Law and Human Behavior, Vol. 11, No. 3, 1987

The Reliability of Eyewitness Identification

The Role of System and Estimator Variables*

Brian L. Cutler,† Steven D. Penrod,† and Todd K. Martens‡

This study examines the effects of 14 estimator variables (e.g., disguise of robber, exposure time, weapon visibility) and system variables (e.g., lineup instructions, exposure to mugshots) on a number of measures of eyewitness performance: identification accuracy, choosing rates, confidence in lineup choice, relation between confidence and identification accuracy, memory for peripheral details, memory for physical characteristics of target, and time estimates. Subjects viewed a videotaped reenactment of an armed robbery and later attempted an identification. Characteristics of the videotape and lineup task were manipulated. Prominent findings were as follows: identification accuracy was affected by both estimator and system variables including disguise of robber, weapon visibility, elaboration instructions, and lineup instructions. Memory for peripheral details was positively correlated with choosing on the identification task but negatively correlated with identification accuracy.

INTRODUCTION

The view that psychologists should provide expert testimony on eyewitness memory in cases in which eyewitness recall or recognition is a primary source of evidence has been advocated by psychologists (Loftus, 1983a; Penrod, Loftus, & Winkler, 1982), lawyers (Frazzini, 1981; Stein, 1981; Woocher, 1977), and judges (Bazelon, 1980; Weinstein, 1981). This view is predicated on the assumption that

* This research was funded by the National Science Foundation under grant SES-8411721 and by the National Institute of Justice under grant No. 84-IJ-CX-0010 to the second author. Carol Krafka and Peter Shapiro were instrumental in the planning of this research. In addition, we wish to thank two anonymous reviewers for their thoughtful comments on an earlier draft of this manuscript, and Mark Bartells, Karla Bishell, James Coward, Michael Hart, Thomas O'Rourke, and Todd Ripple for their assistance with various phases of this research.

† Florida International University.

‡ Harvard University.

0147-7307/87/0900-0233$05.00/0

psychologists have an adequate foundation of empirical literature from which to draw when informing the judge and the jury on psychological factors that influence eyewitness memory. Recently this assumption has been debated in the psychological literature (Bermant, 1986; Konečni & Ebbesen, 1986; Loftus, 1983a, 1983b, 1986; McCloskey & Egeth, 1983a, 1983b; McCloskey, Egeth, & McKenna, 1986; Woocher, 1986; Yarmey, 1986). McCloskey and Egeth (1983a, 1983b; McCloskey et al., 1986) and Konečni and Ebbesen (1986) contend that psychologists should question the empirical literature on which their conclusions and expert testimony are based. This contention, however, leaves the police, the district attorney, the defense attorney, the lay juror, and the judge to their own intuitions as to how eyewitness testimony should be evaluated and weighted in comparison to other forms of evidence. However, survey research (Brigham, 1981; Brigham & Bothwell, 1982; Brigham & Wolfskeil, 1983; Deffenbacher & Loftus, 1982; Rahaim & Brodsky, 1981; Yarmey & Jones, 1983) and experiments involving jury simulations (Cutler, Penrod, & Stuve, in press; Lindsay, Wells, & Rumpel, 1981; Wells, Lindsay, & Ferguson, 1979; Wells, Lindsay, & Tousignant, 1980) clearly demonstrate that lay intuition is often at variance with well-documented experimental findings (see Wells, 1984, for an overview of this research).

The variance between intuition and experimental findings is exemplified in the United States Supreme Court's ruling in *Neil vs. Biggers* (1972). In that case the Court specified variables that jurors may use to evaluate the accuracy of eyewitness identifications. Two of the Supreme Court's criteria, the witness's confidence in his or her identification and the accuracy of the witness' prior description of a suspect, have been carefully scrutinized in the psychological literature (see Wells & Murray, 1983, for an overview). The Supreme Court assumed that the confidence professed by an eyewitness in an identification is directly related to the accuracy of the identification. However, comprehensive reviews of experiments that assess the relation between eyewitness identification accuracy and confidence indicate that this relationship may be quite weak (Deffenbacher, 1980; Deffenbacher, Bothwell, & Brigham, 1986; Leippe, 1980; Wells & Lindsay, 1985; Wells & Murray, 1984). Similarly, the Supreme Court's contention that accuracy of a prior description of a suspect is directly related to the accuracy of a subsequent identification of the suspect receives little support in the eyewitness literature (see, for example, Goldstein, Johnson, & Chance, 1979).

The apparent discrepancy between lay intuitions regarding variables that affect the reliability of eyewitness identifications and recent findings of experiments that test these intuitions raises the question of how to approach the development of an adequate base of knowledge of eyewitness-relevant variables and how to demonstrate or test its adequacy. One possible approach to the problem is suggested by Wells' (1978) distinction between "estimator" and "system" variables. Estimator variables are factors over which the criminal justice system exerts no control. Examples of estimator variables are the level of stress experienced by the witness during the crime; the amount of time the witness had to encode relevant information, such as a perpetrator's face and physical characteristics; and the degree to which the witness was distracted from attending to a perpetrator's characteristics. Wells (1978) terms these estimator variables because, although these variables may be manipulated in the laboratory to influence identification accuracy, they are uncontrollable in the actual criminal situation, and their influence on identification accuracy must be estimated post hoc.

System variables, on the other hand, refer to eyewitness factors that are directly under the control of the criminal justice system. The number of foils in a lineup, the selection of lineup members, context reinstatement procedures, and questioning techniques are examples of system variables. System variables, in contrast to estimator variables, are advantageous in several respects. First, investigators may employ system variables to maximize the reliability of witness reports. Second, the levels of system variables are generally known and do not require post hoc estimation. An investigator typically knows the number of suspects in a lineup, and snapshots are generally taken of the lineup members so fact finders can assess the fairness of the lineup (i.e., the extent to which the foils resembled the physical appearance of the suspect). The levels of estimator variables, on the other hand, must often be determined on the basis of witness testimony; these levels might therefore be subject to error.

Estimator variables quickly fall prey to the criticisms of McCloskey and Egeth (1983a), especially those that have received little experimental attention. Indeed, Wells (1978) concludes that applied eyewitness testimony research should focus primarily on system variables. Among the challenges to the forensic utility of estimator variables are that their levels must be assessed after the fact and objective verification is often impossible; that estimator variables demonstrate inconsistent patterns of results across studies (McCloskey & Egeth, 1983a); and that estimator variables probably interact with one another to affect eyewitness recall and recognition (Wells, 1978).

It may be the case, though, that the pessimism regarding estimator variables is unwarranted. This claim is based on the following arguments pertaining to both estimator and system variables: (1) despite current confusion regarding the effects of various estimator variables, further research may clarify their role in the prediction and postdiction of eyewitness reliability; (2) estimator variables may not generally interact with one another or with system variables; and, if they do interact, the magnitude of the interactions may be negligible; (3) even though the level of estimator variables must be estimated post hoc, information regarding reliable effects of estimator variables is certainly informative, and should therefore lead to a more informed decision than would no information regarding estimator variables; (4) just as jurors are capable of integrating various forms of evidence in their search for facts, they should be capable of integrating information about estimator variables to improve judgments regarding the reliability of a witness; (5) system variables have not convincingly been shown to be more informative than estimator variables; and (6) system variables are subject to important constraints in that new interrogation and lineup procedures, no matter how effective they may be, must be practically and easily integrated into current police procedures.

The six counterarguments to these critiques have a common thread: All are essentially hypotheses and should therefore be subject to empirical investigation. The first argument, that further research may clarify the role of estimator variables, has recently been addressed by Shapiro and Penrod (1986) in a meta-analysis of the facial-identification studies. The meta-analysis is useful for identifying consistencies among a large number of experiments that study similar independent and dependent variables. In Shapiro and Penrod's analysis of 128 experiments (960 experimental conditions, over 16,500 subjects, and over 713,000 separate recognition judgments), some estimator (and system) variables showed reliable and large effects on eyewitness sensitivity (d'). These variables included transformations of the targets' facial features or disguise (less sensitivity when the target's face was transformed in some way, such as the addition or loss of a beard), the amount of time the witness viewed the target (sensitivity increased as exposure time increased), and target distinctiveness (better sensitivity with distinctive targets). The remaining arguments have been given little attention by investigators and clearly merit further investigation. Estimator variables are the primary focus of the present research. The experiment reported here is designed specifically to examine the effects of a variety of estimator variables on eyewitness recall and identification accuracy and to test for interactions between these estimator variables and several system variables.

