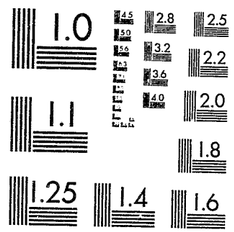

Factors Affecting the Accuracy of Eyewitness Identifications: Investigating the Validity of the U. S. Supreme Court's Guidelines

Final Report, National Institute of Justice Grant No. 82-IJ-CX-0012

John C. Brigham
Department of Psychology
Florida State University
Tallahassee, Florida 32306

Eyewitness identifications play an important role in the American criminal justice system. Analysis of the psychological factors involved in the ability to accurately identify people has been an area of interest to psychologists since the beginning of the century, as evidenced by Muensterberg's 1912 book, On The Witness Stand: Essays on Psychology and Crime, and the series of articles by Whipple (1909; 1912; 1918) on the psychology of observation, report, and testimony. But it is only in the past decade that integrated programs of research on the psychology of eyewitness identification have come into being. This report briefly summarizes the views of legal scholars on the status of eyewitness evidence and presents in detail the results of four scientific research studies designed to identify and clarify factors affecting the accuracy of eyewitness identifications.

Views of Attorneys, Judges, and Law Enforcement Personnel

While eyewitness testimony is a valued prosecutorial tool, many legal scholars have pointed out that eyewitness identifications are frequently inaccurate. For example, Wall (1954, p. 5) has suggested that "the major problem, where actual guilt or innocence is involved, has been and is now the problem posed by the evidence of eyewitness identification." Judge Jerome Frank has suggested that erroneous identification of the accused may constitute the major cause of known wrongful convictions (Frank & Frank, 1957, p.

61). Similarly, Houts (1956, pp. 10-11) has stated that "eyewitness identification is the most unreliable form of evidence and causes more miscarriages of justice than any other method of proof." Judge Nathan Sobel (1972, p. vi) has suggested that judges and lawyers agree that mistaken identification has been responsible for more miscarriages of justice than any other factor -"moreso, perhaps, than all other factors combined."

Despite such warnings, eyewitness evidence is still given great weight in the legal system. Wall has suggested that, "in general, juries are unduly receptive to identification evidence and are not sufficiently aware of its dangers...and...evidence of identification, however, untrustworthy, is taken by the average juryman as absolute proof" (Wall, 1965, p. 19). More recently Woocher (1977, p. 970) has echoed this claim: "Most juries, and even some judges, are unaware of the sources of error in eyewitness testimony and subsequently place undue faith in its veracity." Woocher suggests that this "seriously aggravates the problem" of incorrect identifications.

We might expect that the extent to which criminal attorneys agree with this analysis would depend on their orientation, that is, either prosecution or defense. A recent statewide survey of 235 attorneys in Florida shows that this is the case. Almost 90 percent of the defense attorneys surveyed (attorneys working in the Public Defender's offices or as private criminal attorneys) felt that judges and juries place too much emphasis on eyewitness evidence. Most prosecuting attorneys working in State Attorney's offices disagreed: about 70 percent of them felt that judges and juries place the right amount of emphasis on eyewitness testimony (Brigham, 1981).

A parallel statewide survey was carried out of 201 law enforcement officers, personnel in police departments and county sheriff's departments who

evaluate eyewitness evidence (Brigham & WolfsKeil, 1983). About 65 percent of the law enforcement personnel also felt that judges and juries place an appropriate amount of emphasis on eyewitness evidence. The majority of defense attorneys (64 percent) felt that 75 percent or less of the eyewitness identifications they had encountered were "probably correct," while most prosecuting attorneys (84 percent) and law enforcement personnel (63 percent) responded that 90 percent or more of the identifications they encountered were "probably correct."

Important U. S. Supreme Court Decisions

There had been no coherent body of law concerning appropriate eyewitness identification procedures until 1967 when the U. S. Supreme Court decided on a trilogy of cases: U.S. v. Wade (388 U.S. 218, 87 S.Ct. 1926, 18 L.Ed.2d 1149, 2967), Gilbert v. California (388 U.S. 263, 87 S.Ct. 1951, 18 L.Ed.2d 1178, 1967) and Stovall v. Denno (388 U.S. 293, 87 S.Ct. 1967, 18 L.Ed.2d 1199, 1967). In these rulings, the Court discussed at length the dangers of bias in eyewitness identifications. Because of such bias, the Wade and Gilbert decisions appeared to establish the right to counsel at all pretrial corporeal identifications (live lineups and showups). However, a later Supreme Court decision, Kirby v. Illinois (406 U.S. 682, 925, Ct. 1877, 32 L.Ed. 2d 411, 1982), limited the impact of Wade and Gilbert by holding that the right to counsel applied only after the initiation of adversary judicial criminal proceedings. Although the Kirby decision was not a model of clarity, it implied that such proceedings were initiated only after the holding of an advisory preliminary hearing (which is a relatively rare proceeding in most jurisdictions) or the filing of a formal charge by indictment

