color, hair length, eye color, and complexion. Modal descriptions of each target (along with the target's actual age, height, and weight) were later utilized to assess the correspondence between bank tellers' descriptions and the targets' characteristics.. An additional sample of 44 college students viewed photos of each of the lineup members and rated each member on seven characteristics: age, weight, hair color, hair style, eye color, complexion,, and hair length. Description congruence scores were calculated based on the raters' modal responses.

Description accuracy scores were calculated by comparing the tellers' descriptions of the target with the targets' actual characteristics. Tellers received one point for each description that matched the raters' modal response for that characteristic or was within 2 inches of the target's actual height, 2 years of his actual age, or within 5 pounds of the target's actual weight. A point was deducted for each description that was not accurate and no points were given for characteristics that were not mentioned (*cf.,* Wells

1985). Each teller's total description accuracy score could range from -7 to +7.

Description congruence scores were calculated in a similar manner, except that the teller's description was compared with the modal ratings of the person she chose from the lineup, *whether or not* it was the target. Congruence scores could be calculated only for the 24 tellers who chose someone from the photo lineup.

Completeness scores represented the number of factors the teller listed on the description form, regardless of accuracy. Completeness scores could be as high as 15 since, in addition to the factors listed above, some tellers listed types and color of clothing (shirt, coat, pants, shoes) and a few tellers followed the confederate from the bank and noted the make, color, and license plate of his car.

RESULTS

Identification Accuracy

Tellers correctly identified the target in the target-present lineups slightly less than half the time (47.7 percent), while tellers correctly realized that the customer was not in the target-absent lineups 62.5 percent of the time. Most tellers were moderately confident of their judgment; the mean certainty level was 3.27 on the 5-point scale, as 38 percent of the tellers checked 4 or 5. A 2 (target person) x 2 (type of lineup) multiple contingency analysis performed on the identification accuracy scores yielded no significant main effects or interactions.

TABLE 1
IDENTIFICATION ACCURACY RATES

	Target-present lineup	Target-absent lineup
Target A	38.5%	63.6%
Target B	60.0%	61.5%
Mean accuracy for lineups:	TP: 47.8%	TA: 62.5%
Mean accuracy for target persons:	Target A: 50.0%	Target B: 60.9%
Overall accuracy:	55.3% N=47	

Description Quality

Target A was described as significantly younger, shorter, and heavier than he actually was. Results of t-tests on Target A's age, height, and weight were:

age (t(21)=6.67, p<.005); height (t(21)=7.89, p<.005); and weight (t(21)=3.47, p<.005). Target B was described as significantly older, shorter, and lighter than he actually was; t-test results were: age (t(22)=13.33, p<.005); height (t(22)=9.09, p<.005); and weight (t(22)=2.94, p=.005).

Description congruence scores represent the correspondence between the teller's description and the characteristics of the person chosen from the lineup (whether customer or foil). Description congruence scores, calculated for the 24 tellers who selected someone from the lineup, showed a substantial relationship to description accuracy (r(22)=.67, p<.001). This was expected since 58 percent of the data points were identical (due to the tellers who selected the target, rather than the foil, from the lineup). Completeness scores were not related to description accuracy (r(45)=-.16, ns.) or to description congruence (r(22)=.01, ns.).

Time Estimation Accuracy

In free recall, most tellers overestimated the duration of the interaction. The mean estimate was 4.2 minutes (t(45)=6.34, p<.001); 61.7 percent of the tellers estimated that the (1½ minute) interaction had taken 3 minutes or more. In contrast, mental reconstruction of the event (timed with a stopwatch) produced generally more accurate responses. The overall mean estimate of elapsed time was 67 seconds, which somewhat underestimated the actual 90-second duration of the interaction. The average amount that a teller's elapsed time estimate differed from the actual duration was 39 seconds.

Correlational Data

Description quality and identification accuracy. None of the measures of description quality were significantly related to identification accuracy. The correlation between description accuracy and identification accuracy was not significant (r(45) = -.16, ns.). When only those tellers who picked someone from the lineup ("choosers") were evaluated, there was still no correspondence between description accuracy and identification accuracy (r(22)=.03, ns.). Similarly, there was no significant correspondence between tellers' description congruence scores and the accuracy of their identifications (r(22)=.25, ns.). The final index of teller performance, completeness of description, was also unrelated to identification accuracy (r(45)=.09).

Identification accuracy and confidence. Tellers' confidence in their identification judgments showed a weak relationship to the accuracy of their identications (r(45)=.27, p<.08). However, confidence was

reliably related to accuracy when only the tellers who chose someone from the lineup were analyzed (r(22)=.42, p<.05). For nonchoosers, on the other hand, no relationship between confidence and identification accuracy occurred (r(21=.08, ns.).

Identification accuracy and time estimation. Neither of the time duration accuracy measures related to identification accuracy (r's(45)=-.24 and .13). However, the length of the estimate, irrespective of its accuracy, showed some relation to identification accuracy. Tellers who envisioned a lengthier interaction with the customer tended to be less accurate in their identifications (r(45)=-.32, p<.05). There was also a tendency for tellers who verbally estimated a longer interaction to be less accurate (r)45)=-.24, p<.11) but to give a more complete description (r(45=.33 p<.05).

DISCUSSION

The present data extend the laboratory findings of earlier studies (Cutler, Penrod, and Martens 1987); Pigott and Brigham 1985; Wells 1985) in showing that description congruence, description accuracy, and description completeness are not predictive of identification accuracy in a meaningful field setting. It appears that the Supreme Court was wrong in its emphasis on this "description accuracy" criterion, regardless of whether the Court meant description accuracy or description congruence.