It is plausible that estimator and system variables affect some forms of eyewitness performance but not others. To examine this possibility we test the effects of estimator and system variables on a variety of eyewitness performance measures. Identification accuracy is a most important dependent variable from forensic perspectives, but it has been argued (cf. Malpass & Devine, 1984) that choosing, or criterion, is equally important. In addition, the importance of the confidence–accuracy relationship and the accuracy of prior descriptions of the target were highlighted in the Supreme Court ruling in *Neil vs. Biggers*.

Eyewitness Confidence

Eyewitness confidence and the relation between confidence and identification accuracy are important from both a forensic perspective, as well as from a metacognitive perspective (Cutler & Penrod, 1986a). Recent reviews of experiments testing the eyewitness confidence–accuracy relation (Deffenbacher, 1980; Deffenbacher et al., 1986; Wells & Murray, 1984) indicate that the correlation is negligible. Wells and Murray found the average correlation to be .08. More recently, Deffenbacher et al. (1986) meta-analyzed 40 experiments in which eyewitness confidence and identification accuracy were assessed. They found the average correlation between confidence and accuracy to be .25 (SD = .09). Most notable, however, is the extreme variability in estimates across studies. Deffenbacher (1980) reports confidence–accuracy correlations ranging from .20 to .95 across experiments reported in 25 separate articles and conference presentations, and Wells and Murray report negative correlations between confidence and accuracy. Given this variability, it seems plausible that some identifiable variables moderate the confidence–accuracy relation. For example, previous studies have found that target distinctiveness (Brigham, 1986) and retrospective self-awareness (Kassin, 1985) moderate the confidence–accuracy relation.

Deffenbacher's (1980) "optimality" hypothesis suggests several candidate moderator variables. The optimality hypothesis holds that variables that affect information processing at encoding also affect the reliability of the confidence estimate as a predictor of identification accuracy. Under conditions that promote information processing (encoding), the confidence–accuracy relation is expected to be strong, whereas under conditions that reduce effective information processing, the confidence–accuracy relation is expected to be weak. Deffenbacher specifically predicts that expectation of a lineup task, long exposure of target, short retention interval, low arousal or stressfulness, and unbiased lineup instructions, all should enhance the reliability of eyewitness confidence estimates and thus increase the confidence–accuracy correlation. The optimality hypothesis has been supported in integrative reviews of the confidence literature (Deffenbacher, 1980; Deffenbacher et al., 1986). Factors associated with the optimality of encoding conditions were manipulated in the present study. Although we have argued elsewhere (Cutler & Penrod, 1986a) that the confidence–accuracy relationship might best be examined using a within-subject design, a between-subject design is used here in order to retain external validity.

Another confidence–accuracy issue is raised by research demonstrating that subjects often display considerable overconfidence in their judgments (See Lichtenstein, Fischhoff, & Phillips, 1982 for an overview). Koriat, Lichtenstein, and Fischhoff (1980) posited that subjects, when rating their certainty, attend only to evidence confirming their judgments while ignoring disconfirming evidence. This bias would account for overconfidence. Koriat et al. found that decision makers who were required to write out reasons for their responses were better calibrated than decision makers who did not write out reasons.

Eyewitness Recall Memory

In *Biggers* the Court contended that the accuracy of a witness's prior description of a target is positively correlated with the witness's ability to correctly identify the target. Wells (1985) interprets the court's contention to mean that the witness's description of the target and the characteristics of the identified target are congruent. In the general cognitive literature it has been shown that recall of information is not necessarily correlated with recognition of the information (Broadbent & Broadbent, 1977; Flexser & Tulving, 1978), and in the realm of facial identification, it has been shown that the relationship between facial description accuracy (on recall trials) and recognition performance is at best weakly correlated (Goldstein, Johnson, & Chance, 1979; Howells, 1938; Pigott & Brigham, 1985; Wells, 1985). Wells (1985) found that faces that are more readily described are also more accurately recognized, but the relationship was not strong. In the present investigation we examine another type of recall: estimates of height and weight of the target. If the Supreme Court's contention is valid, then accuracy of height and weight estimates should be positively correlated with identification accuracy.

Witnesses are generally asked to remember details of an incident, especially when police are searching for a suspect. Memory for the physical characteristics of a target or of any peripheral details, as recalled by an eyewitness, may be assessed for accuracy and may also later be used as predictors of identification accuracy. Wells and Leippe (1981) found that memory for peripheral details negatively correlates with identification accuracy, but this finding has not been replicated. The present study attempts to replicate the Wells and Leippe finding in addition to testing the effects of estimator variables on eyewitness recall of peripheral details.

In summary, we manipulate variables associated with the encoding, storage, and retrieval stages of memory, as suggested by Penrod, Loftus, and Winkler (1982). Variables were chosen on the basis of their performance in the Shapiro and Penrod (1986) meta-analysis and their relevance to forensic situations. The effects of the variables and their interactions on six theoretically and forensically relevant dependent variables are examined: identification accuracy, choosing, confidence in lineup choice, confidence–accuracy relation, memory for peripheral details other than the robber's face, and distortion of time estimates. The general procedure was that subjects watch a videotaped reenactment of a robbery and later attempt to identify the robber from a videotaped lineup. In all, 14 estimator and system variables were examined.

In order to manipulate 14 variables (two levels each) a $2^{(7+7)}$ fractional factorial design (Cochran & Cox, 1957; Kenny, 1985) was employed. Seven variables were fully crossed with one another, while the remaining seven were confounded with higher-order four- and five-way interactions. The cost of this design is that it confounds higher-order interactions, but all main effects and two-way interactions, and most three-way interactions, can be assessed without difficulty because they are confounded only with higher-order (four-, five-, and six-way) interactions. Since one purpose of this experiment is to simultaneously manipulate many characteristics of the eyewitness situation and to test the generalizability of main effects across witnessing conditions, we are more concerned with the main effects and lower-order interactions than with the higher-order, five- and six-way interactions. It was decided to obtain at least two subjects per experimental cell. Many researchers advocate the use of no less than 10 subjects per experimental cell. It should be noted that in the current design, although there are only (at least) two subjects per cell, there are at least 64 subjects for each level of each predictor. Two-way interaction cells have at least 32 subjects. There is no problem with using less than 10 subjects per cell in a fractional factorial design such as this one, if higher-order interactions are not analyzed. In fact, Kenny (1985) advocates the use of designs in which there is only one subject per cell if higher-order interactions are ignored. A description of the 14 variables and prior findings regarding each are described below.

Prestimulus Manipulations

Expectation of Crime and Lineup Task. Before viewing the stimulus robbery, half the subjects were informed that they would be viewing a robbery and subsequently making an identification of the perpetrator from a lineup, while the other half were simply told that they would be viewing some videotaped materials and then answering some questions about them. This manipulation was included to test whether other variables have similar effects on memory for incidentally and intentionally encoded stimuli.

Type of Elaboration. Subjects in the facial elaboration condition were instructed to pay attention to the faces they saw, and to make various personality judgments about the faces (e.g., Does this person seem basically friendly or cold?). Subjects in the nonfacial elaboration condition, on the other hand, were instructed to attend to the objects of the videotape, and to decide whether those objects seemed to be appropriate in the context of the videotaped scenario. Customer service workers (especially bank-tellers) are sometimes instructed by their employers to, in the case of a robbery, pay close attention to the perpetrator's face so they might be able to identify him or her among mugshots or from a lineup test. This manipulation was designed to assess the effectiveness of such an instruction. Identification accuracy was affected by a similar manipulation in one previous experiment (Slack & Penrod, 1982).

Distractor Task. Half of the subjects were instructed to count the number of times the word "the" was used in the stimulus videotape, while the remaining half were given no such instructions. It was expected that this task would lead to less effective encoding of the stimulus robbery, and would thus lead to a poorer recall and recognition accuracy.

Stimulus Variables

Number of Additional Bystanders. Before the robber appeared in the videotape, subjects watched either two or five separate interactions between the victim and other actors (henceforth referred to as distractors). Several investigators (e.g., Wall, 1965) have proposed that increasing the number of distractors in a scene should lead to poorer identification accuracy, while others (e.g., Levine & Tapp, 1973) have suggested that comparative processing of multiple distractors would facilitate recognition performance. Clifford and Hollin (1981) found support for the former view: as the number of perpetrators in a crime simulation increased, the accuracy of witness testimony decreased. The present study attempted to extend Clifford and Hollin's (1981) findings. In the Clifford and Hollin experiment the distractors were involved in the crime, whereas in the current experiment the distractors left the scene before the crime was committed.

Exposure Time. In the present study, the robber was visible (exposed) for either 30 or 75 sec. The amount of time for which the target is exposed has been found to influence recognition performance in a variety of studies. Using both photographic stimuli (Laughery, Alexander, & Lane, 1971; Mueller, Carlomusto, & Goldstein, 1978) and videotaped stimuli (Slack & Penrod, 1982), correct recognition of human faces has been facilitated by increasing target face exposure time.