or injunction (Yetter, 1980). Therefore, in "pure" identification cases, wherein the eyewitness identification is the crucial part of the preindictment phase, the suspect does not have benefit of counsel. Yet, as analysts (e.g., Sobel, 1979) have pointed out, this is precisely where counsel may be most needed.

In the Stovall case, the Supreme Court held that due process considerations may require the exclusion of identification evidence produced by a suggestive pretrial identification procedure, even when the right to counsel in inapplicable. The focus of the due process inquiry, the Court indicated, is whether the identification procedure was "unnecessarily suggestive and conducive to irreparable mistaken identification" (Stovall v. Denno, 1967, pp. 301-302.

In Stovall, the victim, a white doctor, underwent major survery to save her life after receiving multiple stab wounds during an unsucc ssful attempt to save her husand, also a doctor, from an assailant. The police arrested Stovall, a black man, for the crime. Shortly after the arrest, five police officers and two prosecutors, all white, brought the handcuffed defendant to the victim's hospital room, where she identified him. The need for immediate action stemmed from uncertainty about whether the victim would survive. She later identified him again at trial and he was convicted. On appeal, the Supreme Court held that the hospital room showup did not violate due process. As the Court viewed it, the showup may have been suggestive, but it was not "unnecessarily" so. In their words,

> Faced with the responsibility of identifying the attacker, with the need for immediate action and with the knowledge that (the doctor) could not visit the jail, the police followed the only feasible procedure and took Stovall to the hospital room. Under these circumstances, the usual police station lineup, which

> Stovall now argues he should have had, was out of the question (Stovall v. Denno, 1967, p. 302).

Despite the Court's acceptance of the identification procedures in Stovall, one law professor has noted, "A procedure more conducive to mistaken identification is hard to imagine, and had there not been such an emergency need for an immediate identification, the due process result in Stovall undoubtedly would have been different" (Grano, 1984, p. 327).

Some observers felt that the Stovall decision meant that unnecessarily suggestive identification procedures would automatically lead to the exclusion of the identification evidence at the trial (Grano, 1984). However, two subsequent U. S. Supreme Court decisions, Neil v. Biggers (409 U.S. 188, 93 S. Ct. 375, 34 L.Ed. 2d 401, 1972) and Manson v. Brathwaite (432 U.S. 98, 97 Ct. 2243, 53 L.Ed. 2d 140, 1977) indicated that, in the Court's view, the due process test rests on more than unnecessary suggestion. The Court emphasized that exclusion of testimony on due process grounds is required only when an unnecessarily suggestive procedure actually creates a substantial risk of mistaken identification.

The basic facts of the Biggers case are as follows. In January, 1965, a youth with a butcher knife assaulted the victim in her kitchen doorway, which was illuminated somewhat by a light in a nearby bedroom. After threatening to kill her if she did not cooperate, the youth walked the victim to a moonlit wooded area and raped her. The entire incident took 15 to 30 minutes. During the next 7 months, the victim observed several lineups and photographic displays without identifying anyone. In mid-August, the police summoned her to the station, and when she arrived, two detectives walked the defendant by her and directed him to say, "Shut up or I'll kill you." The

victim identified the defendant. The defendant's subsequent conviction rested almost entirely upon the victim's identification of him. The state appellate courts upheld the conviction, but a federal court, in a subsequent habeas corpus proceeding ordered the defendant's retrial because of the unnecessarily suggestive showup. The federal Court of Appeals upheld this action, but the U. S. Supreme Court found it erroneous.