The present study replicates findings of several recent studies regarding the relationship between confidence and identification accuracy. A meta-analysis (Brigham 1988) of six staged-event eyewitness studies found the mean confidence-accuracy correlation for choosers (mean weighted r=.37, N=533) to be significantly higher (z=4.51, p<.001) than the mean confidence-accuracy correlation for nonchoosers (r=.07, N=330). It appears that witnesses' confidence in their positive identifications may carry considerably more information about the likely accuracy of those identifications than do witnesses' confidence in their decision that the target was not in the lineup. This is an important distinction because a chooser is far more likely to testify in court than a nonchooser.

The finding of negative correlations between estimates of exposure duration and identification accuracy is interesting. Exposure duration can be seen as a central component of the "opportunity to view" criterion, and a meta-analysis of facial recognition studies (Shapiro and Penrod 1986) found that exposure durations were, in general, positively correlated with recognition accuracy. However, in most crime situations there is no independent measurement of exposure duration, and law enforce-

ment and the courts must depend on eyewitnesses' subjective estimates of duration. The present data illustrate the danger in the reliance on individuals' estimates. Clearly, this issue is worthy of further empirical investigation as is the finding that the general tendency to overestimate the duration of the initial incident can be eliminated by having eyewitnesses cognitively review the incident.

In sum, this study provides evidence of the independence of identification accuracy and description quality in a forensically relevant field setting. This finding is directly applicable to the legal system which currently relies on description quality as a predictor of eyewitness accuracy. The courts have shown a general disinclination to apply findings of laboratory studies to real-world situations (for example, *Lockhart* v. *McCree* 1986) but a number of courts have stressed that social science research carried out in a forensically meaningful situation can be of considerable value to the courts' decision-making duties (for example, *State of Arizona* v. *Chapple* 1983; *People* v. *McDonald* 1984).

ACKNOWLEDGEMENT

REFERENCES

Alper, A.; Buckout, R.; Chern, S.; Harwood, R.; and Slomovits, M. 1976. Eyewitness identification: Accuracy of individual vs. composite recollections of a crime. *Bulletin of the Psychomic Society* 8:147-149.

Bothwell, R.K.; Deffenbacher, K.; and Brigham, J.C. 1987. Correlation of eyewitness accuracy and confidence: Optimality hypothesis revisited. *J. Appl. Psych.* 72:691-695.

Brigham, J.C. 1988. Is witness confidence helpful in judging eyewitness accuracy? In *Practical aspects of memory: Current research and issues* Vol. 1, pp. 72-82, edited by M.M. Gruneberg, P.E. Morris, and R.N. Sykes. Chichester: Wiley.

Clifford, B.R. 1978. A critique of eyewitness research. *In practical aspects of memory*, pp.199-209, edited by M.M. Gruneberg, P.E. Morris, and R.N. Sykes. London: Academic Press.

Cutler, B.L.; Penrod, S.D.; and Martens, T.K. 1987. The reliability of eyewitness identification: The role of system and estimator variables. *Law and Human Behavior* 11:233-258.

Devlin, Hon. Lord Patrick. 1976. *Report to the Secretary of State for the home department of the departmental committee on evidence of identification in criminal cases.* London: Her Majesty's Stationery Office.

Fleet, M.L.; Brigham, J.C.; and Bothwell, R.K.; 1987. The effects of choosing on the confidence-accuracy relationship in eyewitness identification. *J. Applied Soc. Psych.* 17:171-187.

Krafka, C., and Penrod, S.D. 1985. Reinstatement of context in a field experiment on eyewitness identification. *J. Personality and Social Psych.* 49:58-69.

Lockhart v. *McCree*, 106 S. Ct. 1758 (1986).

Loftus, E.F.; Schooler, J.W.; Boone, S.M.; and Kline, D. 1987. Time went by so slowly: Overestimation of event duration by males and females. *Applied Cognitive Psych.* 1:3-13.

Malpass, R.S. 1981. Effective size and defendant bias in eyewitness identification lineups. *Law and Human Behavior* 5:299-309.

McCloskey, M.; Egeth, H.; and McKenna, J. 1986. The experimental psychologist in court: The ethics of expert testimony. *Law and Human Behavior* 10:1-13.

Neil v. *Biggers*, 409 U.S. 188, 93 S. Ct. 2243, 53 L. Ed. 2d (1972).

P.B. Johnson v. *State of Florida*, 438 So. 2d 774 (1983).

People v. *McDonald*, 37 Cal, 3d 351, 690 P.2d 709, 208 Cal. Rptr. 236 (1984).

Pigott, M.A., and Brigham, J.C. 1985. Relationship between accuracy of prior description and facial recognition. *J. Appl. Psych.* 70(3):547-555.

Schiffman, H.R., and Bobko, D.J. 1974. Effects of stimulus complexity on brief temporal events. *J. Experimental Psych.* 103:156-159.

Shapiro, P.N., and Penrod, S. 1986. A meta-analysis of facial identification studies. *Psychological Bulletin* 100:139-156.

State of Arizona v. *Chapple*, 135 Ariz. 281, 660 P.2d, 1208 (1983).

United States v. *Fosher*, 590 F. 2d 381 (1st Cir., 1979).

Wells, G.L. 1985. Verbal descriptions of faces from memory: Are they diagnostic of identification accuracy? *J. Appl. Psych.* 70:619-626.

Wells, G.L.; Leippe, M.R.; and Ostrom, T.M. 1979. Guidelines for empirically assessing the fairness of a lineup. *Law and Human Behavior* 3:285-293.