Violence. In half of the videotaped robberies, the robber threatened and manhandled the clerk, fired his handgun into the floor, and threw the victim to the floor before escaping. In the remaining versions of the robbery, the robber appeared calm and nonthreatening throughout the robbery, and left without ever manhandling the victim. The volume of the exchange in the high-violence robbery was considerably louder than the volume of the exchange in the low-violence robbery. Increased violence in videotaped reenactments of crimes has been shown to lead to decrements in both identification accuracy and eyewitness recall (Clifford & Hollin, 1981; Clifford & Scott, 1978; Johnson & Scott, 1976; Sanders & Warnick, 1980), but this finding is not universally obtained (Sussman & Sugarman, 1972). One goal of the present study was to identify conditions under which violence in a witnessed event impairs performance on recall and recognition tests.

Weapon Visibility. In half of the videotapes, the robber outwardly brandished his weapon during the entire robbery (and pointed it at the victim), while in the remaining half, the weapon was hidden under the robber's coat and thus remained invisible throughout most of the robbery. This manipulation was designed to facilitate "weapon focus" (Loftus, 1979), which refers to an attentional phenomenon wherein the witness's attention is concentrated on a brandished weapon. Since more of the witness's attention is focused on the weapon, there is less attention given to and less encoding of the robber's face or of other relevant details. The weapon focus effect has been given little empirical attention, although there is a theoretical basis for expecting such an effect (Esterbrook, 1959). Poorer recall and recognition of the robber were expected in a high-weapon-visibility group as compared with the low-weapon-visibility group.

Disguise. In half of the videotapes the robber wore a knit pullover hat that fully covered his hair. In the remaining videotapes the robber wore no hat. As noted earlier, a meta-analysis of facial recognition studies (Shapiro & Penrod, 1986), found that transformation of target characteristics between encoding and retrieval situations had large effects on eyewitness performance across 20 studies in which this variable was examined. It was expected that disguising the robber in this manner would lead to less accuracy on the identification test.

Crime. A stimulus film manipulation was employed to examine whether the results obtained with one crime situation would generalize to a different crime situation with a different robber and a different location. While dialogue remained essentially the same in the two robberies, one occurred in a liquor store, while the other occurred on a path near a lake. The victim was the same in both videotapes.

Storage Variables

Exposure to Mugshots. After viewing the videotaped robbery, half the subjects in the present study viewed 44 mugshot slides. The other half viewed no mugshots. Before viewing the slides, subjects were instructed to search the series of mugshots for the perpetrator. Immediately after the mugshot task subjects were informed that the robber had not been among the faces in the mugshot array. Several investigators have found performance decrements attributable to mugshot searches (Davies, Shepherd, & Ellis, 1979; Deffenbacher, Leu, & Brown, 1979; Devlin, 1976).

Retention Interval. Identifications were made either within 30 min after viewing the crime, or after one week. Though retention interval often reduces facial recognition accuracy (Courtois & Mueller, 1981; Cutler, Penrod, O'Rourke, & Martens, 1986, Experiment II; Krouse, 1981; Slack & Penrod, 1982; Brigham, Maass, Snyder, & Spaulding, 1982), some studies fail to find retention interval effects (Krafka & Penrod, 1985; Laughery, Fessler, Lenorovitz, & Yoblik, 1974). It is probably the case that retention interval effects are mediated by other factors (Barkowitz & Brigham, 1982; Cutler, Penrod, O'Rourke, & Martens, 1986).

Retrieval Variables

Lineup Instructions. Half of the subjects were explicitly offered the choice of rejecting the lineup, whereas the remaining subjects were allowed, but not explicitly offered, the choice of rejecting the lineup. Buckhout and associates (Buckhout, 1974; Buckhout, Alper, Chern, Silverberg, & Slomovitz, 1974; Buckhout, Figueroa, & Hoff, 1975) have found that instructions that pressure the witness to make an identification from the lineup (henceforth referred to as "biased instructions") increase the number of false identifications, but do not affect the correct identification rate. Similar findings have been noted by Malpass and Devine (1981) and Warnick and Sanders (1980).

Type of Lineup. Half of the subjects were shown an offender-present lineup, while the remaining subjects were shown an offender-absent lineup. Much of the eyewitness research has been criticized for examining identification accuracy using only offender-present lineups (Malpass & Devine, 1981, 1984; Wells & Lindsay, 1985). Offender-absent, or "blank" lineups, are forensically relevant. The police generally do not know whether the offender is actually in the lineup (although they could intentionally provide the witness with a blank lineup as an additional test); otherwise, there would be no reason to carry out the lineup parade. Identification accuracy involves both identifying the perpetrator when present and rejecting the lineup when the perpetrator is absent from the lineup.

Reasons for Confidence. The final variable manipulated in the present study was implemented after subjects had made their lineup decisions and before they rated the confidence in their lineup decisions. Following the procedures of Koriat et al., subjects in the "reasons for confidence" condition wrote down two reasons in support of their choices and two reasons contradicting their choices. It was hypothesized that subjects who gave reasons for their decisions would show a stronger confidence–accuracy correlation than subjects who were not asked for reasons.

METHOD

Subjects. Subjects ($N = 165$) were volunteers from the University of Wisconsin—Madison Introductory Psychology subject pool, who received extra credit points for their participation. Subjects were randomly assigned to conditions. There were two to three subjects per cell. Each cell necessitated a separate encoding session, but one-week retrieval sessions were run in groups of 8 to 15 subjects.

Instructions. Instructions concerning expectation of a lineup test, elaboration, and distraction (described in detail above) were given in writing. Subjects in the mugshot condition were instructed to study each slide and search for the robber while the experimenter read a number aloud for each mugshot. At the end of the presentation, they were to indicate in writing the corresponding number of the suspect whom they thought was the robber or to indicate that the robber was not among the mugshots (unbiased instructions). Neutral and biased lineup instructions were also given in writing.

Subjects in the reasons for confidence condition were instructed to write down two reasons supporting and two reasons contradicting their lineup decision. Subjects in the no reasons condition completed an innocuous imagery-ability questionnaire.

Questionnaires. The interrogation questionnaire contained 19 questions pertaining to descriptions of the robber (height, weight, clothing, appearance), the event, and the setting. Subjects were also asked questions pertaining to peripheral details. These questions concerned the hand in which the robber held his weapon, the color of the victim's sweater, and the number of people that interacted with the victim before she was accosted. A third category of questions asked subjects to estimate the length of time for which the robber was visible. Finally, manipulation checks were also included. Subjects were asked to indicate whether they paid particular attention to faces or to background objects. Subjects also rated the violence of the robbery on 16 separate bipolar items, such as: the victim was "calm, hysterical," "not very nervous, very nervous," "not threatened, threatened," "uninjured, injured," and the robber was "not violent, violent." Each pair of descriptors was separated by a nine-point scale, with higher numbers indicating greater violence.

Confidence in lineup choice was measured with three separate items. Subjects first rated their confidence in the accuracy of their choices on a nine-point bipolar scale ("not at all confident, very confident") and on the same scale rated their willingness to sign a sworn statement regarding their choices ("not at all willing, very willing"). Subjects then indicated the probability, from .01 to 1.00, that their choices were correct.

Stimulus Materials. Six variables were fully crossed within 64 stimulus tapes. This was accomplished by repeated filming and careful editing of the videotapes so that they reflected appropriate experimental conditions. The plot of the vignette concerned a woman who, after interacting with two or five persons, was confronted and robbed. Each videotape lasted approximately 3 min 50 sec, and the robbery itself lasted approximately 75 sec. The videotape, which was high quality ¾-in. videocassette, was shown on a large projector screen (a 64-in. diagonal Kloss Nova Beam, Model 2).

Lineup Materials. Subjects were shown a videotaped, color, eight-person lineup. Serial presentation was employed; this method has been shown to reduce false identifications without affecting correct identifications (Cutler & Penrod, 1986b; Lindsay & Wells, 1985). All lineup members were presented from a front, three quarter profile, and full profile position, and were shown in closeup (face and shoulders) and full frontal views. Besides the replacement of the offender in the blank lineup condition, the seven foils were identical in all conditions. The position of the offender in the lineups in which he appeared was counterbalanced.

Procedure. Subjects participated in either one or two sessions. Prestimulus instructional manipulations were implemented in the following order: expectation, elaboration, and distractor task. Subjects were then shown the stimulus videotape. Immediately following the videotape, subjects completed the interrogation questionnaire. Subjects who were not in the mugshot exposure condition were either sent to a different room for the lineup test (the immediate recall condition) or were dismissed (the delayed condition). Subjects in the mugshot presentation condition remained to view 44 mugshots, each of which was shown for 2 sec. After subjects viewed the mugshot slides and made a decision regarding whether the robber had been among the mugshots, they were informed that the robber had not been among the slides presented. Subjects then either began the retrieval phase of the study or were dismissed, depending on their retention interval condition.