The Supreme Court agreed with the lower federal court that the showup was suggestive - as all one-person showups are - and unnecessarily so. The police had claimed that their small community made it impossible for them to assemble men comparable to the defendant so that a fair lineup could be held, but the Court did not find this excuse for holding a one person showup persuasive. Moreover, unlike the situation in Stovall, the police had no emergency need for an immediate identification. Nevertheless, the Court found little danger of mistaken identification. The Court emphasized that the victim had an adequate opportunity to view the rapist both in the artificial light of her hallway and in the moonlit wooded area where the rape occurred. The Court also observed that the victim had given a "more than ordinarily thorough" description of the rapist and that she was positive in her identification of the defendant. The Court thought that the 7-month delay between the crime and the identification was a negative factor, but during that interval the victim had never identified anyone else at various identification procedures she had attended, and this, in the Court's view, bolstered the reliability of her identification (Neil v. Biggers, 1972, pp. 200-201).

Like Biggers, Manson reaffirmed that the due process test focuses on unnecessarily suggestive identification procedures that create a substantial likelihood of mistaken identification. The Court once again rejected arguments that it should create a rule requiring automatic exclusion of identification evidence whenever the police employ an unnecessary suggestive procedure. Such a rule would go too far, the Court explained, for it would keep from the trier of fact identification evidence even when it is fully reliable and relevant. The Court acknowledged that such a rule would have more deterrent effect on the police, but it felt that its rule, focusing on the likelihood of mistaken identification, would also influence police behavior, for the police would not be inclined to take the risk that a court would suppress evidence under the more lenient due process standard. Summarizing its position, the Court said that it did not view with "unlimited enthusiasm" those "inflexible rules of exclusion that may frustrate rather than promote justice" (Manson v. Brathwaite, 1977, p.113).

Photographic identification techniques may be particularly susceptible to bias, yet in 1973 the U. S. Supreme court held in United States v. Ash (413 U.S. 300, 93 S.Ct. 2568, 37 L.Ed.2d 619, 1973) that there is no right to counsel at any photographic identification procedures. Sobel suggests that the Ash decision

> indicates a frightening unawareness of the basic dangers of photo identification. It is a commonplace that witnesses are less reluctant to make identifications and fasten guilt in photo identification procedures than in corporeal lineup procedures. It is precisely because defendant is not present that the witness is complacent about overcoming doubt ans fastening guilt--often with the observation that this looks "most like the man" (Sobel, 1979, p. 84).

Factors Affecting the Accuracy of Eyewitness Identifications

The Biggers and Manson cases are particularly important, from a psychological point of view, because in them the Supreme Court enumerated five con-

ditions to be considered in evaluating eyewitness identification evidence: (1) the opportunity of the witness to view the criminal at the time of the crime; (2) the length of time between the crime and the identification; (3) the level of certainty demonstrated by the witness at the identification; (4) the accuracy of the witness's prior description of the criminal; and (5) the witness's degree of attention during the crime. These same five factors were restated and reaffirmed several years later in Manson v. Brathwaite (1977). Manson added that these factors should be weighted against the corrupting influence of the suggestive identification procedure. The Court did not explain how this weighing is to be done.

These five factors are of crucial importance not only because they affect the evaluation of eyewitness evidence in the courtroom but also because they are likely to determine the ways in which criminal justice personnel gather and utilize eyewitness evidence in the first place. Hence it is terribly important that these five criteria which now guide the gathering and utilization of eyewitness evidence fit the facts of the situation as closely as possible. And what are the facts?

In the early 1970s when the court was creating the Biggers decision almost no empirical research was available on factors affecting the accuracy of eyewitness identifications (the Malpass and Kravitz 1969 study is an exception). Thus the Court had to rely on folk wisdom or "common sense" in coming to its decision. The problem with such an approach is that folk wisdom or "common sense" can be flagrantly wrong in some areas. If the "educated guesses" made by the Supreme Court in the early 1970s are incorrect, then they might even increase the frequency of miscarriages of justice created by inappropriate evaluations of eyewitness evidence.

Therefore, there is a strong need for basic scientific research which can shed light on the accuracy of the Court's five "educated guesses" and, further, identify other criteria which may also be valuable for evaluating eyewitness evidence. The studies carried out under this grant were designed to gather systematic data on the influence of three of these factors - witness certainty, arousal (attention), and accuracy of prior desriptions - on the accuracy of eyewitness identifications.