In the retrieval phase of the study, subjects were given the appropriate lineup instructions and were shown the lineup. Each lineup member was shown for 35 sec and the entire lineup was shown twice before subjects made their decisions. Subjects were then handed either the reasons for confidence instructions or innocuous imagery questionnaire. Subjects then completed the confidence questionnaire.

RESULTS

Manipulation Checks

Subjects who viewed the high-violence stimulus tape gave significantly higher ratings on each of the 16 violence items than subjects who viewed the low-violence videotape; F-values (each on 1 and 162 degrees of freedom) ranged from 26.66 to 247.90, and all were statistically significant ($p < .01$). Subjects clearly found the high-violence stimulus tape more violent than the low-violence stimulus tape. The other manipulation that was checked was elaboration. Of the 165 subjects, 40 claimed to have attended to the background objects, 64 to the faces of the people in the stimulus tape, and 61 reported attending to both. Subjects from the two conditions were equally likely to have reported attending to both stimulus characteristics. Responses of the 104 subjects who reported attending to either faces or background objects were analyzed. Among the 52 subjects assigned to the nonfacial elaboration condition, 30 reported attending to objects while 22 reported attending to faces. Among the 52 subjects assigned to the facial elaboration condition, 10 reported attending to objects while 42 reported attending to faces; this asymmetry in responses was significant, $\chi^2\ (1,102) = 16.25; p < .01$. Although the manipulation was not overwhelmingly effective, it succeeded in diverting the attention of an appreciable proportion of subjects in the planned direction.

Lineup Identification

Identification Accuracy. Of the 165 subjects, 93 were shown offender-absent lineups while 72 were shown offender-present lineups. Among the subjects shown offender-absent lineups, 32% (30) correctly rejected the lineup while 68% (63) falsely identified a foil. Among the subjects shown offender-present lineups, 43% (31) correctly identified the target, 52% (37) falsely identified a foil, and 5% (4) incorrectly rejected the lineup. To examine the influence of the independent variables on overall identification accuracy, correct identifications and correct rejections were scored one and false identifications and misses were scored zero (henceforth referred to as the identification performance score). All independent variables were given orthogonal codes (-1 and 1), and the two-way interactions were represented by cross-products. A hierarchical regression analysis was then performed with identification performance score as the dependent variable. The 13 independent variables were entered on the first step and the 78 pairwise interactions between the 13 variables were entered on the second step. The one variable that was not examined as a predictor of identification accuracy (as a main effect or as an interactant) was whether or not subjects gave reasons for their lineup choice. This variable was not expected to influence performance since it was manipulated after the lineup choice was made.

The results of the regression analysis are displayed in column 1 of Table 1. Of the 13 variables, seven significantly influenced identification accuracy. The only prestimulus variable that significantly affected identification accuracy was elaboration. Facial elaboration instructions improved identification accuracy relative to the nonfacial elaboration instructions. Three stimulus variables significantly influenced performance. Identification accuracy was better under low disguise and low weapon visibility conditions. Performance on the lineup test was better for one target (the robber in the liquor store robbery) than for the other. Among the storage variables, retention interval had a counterinituitive effect on identification accuracy (better performance in the one-week condition than in the immediate condition). Biased lineup instructions significantly reduced the accuracy of identification judgments. Performance on the lineup tests was significantly better in offender-present lineup condition than in offender-absent lineup conditions. In all, these predictors accounted for 37% (adjusted) of the variance in performance on the lineup test.

To further substantiate the effects of biased lineup instructions, we computed the diagnosticity ratio (Malpass & Devine, 1984) on data from the two instruction conditions separately. The diagnosticity ratio is the ratio of correct identifications in offender-present lineups to false identifications in offender-absent lineups. Larger ratios indicate better diagnosticity. In the neutral instruction condition the correct identification rate was .46 and the false identification-rate was .45, thus yielding a diagnosticity ratio of 1.02. The corresponding means for the biased lineup instructions were .43 and .90, respectively, which yields a diagnosticity ratio of .48. Clearly, lineup tests given with biased instructions are less diagnostic than lineup tests given with unbiased instructions.

Among the 78 two-way interactions, only four were significant at the .05

Table 1. Eyewitness Factors and Eyewitness Performance

		Identification accuracy		
Variable	Levels	M_1	M_2	t
Expectation	Low, high	.38	.34	−.52
Distraction	Low, high	.42	.30	−1.73
Elaboration	Nonfacial, facial	.28	.44	2.73[a]
Arousal	Low, high	.38	.34	−.68
Disguise	Low, high	.45	.27	−3.10[b]
Weapon visibility	Low, high	.46	.26	−3.29[b]
Target exposure	30, 75 sec	.41	.31	−1.53
Crime	Mugging, store	.17	.55	6.27[b]
Number of bystanders	Two, five	.38	.34	−.70
Exposure to mugshots	None, viewed	.31	.41	1.58
Retention interval	Zero, seven days	.30	.42	2.04[a]
Lineup instructions	Neutral, biased	.48	.24	−4.06[b]
Offender in lineup	Absent, present	.29	.43	2.30[a]
Reasons for confidence	Not given, given	—	—	—
Identification accuracy	Incorrect, correct	—	—	—
Grand mean			.36	
R			.65	
R^2			.37	
MSe			.15	
DFe			151	
F			8.34[b]	

	Choosing			Confidence			Correlation		
Variable	M_1	M_2	t	M_1	M_2	t	M_1	M_2	t
Expectation	.81	.83	.31	.04	−18	−1.51	.16	−.11	−.78
Distraction	.78	.86	1.24	−.11	.03	.57	.26	−.21	−1.25
Elaboration	.87	.77	−1.89	−.21	.07	1.98[a]	−.13	.17	.79
Arousal	.78	.86	1.54	.05	−.19	−1.70	.02	.02	.01
Disguise	.73	.91	3.33[b]	−.11	−.03	.53	.06	−.01	−.19
Weapon visible	.78	.86	1.54	−.08	−.06	.15	.00	.05	.15
Exposure	.82	.82	.08	−.18	.04	1.59	.10	−.06	−.42
Crime	.88	.76	−2.21[a]	−.12	.02	.62	−.31	.35	1.43
Bystanders	.81	.83	.54	−.12	−.02	.72	−.04	.09	.38
Mugshots	.87	.77	−1.81	−.01	−.13	−.90	−.37	.40	2.34[a]
Retention	.84	.80	−.88	−.01	−.13	−.85	.02	.02	−.01
Instructions	.67	.97	5.70[b]	−.02	−.12	−.72	−.11	.16	.74
Lineup	.69	.95	4.68[b]	−.26	.12	2.59[b]	−.18	.23	1.11
Reasons	—	—	—	−.09	−.05	.25	−.28	.21	1.32
Accuracy	—	—	—	−.34	.20	1.41	—	—	—
Grand mean		.82			−.07			.31	
R		.61			.42			.10	
R2		.31			.09			.01	
MSe		.11			.74			.74	
DFe		151			141			127	
F		6.77[b]			2.06[a]			1.09	

[a] $p < .05$.
[b] $p < .01$.

level, which is about what would be expected by chance alone. These interactions are therefore not discussed. The finding that estimator and system variables did not interact significantly with type of lineup (offender-absent or -present) indicates that the variables affected identification accuracy similarly across lineup conditions; that is, variables that affected the likelihood of a correct identification in the offender-present condition also affected, with equivalent magnitude, the likelihood of a correct rejection in the offender-absent condition. The lack of significant interactions also implies that although the marginal performance levels differed as a function of crime situation, the eyewitness factors had similar effects on identification performance in both situations.

Choosing. All positive identifications, whether correct or incorrect, were scored one, and lineup rejections were scored zero (henceforth referred to as choosing score). Another regression analysis, identical to the one described above, was performed on the data with choosing score as the dependent variable. Results of this analysis are displayed in Table 1. Four variables significantly affected choosing. Subjects were less likely to make a positive identification in the low disguise condition than in the high disguise condition, and more likely to make an identification of the perpetrator of the armed mugging than of the liquor store robbery. Biased lineup instructions led to significantly more choosing than neutral instructions. Finally, and not surprisingly, subjects were more likely to attempt an identification when the offender was present as compared with when he was absent. In all, 31% (adjusted) of the variance in choosing was explained by these predictors. The number of significant two-way interactions did not exceed that which would be expected by chance, and these interactions are therefore not discussed. It is again worth noting that the absence of significant lineup type (offender-absent vs. -present) by estimator or system variable interactions indicates that choosing rate was affected similarly by estimator and system variables across the two types of lineup.

Confidence. Each subject responded to three questions pertaining to their judgments in the lineup task. Subjects' responses to the question pertaining to confidence in lineup choice ranged from 1 to 9 and had a mean of 5.79 (SD = 2.51). Subjects' responses to the question regarding willingness to sign a sworn statement that their choices were correct ranged from 1 to 9 and had a mean of 3.94 (SD = 2.88). Subjects also rated the probability that their judgments were correct. These probabilities ranged from 0 to 1.00 and had a mean of .61 (SD = .31). In order to obtain the most reliable estimate of confidence, responses to the three items were standardized and aggregated to form a confidence scale. The intercorrelations between confidence measures, performance, and choosing are displayed in Table 2. It is noteworthy that the confidence–accuracy correlation was .20.

Moderators of the Confidence–Accuracy Relation. In order to examine the independent variables' effects on the confidence–accuracy correlation (i.e., to test for moderators of the confidence–accuracy relation), a moderated regression analysis was performed. The 15 predictors (the 14 independent variables and identification performance) were entered on Step 1 and the interaction (cross-

Table 2. Intercorrelations among Variables[a]

	2	3	4	5	6	7	8	9
1. Confidence in choice	.72	.85	.95	.18	.01	.02	−.10	.08
2. Willingness to sign		.62	.86	.24	−.09	.01	−.02	.00
3. Probability of correct			.91	.12	.07	.03	−.02	.08
4. Aggregate confidence score				.20	.00	.02	−.05	.06
5. Identification accuracy					−.50	−.21	−.04	.16
6. Choosing						.22	−.03	.08
7. Peripheral details							.03	.02
8. Size estimation								.11
9. Time estimation								

[a] Note. $N = 147$. Correlation coefficients .16 or above are significant at the .05 level.

product) between each variable and identification performance was entered in Step 2. The dependent variable was the aggregate confidence measure. Confidence–accuracy correlations for each subgroup were calculated from the B-values obtained from the regression output; these coefficients are displayed in the fourth column of Table 1. A significant interaction term indicates heterogeneity of regression, which means that the correlation between identification performance and confidence is significantly stronger at one level of the independent variable than at the other. Only one variable (exposure to mugshots) moderated the confidence–accuracy relation. The correlation between confidence and accuracy was .05 for subjects who were not exposed to mugshots and .38 ($p < .01$) for subjects who were exposed to mugshots.

Memory for Peripheral Details

Subjects' recall of details of the videotape were analyzed. Three questions required subjects to report the number of people (distractors) with whom the victim interaction before being robbed (two or five, depending on the condition), the color of the victim's sweater (yellow), and the hand in which the robber held his weapon (right for both videotapes). Correct responses were scored one and incorrect responses were scored zero. Scores for the three items were summed to form a peripheral detail score. Forty-four percent of the subjects accurately identified the number of individuals who interacted with the victim before the robbery occurred, 79% correctly identified the color of the victim's sweater, and 92% correctly identified the hand in which the robber held his weapon. The peripheral detail score ranged from zero to three and had a mean of 2.15 and a standard deviation of .73. This score was used as the dependent variable in a hierarchical regression equation in which nine variables (the three prestimulus and the six stimulus variables) and the subsequent 36 two-way interactions served as predictors. (The storage and retrieval manipulations were instituted after the interrogation questionnaire was completed and were therefore not expected to influence memory for peripheral details.)

Subjects were more accurate at identifying peripheral details in the two-distractor condition than in the five-distractor condition, $t = -10.15, p < .01$; this is understandable, as one of the questions deals specifically with the number of distractors. Subjects were less able to keep track of the number of distractors when the number exceeded two. Subjects reported peripheral details more accurately if they viewed the outdoor mugging than if they viewed the liquor store robbery, $t = -3.15, p < .01$. Finally, there was a marginal effect for level of disguise, $t = 1.87, p < .06$, with subjects accurately reporting more details in the high-disguise condition than in the low-disguise condition. In all, 44% (adjusted) of the variance in memory for peripheral details was accounted for by these predictors. Only four of the 36 interactions were statistically significant, but the interactions were difficult to interpret.

Perception of Time

Subjects' estimates of the robber exposure time were regressed over the nine prestimulus and stimulus variables (entered on Step 1) and the 36 two-way interactions (entered on Step 2). Two main effects were significant. Subjects in the short-exposure condition estimated less viewing time as compared with subjects in the long-exposure condition ($p < .01$). Although this finding may be viewed merely as a manipulation check, it should be noted that, on the average, subjects overestimated the viewing time in both conditions. The robber was exposed for 30 sec in the short-exposure condition, yet average estimated viewing time was 82.21 sec. Likewise, the target was visible for 75 sec in the long-exposure condition, yet the average estimated viewing time was 105.57 sec. In the short-exposure condition, subjects overestimated the viewing time by 174%, whereas in the long-exposure condition, subjects overestimated the viewing time by only 41%. The violence level of the robbery also affected time estimates. Subjects in the low-violence condition reported more viewing time than subjects in the high-violence condition ($p < .01$). Ten percent of the adjusted variance in time estimates were accounted for by the main effects mentioned above. Seven of the 36 two-way interactions were significant at the .05 level. These interactions were difficult to interpret, and are therefore not discussed.[1]

Distortions of Height and Weight Estimates

Subjects estimated the height and weight of the robber. The deviation between height estimate and actual heights (5 ft 10.5 in. for the robber of the liquor store and 6 ft 2 in. for the perpetrator of the armed mugging) were computed and these deviations were squared. Squared weight deviations were also computed (the actual weights of the liquor store and mugger were 147 and 175 lb, respectively). The two deviation scores were then standardized and summed to form a

---

[1] Correspondence concerning these interactions or other analyses can be directed to the first author.

"distortion in size estimate" score. This score was then regressed over the nine prestimulus and stimulus variables and corresponding 36 two-way interactions. Target exposure was the only variable that had a significant effect on size estimates. Size estimates were more distorted when the robbers were exposed for 30 sec as compared with when the robber was exposed for 75 sec ($t = -2.42$, $p < .01$). The number of two-way interactions that were significant did not exceed that which would be expected by chance.

Intercorrelations among Dependent Measures

Pearson product–moment correlations were computed between all dependent measures. List-wise deletion of data was used. These correlations appear in Table 2. That choosing was so highly negatively correlated with identification accuracy ($r = -.50$; $p < .001$) can be attributed to the fact that subjects used very low criteria and half of the subjects viewed an offender-absent lineup. Contrary to the findings of Malpass and Devine (1981), who report a correlation between confidence and choosing of .86, confidence was unrelated to choosing ($r = .00$). It should be noted that other researchers (e.g., Hosch, Leippe, Marchiani, & Cooper, 1984) have reported a negative correlation between confidence and choosing.

An interesting pattern of results emerged within the intercorrelations of choosing, identification accuracy, and memory for peripheral details. Memory for peripheral details was significantly correlated with choosing ($r = .22$; $p < .01$); that is, subjects who remembered more peripheral details were more likely to make a positive identification than subjects who remembered fewer peripheral details. Given the strong negative correlation between choosing and identification accuracy, it is not surprising that memory for peripheral details also yielded a significant negative correlation with identification accuracy ($r = -.21$; $p < .01$). Distortions in size estimates were found to be unrelated to all other variables; all correlations were less than .05 in magnitude.

These correlations were also computed on data from offender-present and offender-absent lineup conditions separately. Results indicate that in data from the offender-present lineup condition, choosing was not significantly correlated with identification accuracy ($r = .12$) or with confidence ($r = .08$). In data from the offender-absent lineup condition, choosing was not significantly correlated with confidence ($r = -.12$), but, of course, choosing was perfectly inversely correlated with identification accuracy. With these exceptions, all other relations were similar across lineup conditions.

DISCUSSION

McCloskey and Egeth (1983a) claim that the literature on which psychologists rely when giving expert advice is fraught with inconsistent findings and unreliable effects, especially with respect to estimator variables. With use of a fractional factorial design we examined a large number of estimator variables and some system variables as predictors of eyewitness identification accuracy, choosing, confidence in identification, the relation between confidence and identification accuracy, memory for peripheral details, time estimates, and accuracy of physical description. This detailed analysis allowed us to evaluate two of the Supreme Court's (*Neil vs. Biggers,* 1972) criteria for evaluating eyewitness identification accuracy.