Recent research has generally supported the importance and relevance of the first factor, the witness's opportunity to view the subject. With respect to the second factor, the time elapsed before the identification occurs, research data are somewhat mixed (see, for example, Barkowitz & Brigham, 1982; Loftus, 1979; Shepherd, 1983; Yarmey, 1979), although the majority of studies have found that recognition becomes worse after longer time delays, as the Supreme Court predicted. Investigations of the third factor, the witness's certainty about his or her identifications, have produced even more inconsistent findings with respect to accuracy. Most research studies have found either a weak positive relationship or no relationship at all between an eyewitnesses' degree of certainty in his or her identification and the actual accuracy of that identification (for reviews of these findings see Deffenbacher, 1980; Leippe, 1980; Wells & Murray, 1983). Research data are virtually nonexistent with respect to the fourth factor, accuracy of prior descritions.

Studies of the fifth factor, the witnesses' "degree of attention" are rare. However one factor related to attention has received some attention: amount of arousal or stress that a witness experiences. Psychologists have predicted, in line with the Yerkes-Dodson Law, that facial recognitions

should be relatively poor under very low levels of arousal (e.g., boredom, sleepiness) and very high levels of attention (e.g., high stress, fear, presence of a weapon) (Deffenbacher, 1982). Attempting to assess the impact of arousal on recognition, Johnson and Scott (1976) set up two scenarios. In their high-arousal condition, subjects overheard an argument and then saw the target person rush out of a room with a letter opener in his blood-covered hands. Subjects in their low-arousal condition overheard a discussion and saw a man leave with a pen in his grease-covered hands. Johnson and Scott found no direct effect for level of arousal on recognition accuracy. Brigham, Maass, Martinez, and Whittenberger (1973) manipulated arousal by exposing subjects to eletric shock or no shock while they viewed photos of faces. They found a significant main effect for arousal--recognition accuracy was poorer under high-arousal conditions. Ethical and practical considerations make it very difficult to study the impact of very high stress on eyewitness accuracy (see, for example, Malpass & Devine, 1980). Nevertheless, the importance of this factor in these preliminary research results suggests that additional research in this area would be extremely valuable.

In any research endeavor in social science, there are necessary trade-offs between considerations of experimental realism (the impact that the research study has on the participants), mundane realism (the extent to which the research situation is similar to everyday events), and external validity (the extent to which the findings can be generalized to other situations and samples of people). (See Carlsmith, Ellsworth, and Aronson, 1976, for detailed discussion of these concepts). In our studies we attempted to achieve relatively high levels of all three factors, within ethical and practical constraints. Study 1 employs a "staged crime" setup, involving high experimental realism and high mundane realism. Study 2 investigated ability to identify a number of blacks and whites, increasing the external validity of the findings. Studies 3a and 3b involve a person-to-person situation of moderately high experimental and mundane realism. The purpose, design, and outcome of each of these studies is presented below.

Study 1: Further Clarifying the Relation Between Eyewitness Certainty and Accuracy

As discussed above, the U. S. Supreme Court listed "level of certainty" of the eyewitness as one of the five factors affecting the likelihood that an eyewitness identification will be accurate. The Court felt that an eyewitness who is certain is more likely to be accurate than one who is uncertain. And, indeed, it seems intuitively reasonable that people who are more confident in the accuracy of their judgements would be more accurate, on the average. Wells, Lindsay, and Ferguson (1979) have shown that mock jurors are more likely to <u>believe</u> highly confident witnesses than they are to believe less confident witnesses. Yet, several studies (e.g., Brown, Deffenbacher, & Sturgill, 1977; Clifford & Scott, 1978; Leippe, Wells, & Ostrom, 1978) have found no significant relationship between witness confidence and accuracy. Several other studies (e.g., Lindsay, Wells, & Rumpel,1981; Wells, Lindsay, & Ferguson, 1979) have found confidence accuracy relationships which were statistically significant but still quite small, accounting for less than 10 percent of the variance.

The prevailing view of the accuracy-confidence research is exemplified in a <u>Psychology Today</u> article (Bazelon, 1980) in which the large subheading

claims, "Recent studies show no relationship between an eyewitness's confidence and the accuracy of his testimony." This is not really true; the data are mixed. Deffenbacher (1980) has reviewed some 40 separate studies and experimental conditions within larger studies that have addressed this question. He notes that about half the studies have found a relationship between confidence and accuracy that was statistically significant. Deffenbacher suggests that the majority of the studies obtaining this significant relationship are those in which the identification took place in relatively optimal conditions. These conditions include: prior warning of an impending memory test, situational stress that is nondebilitating but adequate to promote a high level of vigilance, a good opportunity to observe the target, high familiarity with the target, a brief retention interval, similarity of viewing and testing conditions, low similarity between the target and forced-choice test distractors, additional consistent information presented during the retention interval, and a forced-choice test of memory with unbiased instructions. Clearly, many of these conditions are not likely to be present in real-world criminal identification situations (see also Leippe, 1980).