Identification accuracy was affected by both estimator and system variables. The finding that type of elaboration affected identification accuracy supports previous findings using levels of processing and similar types of manipulations (Craik & Lockhart, 1972; Mueller, Carlomusto, & Goldstein, 1978; Slack & Penrod, 1982). Consistent with the results of Shapiro and Penrod's (1986) meta-analysis, disguise impaired identification accuracy while expectation of the lineup test did not affect identification accuracy. Duration of exposure to the target did not influence identification accuracy. In this experiment we examined two levels of exposure, 30 and 75 sec. If witnesses' attention is directed to a scene containing only a few actors than perhaps 30 sec is sufficient time to encode the target characteristics, for performance can be high even with brief viewing. For example, Cutler and Penrod (1986a) found that when subjects viewed each of 60 faces for only 10 sec, they were, on the average, 85% correct in a recognition task two days later.

Previous research (Clifford & Hollin, 1981; Loftus, 1976) found that identification accuracy was impaired when other people, or distractors from the crime simulation, were included in the lineup. The current experiment examined whether this finding extends to situations in which others were present at the scene but were not in the lineup. The number of distractors did not significantly affect identification accuracy. Perhaps the results would have differed if the distractors were in view while the robber was present or if the distractors appeared in the lineup together with the robber.

Violence level also failed to significantly affect identification accuracy, despite the fact that our manipulation checks showed that one version of the crime was clearly more violent than the other. These results fail to support the findings of Johnson and Scott (1976) and Clifford and Scott (1978). Perhaps the violence manipulation, though appreciable, was not powerful enough to affect encoding processes. Witnesses might well recognize that the scenes they viewed were arousing and exciting even though they, themselves, were not any more aroused than they would be if watching a televised crime drama.

Our findings regarding retention interval are counterintuitive. Subjects performed better on the lineup task after one week than if they had attempted an identification immediately. Deffenbacher, Carr, and Leu (1981) also found better facial recognition after two weeks than after 2 min and refer to this finding as a reminiscence effect. Perhaps in the immediate identification condition subjects have a very clear image of the target, and if the view of the target in the lineup does not precisely match the view of the target in the robbery, subjects either miss or make a false identification. Further research is clearly needed to clarify the effects of very short retention intervals on identification accuracy.

In contrast to the results of previous experiments (e.g., Davies et al., 1979; Deffenbacher, Leu, & Brown, 1979; Devlin, 1976), exposure to mugshots also had little effect on identification accuracy. In the present experiment, half of the subjects searched a series of 41 mugshot slides for the robber. This procedure differed from others in that subjects in the present study were, upon completing the task, informed that the robber was not among the mugshots. This finding supports Davies et al.'s hypothesis that the performance decrement attributable to mugshot exposure obtained in their experiment might be eliminated by informing subjects immediately after the mugshots procedure that the target was not among the mugshots. Perhaps if subjects were told much later, or were not told at all, a decrement in performance would have been observed among subjects who viewed mugshots. In a forensic setting it is unlikely that the police would know whether the mugshots include a suspect; thus, exposure to mugshots without immediate feedback is probably a more forensically relevant manipulation. One implication of this finding is that perhaps police investigators should inform witnesses if they mistakenly identify a known innocent from mugshots.

It is noteworthy that the main effects on identification accuracy were not qualified by two-way interactions, as suggested by Wells (1978). Of the 78 two-way interactions between the 13 variables, only four were significant; and, this is roughly the number of interactions that may be expected to be significant by chance when Type I error is set at .05. Although these results do not rule out large numbers of possible interactions among variables not examined here, the lack of two-way interactions indicates that an emphasis on the main effects of estimators is justified. In fact, an appreciable proportion of variance (37%, adjusted) in identification accuracy was explained by this set of main effects. It is also interesting that the presence or absence of the target in the lineup did not interact with other predictors. This finding indicates that the estimator and system variables studied here had similar effects on both types of lineups.

Four variables significantly affected choosing of a lineup suspect. Subjects were more likely to choose in the high-disguise condition, in the armed mugging condition, if they viewed an offender-present lineup, and if they were given instructions that encourage choosing. Our instruction finding replicates those of Malpass and Devine (1981) as well as those of Buckhout and his colleagues (e.g., Buckhout et al., 1974). In our experiment overall identification accuracy and diagnosticity were adversely affected by biased lineup instructions.

The current experiment found that, like other experiments on eyewitness identification (e.g., Malpass & Devine, 1981), choosing rate was quite high (82% identified a lineup member as the robber). This high choosing rate may reflect a generally high a priori belief that the lineup contains the perpetrator or may indicate a judgmental strategy in which witnesses identify the lineup member who best resembles the perpetrator (Lindsay & Wells, 1985). In either case, a lineup rejection should occur only if the witness has a relatively vivid image of the perpetrator and no close match is found between the lineup members and the witness's image of the perpetrator. Consistent with this conjecture, it was found that factors that degrade memory increase the likelihood that a lineup member is identified (when controlling for the presence of the offender in the lineup). Subjects made more positive identifications under conditions of high disguise and high weapon visibility, and if they viewed the path robbery. There was also a trend toward increased positive identification rate if nonfacial elaboration instructions were given.

An analysis of the variables that affected level of confidence provides some insights into the metacognitive processes associated with eyewitness identification. Subjects were somewhat sensitive to the variables that affected their identification accuracy. The finding that elaboration instructions influenced confidence might be explained in two ways. This result could be another manifestation of subjects' sensitivity to their memories (subjects in the nonfacial elaboration condition had poorer memories of the target than subjects in the facial elaboration condition), or it might indicate that subjects base their confidence ratings in part on the conditions under which the crime was witnessed. In other words, subjects in the facial elaboration condition may have gauged their confidence in part on the depth of processing strategy employed, rather than completely on their belief in their memories.

Further research comparing the effects of variables on both confidence and identification accuracy should yield some useful findings regarding witness' strategies for rating confidence. It is evident from our research that subjects were sensitive to some of the variables that affected their identification accuracy, such as elaboration instructions, but were relatively insensitive to a number of other variables that also affected their identification accuracy, such as disguise, weapon visibility, and biased lineup instructions. Such sensitivity contributes to the overall confidence–accuracy correlation.

The only variable that moderated the confidence–accuracy relation was exposure to mugshots. Subjects who viewed mugshots and were instructed to search for the target were better able to predict their performance on the identification task with their confidence than were subjects who did not view mugshots. Some caution must be exercised in drawing conclusions from this finding given that the mugshot search manipulation studied here was unlike realistic procedures in several important respects. Indeed it is conceivable that the feedback given to subjects immediately after the mugshots procedure (which would normally be unavailable) might have improved the confidence–accuracy correlation.

Deffenbacher's (1980) optimality hypothesis is relevant here in that some variables affected identification accuracy presumably by enhancing or diminishing the quality of information-processing conditions, but none significantly affected the confidence–accuracy relation. However, apart from the one significant moderator effect, little can be concluded regarding the null effects of other variables on the confidence–accuracy relation, since, in general, the confidence–accuracy correlation in this study was so weak.

The one variable specifically intended to improve the confidence–accuracy relation—having subjects list reasons both for and against their decisions—did not, in fact, help to increase the predictive validity of confidence ratings. This finding suggests the possibility that the cause of the low confidence–accuracy correlation is the subjects' inabilities to discriminate correct from incorrect judgments by differentially assigning confidence levels to the two types of judgments, and not to differential calibration of confidence estimates. Further research is needed to substantiate this claim.

Memory for peripheral details was best with conditions that degraded recognition accuracy (high disguise, path robbery). Perhaps this result reflects the subjects' desire to encode information (especially if they know their memories will be tested); but if subjects are prevented from encoding the target's physical characteristics (e.g., by a disguise), they attend to peripheral information instead. This conjecture would also explain the inverse relationship between memory for peripheral details and identification accuracy obtained in the present experiment and in Wells and Leippe (1981).

Little is known about eyewitnessing variables that affect time estimates, although it is generally contended that people tend to overestimate when estimating duration. Subjects in this experiment overestimated the time for which the robber was in view. Interestingly, the extent of overestimation was inversely related to the actual exposure duration. Exposure duration exemplifies the difficulty with estimator variables alluded to by Wells (1978). A witness's estimate of the time for which a perpetrator was visible is apparently subject to systematic distortions. Since exposure duration is an important predictor of identification accuracy (Shapiro & Penrod, 1986) and of the accuracy of height and weight descriptions (as evidenced in the current experiment), then objective assessments of exposure duration are most desirable, but also most difficult to obtain. Nevertheless, the factfinders can be informed about systematic distortions such as time overestimation, and could weigh this evidence accordingly.

A final set of analyses examined the intercorrelations among the dependent measures. Several findings merit additional research attention. First, it has been posited (Malpass & Devine, 1981) that confidence level drives the decision to identify a lineup member or to reject the lineup. Our results fail to support this finding; the correlation between choosing and confidence was zero in both the offender-present and offender-absent lineup conditions. An experiment by Hosch, Leippe, Marchiani, and Cooper (1984) also failed to show a positive correlation between choosing and confidence. One possible explanation for this discrepancy is that perhaps choosing may be unrelated to confidence. Another possible explanation is that studies that employ offender-absent lineups (including both ours and Malpass and Devine's, 1981), fail to determine the basis of lineup rejections. A lineup might be rejected because the witness feels that his or her memory is too weak to make an identification, or a lineup can be rejected because there is a firm belief that the perpetrator is absent from the lineup. Perhaps an attempt to distinguish between these rejections would clarify the confidence–choosing relationship.

Memory for peripheral details was significantly correlated with choosing in these data. As previously mentioned, Wells and Leippe (1981) showed that memory for peripheral details was negatively correlated with identification accuracy. An inspection of Wells and Leippe's results indicates that they, too, found a positive relationship between memory for peripheral details and choosing. The average number of correct items (out of 11) was 6.36 for subjects who made a correct identification, 8.50 for subjects who made an incorrect identification, and 5.11 for subjects who made no identification. In both our study and Wells and Leippe's study, memory for peripheral details was found to be significantly negatively correlated with identification accuracy. This suggests the possibility that subjects base their decision to identify a lineup member as the perpetrator on their ability to recall information about the crime. This reasoning is especially troublesome given Wells and Leippe's (1981; Experiment 2) finding that mock jurors discredited witness' identifications if the witnesses were found to have incorrectly recalled peripheral details. The relationship between peripheral detail recall and identification accuracy indicates that jurors should do just the opposite! Both memory for peripheral details and choosing are important variables from a forensic perspective. As Malpass and Devine (1981) point out, considerable attention should be paid to the variables that mediate choosing. Based on our results, ability to recall peripheral details is a strong candidate for additional study.

One further finding deserves special attention, as its relevance to the *Neil vs. Biggers* (1972) criteria for distinguishing accurate from inaccurate witnesses is evident. The Court held that accuracy (or, more appropriately, congruence; Wells, 1985) of the witness's prior description of the target may be used to predict the likelihood of a correct identification. The correlation between distortion of size (height and weight) estimates and identification accuracy was marginally significant in the direction opposite to that predicted by the Court. Subjects who distorted the size of the robber were more likely to make a correct identification than subjects who gave more accurate height and weight estimates. This finding is consistent with those of other investigators (Goldstein, Johnson, & Chance, 1979; Pigott & Brigham, 1985; Wells, 1985), who found that accuracy of prior facial description was at best weakly related to identification accuracy, and with findings in the general cognitive literature that recall and recognition abilities are uncorrelated (Broadbent & Broadbent, 1977; Flexser & Tulving, 1978).

In conclusion, this experiment helps to reinstate the role of estimator variables in the study of eyewitness testimony. McCloskey and Egeth's (1983a) claim that prior research on variables that affect eyewitness recall and recognition fails to reveal consistent patterns of findings about which to offer expert testimony can be weakened by additional, careful research. The legal system is in need of valid criteria for evaluating the reliability of eyewitness identifications, especially given evidence from this study and others (see Wells & Murray, 1983) concerning the *Neil vs. Biggers* criterion. Although we agree with Wells (1978) that system variables are in need of study, we feel that estimator variables also merit a role in the evaluation of eyewitness evidence.

We have shed some light on the metacognitive processes associated with eyewitness behavior by examining the interrelations between confidence, choosing, memory for peripheral details, height and weight estimates, and identification accuracy. Though we have presented a considerable number of experimental findings from the current experiment, we have also identified a substantial number of productive avenues that eyewitness research can follow.

REFERENCES

Barkowitz, P., & Brigham, J. C. (1982). Recognition of faces: Own-race bias, incentive, and time delay. *Journal of Applied Social Psychology, 12,* 255–268.

Bazelon, D. L. (1980). Review of "Eyewitness Testimony" by Elizabeth Loftus and "The Psychology of Eyewitness Testimony" by A. Daniel Yarmey. *Psychology Today, 13,* 102–106.

Bermant, G. (1986). Two conjectures about the issue of expert testimony. *Law and Human Behavior, 10,* 97–100.

Bothwell, R. K., Deffenbacher, K. A., & Brigham, J. C. (in press). Correlation of confidence and accuracy: The optimality hypothesis revised. *Journal of Applied Psychology.*

Brigham, J. C. (1981). The accuracy of eyewitness evidence: How do attorneys see it? *The Florida Bar Journal, 55,* 714–721.

Brigham, J. C. (1986). *Distinctiveness of appearance as a moderator variable in the confidence–accuracy relationship in facial identifications.* Unpublished manuscript, Florida State University.

Brigham, J. C., & Bothwell, R. K. (1982). The ability of prospective jurors to estimate the accuracy of eyewitness identifications. *Law and Human Behavior, 1,* 19–30.

Brigham, J. C., & Wolfskeil, M. P. (1983). Opinions of attorneys and law enforcement personnel on the accuracy of eyewitness identifications. *Law and Human Behavior, 9,* 337–349.

Brigham, J. C., Maass, A., Snyder, L. D., & Spaulding, K. (1982). Accuracy of eyewitness identifications in a field setting. *Journal of Personality and Social Psychology, 42,* 673–680.

Broadbent, D. E., & Broadbent, M. H. P. (1977). Effects of recognition on subsequent recall: Comments on "Determinants of recognition and recall: Accessibility and generation" by Rabinowitz, Mandler, and Patterson. *Journal of Experimental Psychology: General, 85,* 330–335.

Buckhout, R. (1974). Eyewitness testimony. *Scientific American, 231,* 23–31.

Buckhout, R., Alper, A., Chern, S., Silverberg, G., & Slomovits, M. (1974). Determinants of eyewitness performance on a lineup. *Bulletin of the Psychonomic Society, 4,* 191–192.

Buckhout, R., Figueroa, D., & Hoff, E. (1975). Eyewitness identification: Effects of suggestion and bias in identification from photographs. *Bulletin of the Psychonomic Society, 6,* 71–74.

Clifford, B. R., & Hollin, C. (1981). Effects of the type of incident and the number of perpetrators on eyewitness memory. *Journal of Applied Psychology, 66,* 364–370.

Clifford, B. R., & Scott, J. (1978). Individual and situational findings in eyewitness testimony. *Journal of Applied Psychology, 63,* 352–359.

Cochran, W. G., & Cox, G. M. (1957). *Experimental designs* (2nd ed.). New York: Wiley.

Courtois, M. R., & Mueller, J. H. (1981). Target and distractor typicality in facial recognition. *Journal of Applied Psychology, 66,* 639–645.

Cutler, B. L., & Penrod, S. D. (1986a). *Calibrating confidence and accuracy: Recognizing faces.* Manuscript submitted for publication.

Cutler, B. L., & Penrod, S. D. (1986b). *Improving the reliability of eyewitness identifications: Lineup construction and presentation.* Manuscript submitted for publication.

Cutler, B. L., Penrod, S. D., O'Rourke, T. E., & Martens, T. K. (1986). Unconfounding the effects of context cues on eyewitness identification accuracy. *Social Behavior, 1,* 113–134.

Cutler, B. L., Penrod, S. D., & Stuve, T. E. (in press). Jury decisionmaking in eyewitness identification cases. *Law and Human Behavior.*

Davies, G., Shepherd, J., & Ellis, H. (1979). Effects of interpolated mugshot exposure on accuracy of eyewitness identification. *Journal of Applied Psychology, 64,* 232–237.

Deffenbacher, K. A. (1980). Eyewitness accuracy and confidence: Can we infer anything about their relationship? *Law and Human Behavior, 4,* 243–260.

Deffenbacher, K. A., & Loftus, E. F. (1982). Do jurors share a common understanding concerning eyewitness behavior? *Law and Human Behavior, 6,* 15–30.

Deffenbacher, K. A., Leu, J. R., & Brown, E. L. (1979). *Remembering faces and their immediate context.* Paper presented at the meeting of the Psychonomic Society, Phoenix.

Deffenbacher, K. A., Carr, T. H., & Leu, J. R. (1981). Memory for words, pictures, and faces: Retroactive interference, forgetting, and reminiscence. *Journal of Experimental Psychology: Human Learning and Memory, 7,* 299–305.

Devlin, Rt. Hon. Lord Patrick (chair). (1976). *Report to the secretary of state for the house department of the departmental committee on evidence of identification in criminal cases*. London: Stationery Office.

Easterbrook, J. A. (1959). The effect of emotion on the utilization and organization of behavior. *Psychological Review, 66,* 183–201.

Flexser, A. J., & Tulving, E. (1978). Retrieval independence in recognition and recall. *Psychological Review, 85,* 153–171.

Frazzini, S. F. (1981). Review of eyewitness testimony. *The Yale Review, 70,* XVIII–XX.