There is another factor which may be of considerable importance in affecting whether or not accuracy will be related to confidence--namely, the consequences for the suspect of being identified. In the majority of eyewitness identification studies carried out thus far, witnesses know that their responses will have no real consequences for the person being identified, since it is "only" an experiment. If, on the other hand, witnesses believe that the situation is real and their responses will have an impact on the suspect, it seems possible that witnesses will be more careful abot their

responses and about how certain they say they are. In general, people appear to be certain about the accuracy of their judgments, whether or not the judgments are in fact accurate (Einhorn & Hogarth, 1978). Researchers have argued that the same appears to be the case in the eyewitness context. Leippe, Wells and Ostrom (1983, p. 348), for example, argued that "witnesses only were willing to make an identification when they were fairly certain. Whether or not their identification was correct, they were confident about it." Malpass and Devine (1981) have argued similarly, "The confidence-accuracy relationship is generally small or absent, but witnesses are confident in whatever choice they make" (Malpass & Devine, 1981, p. 488).

As an example of the high certainty people may place on their judgements, Leippe, Wells, and Ostrom (1978) staged a theft in front of college students and then had the students try to identify the thief in a photograph lineup. They found only a low, nonsignificant correlation between certainty and accuracy of identification. Leippe and his colleagues accounted for this low correlation in terms of an absence of variability in confidence judgements given by subjects who were willing to make an identification. More than 88% of the subjects who made an identification attached at least a 7-point confidence judgement to their choice. Confidence judgements were measured on a 10-point scale with 10 labelled as "completely certain he was the thief."

Malpass and Devine (1981) expanded upon the results of Leippe et al. (1978) by examining the responses of choosers and nonchoosers to a target-present or target-absent live lineup under either biased or unbiased lineup constructions. Biased instructions strongly implied that the thief was in the lineup, unbiased instructions stated that the thief might or might not be

in the lineup. Malpass and Devine found a strong positive relationship between choosing and certainty across and within conditions: certainty was significantly higher among witnesses who chose than among those who did not choose. The overall correlation between certainty and accuracy, however, was small and insignificant. When the target was in the lineup, confidence and accuracy were positively correlated, but confidence and accuracy were negatively correlated when the target was not in the lineup.

Malpass and Devine (1981) interpreted these results by arguing that the confidence-accuracy relationship was mediated by choosing. In the target-present lineups, a high proportion of choices were made, and a large portion of these were accurate. Because witnesses were confident in their choice, confidence and accuracy were positively related. On the other hand, in the target-absent lineups witnesses who made a choice were both confident and wrong. Hence, confidence and accuracy were negatively correlated.

Leippe (1980) has pointed out that social psychological research suggests that the act of identification is a social behavior that may have important cognitive consequences. The first important cognitive consequence of making an identification involves commitment. Social psychologists (e.g., Brehm & Cohen, 1962; Kiesler, 1971) have repeatedly found that public, verbal commitment to a position generally strengthens one's belief in that position and promotes resistance to discrepant information, particularly if the commitment is voluntary. Accordingly, Leippe argued that "eyewitnesses who publicly commit themselves to a lineup choice will express an elevated level of confidence (though certainly not of accuracy) and possibly become less receptive to the idea that they might be mistaken" (p. 268).

The second important psychological consequence of making an identification involves what Bem (1972) has labelled self-perception. Self-perception theory proposes that perceptions of one's own behavior can create an emotion or an attitude. Applied to the lineup situation, this suggests that an attitude of confidence will be formed as a result of making an identification: "Eyewitnesses should report a positive sense of confidence in memory after they make an identification, as if they were saying to themselves, 'I really must be sure that was the person, since I was willing to choose that person'" (Leippe, 1980, p.269).

In accordance with self-perception theory, Murray and Wells (1982) hypothesized that a predicision measure of confidence (i.e., confidence that one will be able to make an accurate identification) might be better correlated with accuracy than postdecision measures. In order to test this hypothesis, they exposed subjects to a staged theft followed by a photo lineup. For some subjects, confidence was measured before viewing the photo lineup (predecision confidence); for others, confidence was assessed after viewing the lineup and making a decision (postdecision confidence). Murray and Wells (1982) found no support for the self-perception prediction. First, choosers and nonchoosers were equally confident in their decisions. Second, when witnesses were informed prior to the lineup that the theft had been staged, postdecision confidence was significantly related to accuracy, but predecision confidence was not. Furthermore, when witnesses were not debriefed prior to the lineup, neither predecision confidence nor postdecision confidence were significantly related to accuracy.