Goldstein, A. G., Johnson, K. S., & Chance, J. (1979). Does fluency of face description imply superior face recognition? *Bulletin of the Psychonomic Society, 13,* 15–18.

Hosch, H. M., Leippe, M. R., Marchioni, P. M., & Cooper, D. S. (1984). Victimization, self-monitoring, and eyewitness identification. *Journal of Applied Psychology, 69,* 280–288.

Howells, T. H. (1938). A study of ability to recognize faces. *Journal of Abnormal Social Psychology, 33,* 124–127.

Johnson, C., & Scott, B. (1976). *Eyewitness testimony and suspect identification as a function of arousal, sex of witness, and scheduling of interrogation*. Paper presented at the meeting of the American Psychological Association, Washington, D.C.

Kassin, S. M. (1985). Eyewitness identification: Retrospective self-awareness and the accuracy–confidence correlation. *Journal of Personality and Social Psychology, 49,* 878–893.

Kenny, D. A. (1985). Quantitative methods for social psychology. In G. Lindzey and E. Aronson (Eds.), *The handbook of social psychology* (3rd Ed., Vol. 1, pp. 487–508). New York: Random House.

Konečni, V. J., & Ebbesen, E. B. (1986). Courtroom testimony by psychologists on eyewitness identification issues: Critical notes and reflections. *Law and Human Behavior, 10,* 117–126.

Koriat, A., Lichtenstein, S., & Fischhoff, B. (1980). Reasons for confidence. *Journal of Experimental Psychology: Human Learning and Memory, 6,* 107–118.

Krafka, C., & Penrod, S. D. (1985). Reinstatement of context in a field experiment of eyewitness identification. *Journal of Personality and Social Psychology, 49,* 58–69.

Krouse, F. L. (1981). Effects of pose, pose change, and delay on face recognition performance. *Journal of Applied Psychology, 66,* 651–654.

Laughery, K. R., Alexander, J. F., & Lane, A. B. (1971). Recognition of human faces: Effects of target exposure time, target position, and type of photograph. *Journal of Applied Psychology, 59,* 490–496.

Laughery, K. R., Fessler, P. K., Lenorovitz, D. R., & Yoblick, D. A. (1974). Time delay and similarity effects in facial recognition. *Journal of Applied Psychology, 59,* 490–496.

Leippe, M. R. (1980). Effects of integrative memorial and cognitive processes on the correspondence of eyewitness accuracy and confidence. *Law and Human Behavior, 4,* 261–273.

Levine, F. J., & Tapp, J. L. (1973). The psychology of criminal identification: The gap from Wade to Kirby. *University of Pennsylvania Law Review, 121,* 1079–1131.

Lichtenstein, S., Fischhoff, B., & Phillips, L. D. (1982). Calibration of probabilities: State of the art to 1980. In D. Kahneman, P. Slovic, and A. Tversky (Eds.), *Judgment under uncertainty: Heuristics and biases*. New York: Cambridge University Press.

Lindsay, R. C. L., & Wells, G. L. (1985). Improving eyewitness identifications from lineups: Simultaneous versus sequential lineup presentation. *Journal of Applied Psychology, 70,* 556–564.

Lindsay, R. C. L., Wells, G. L., & Rumpel, C. M. (1981). Can people detect eyewitness identification accuracy within and across situations? *Journal of Applied Psychology, 66,* 79–89.

Loftus, E. F. (1976). Unconscious transference in eyewitness identification. *Law and Psychology Review, 2,* 93–98.

Loftus, E. F. (1979). *Eyewitness testimony*. Cambridge, Massachusetts: Harvard University Press.

Loftus, E. F. (1983a). Silence is not golden. *American Psychologist, 38,* 564–572.

Loftus, E. F. (1983b). Whose shadow is crooked? *American Psychologist, 38,* 576–577.

Loftus, E. F. (1986). Experimental psychologist as advocate or impartial educator. *Law and Human Behavior, 10,* 63–78.

Malpass, R. S., & Devine, P. G. (1981). Eyewitness identification: Lineup instructions and the absence of the offender. *Journal of Applied Psychology, 66,* 482–489.

Malpass, R. S., & Devine, P. G. (1984). Research on suggestion in lineups and photospreads. In G. L. Wells & E. F. Loftus (Eds.), *Eyewitness testimony: Psychological perspectives*. New York: Cambridge University Press.

McCloskey, M., & Egeth, H. (1983a). Eyewitness identification: What can a psychologist tell a jury? *American Psychologist, 38,* 550–563.

McCloskey, M., & Egeth, H. (1983b). A time to speak, or a time to keep silence? *American Psychologist, 38,* 573–575.

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

Mueller, J. H., Carlomusto, M., & Goldstein, A. G. (1978). Orienting task and study time in facial recognition. *Bulletin of the Psychonomic Society, 11,* 313–316.

Neil vs. Biggers, 409 U.S. 188, 93 S. Ct. 375; 34 L. Ed. 2d 401 (1972).

Penrod, S., Loftus, E. F., Winkler, J. D. (1982). Eyewitness reliability. In R. Bray, & N. Kerr (Eds.), *The psychology of the courtroom*. New York: Academic.

Pigott, M. A., & Brigham, J. C. (1985). The relationship between accuracy of prior description and facial recognition. *Journal of Applied Psychology, 70,* 547–555.

Rahaim, G. L., and Brodsky, S. L. (1981). *Empirical evidence versus common sense: Juror and lawyer knowledge of eyewitness accuracy*. Unpublished manuscript, University of Alabama.

Sanders, G. S., & Warnick, D. (1980). Some conditions maximizing eyewitness accuracy: A learning memory model. *Journal of Criminal Justice, 8,* 395–403.

Shapiro, P., & Penrod, S. D. (1986). A meta-analysis of the facial identification literature. *Psychological Bulletin, 100,* 139–156.

Slack, A. T., & Penrod, S. (1982). *Facial recognition in eyewitness testimony*. Unpublished manuscript, University of Wisconsin at Madison.

Stein, J. A. (1981). Review of eyewitness testimony. *Trial Diplomacy Journal, 4,* 61–63.

Sussman, E. D., & Sugarman, R. C. (1972). The effect of certain distractions on identification by witnesses. In A. Zavala, J. J. Paley, & R. R. J. Gallati (Eds.), *Personal appearance identification*. Springfield, Illinois: Charles C. Thomas.

Wall, P. M. (1965). *Eyewitness identification in criminal cases*. Springfield, Illinois: Charles C. Thomas.

Warnick, D. H., & Sanders, G. S. (1980). Why do eyewitnesses make so many mistakes? *Journal of Applied Social Psychology, 10,* 362–366.

Weinstein, J. (1981). Review of eyewitness testimony. *Columbia Law Review, 81,* 441–457.

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. (1985). Verbal descriptions of faces from memory: Are they diagnostic of identification accuracy? *Journal of Applied Psychology, 70,* 619–626.

Wells, G. L., & Leippe, M. R. (1981). How do triers of fact infer the accuracy of eyewitness identifications? Using memory for peripheral detail can be misleading. *Journal of Applied Psychology, 66,* 682–687.

Wells, G. L., & Lindsay, R. C. L. (1985). Methodological notes on the accuracy-confidence relationship in eyewitness identifications. *Journal of Applied Psychology, 70,* 413–419.

Wells, G. L., & Murray, D. M. (1983). What can psychology say about the *Neil v. Biggers* criteria for judging eyewitness accuracy. *Journal of Applied Psychology, 68,* 347–362.

Wells, G. L., & Murray, D. M. (1984). Eyewitness confidence. In G. L. Wells, & E. F. Loftus, (Eds.), *Eyewitness testimony: Psychological perspectives*. New York: Cambridge University Press.

Wells, G. L., Lindsay, R. C. L., & Ferguson, T. J. (1979). Accuracy, confidence, and juror perceptions in eyewitness identification. *Journal of Applied Psychology, 64,* 440–448.

Wells, G. L., Lindsay, R. C. L., & Tousignant, J. P. (1980). Effects of expert psychological advice on human performance in judging the validity of eyewitness testimony. *Law and Human Behavior, 4,* 275–285.

Woocher, F. D. (1977). Did your eyes deceive you? Expert psychological testimony on the unreliability of eyewitness identification. *Stanford Law Review, 29,* 969–1030.

Woocher, F. D. (1986). Legal principles governing expert testimony by experimental psychologists. *Law and Human Behavior, 10,* 47–61.

Yarmey, A. D. (1986). Ethical responsibilities governing the statements experimental psychologists make in expert testimony. *Law and Human Behavior, 10,* 101–115.

Yarmey, A. D., & Jones, H. P. T. (1983). Accuracy of memory of male and female eyewitnesses to criminal assault and rape. *Bulletin of the Psychonomic Society, 2,* 89–92.