The Murray and Wells (1982) findings suggest that the hypothesis that witnesses are confident in an identification regardless of whether that identification is correct or incorrect should be examined more critically. The

present experient was designed to empirically assess the effects of choosing (someone from a lineup) on the confidence-accuracy relationship in eyewitness identifications. Following a staged theft, witnesses were exposed to either a target-present or target-absent photo lineup under biased (in a positive or negative direction) or unbiased lineup instructions. Both predecision and postdecision certainty were assessed for half the subjects and only post-decision confidence for the other half. Finally, data were assessed using three different officer-thief teams so that the generalizability of the results could be evaluated across different interrogator-target pairings.

Method

Overview

Subjects (research participants) witnessed a staged theft in groups of six and then were individually interviewed by a "campus policeman." During this interview the identification procedure was carried out. Each subject was randomly assigned to one of twelve conditions in which s/he received either positively biased, negatively biased, or unbiased instructions, and the thief's picture was either present or absent in the photo lineup. In addition, subjects' certainty in their choice (or lack of one) was rated either before the identification or after the identification. After all subjects had been interviewed they were debriefed together.

Subjects

Subjects were introductory psychology students at Florida State University who participated in partial fulfillment of a course research participation requirement. There were 148 subjects in the present study, but the data of four subjects were dropped: two subjects admitted recognizing lineup photos of acquaintances, one subject saw the confederate-thief

discussing something with the experimenter before the experiment began, and one subject indicated prior knowledge of the experimental manipulations. The remaining 144 subjects were randomly assigned to conditions in the 3 x 2 x 2 design, with 11 to 13 subjects per cell. There were 12 possible combinations ("cells") of the 3 variables (3 x 2 x 2) and each subject was exposed to only one combination of factors.

Photograph Lineups

Photograph lineups for each of the three confederate/thieves employed in the present study were constructed by personnel at the Florida State University Police Department who construct such lineups for actual criminal cases. The photographs were all Polaroid color prints, 7.25 cm by 9.5 cm "bust shots" taken from a distance of 1.10 meters. To construct the lineup, the picture of the "thief" was given to a police officer who then selected six distractor photos from a larger group of photos of similar-appearing persons of approximately the same age (all photos were of white males since the three thieves were white males). Police judgments of distractors' similarity to the thieves were based on hairstyle, hair color, skin color, and general appearance. In target-present conditions, the confederate thief's photo appeared in the lineup. In target-absent conditions, the thief's photo was replaced by an additional distractor photo. Lineups in all cases contained six photos.

Procedure

The theft. Subjects entered the laboratory in groups of seven (six subjects and an accomplice of the experimenter). The experimenter presented a brief introduction to the study, telling subjects it was a study on self-presentation and pointing out some of the audiovisual equipment to be used in

the experiment. After one of the subjects volunteered to be videotaped first (the accomplice volunteered if no one else did), the experimenter turned to start the video recorder only to notice that there was no tape in the machine. He excused himself to go to his office and find a videotape.

Approximately two minutes after the experimenter's departure, a second accomplice slowly opened the door to the laboratory and peeked in at the group of subjects. He then entered the room, walked to the video equipment and handled the video camera, apparently inspecting it carefully. If no subject by this time had questioned the accomplice-thief's purpose in handling the equipment, the accomplice-subject did so, making sure that the other subjects were now paying attention to the thief. The thief did not answer the question but instead hastily unscrewed the camera from its tripod and ripped the electrical cord from its socket. As he ran from the room, he knocked a metal ashtray to the floor, underscoring his urgency and making a loud noise. This served as a further cue that the situation was a robbery and not merely the case of another graduate student coming in to borrow the camera. The thief quickly entered a nearby office along the outside hall after leaving the laboratory to insure against being followed (several groups of subjects, apparently caught up in the situation, did attempt to follow the thief but were never quick enough to catch him). The thief was in the laboratory in full view of the subjects for approximately thirty seconds and stood no more than six feet away from them when "inspecting" the camera.

The identification. Approximately three minutes after the theft took place, the experimenter returned to the laboratory, videotape in hand. One of the subjects (or the accomplice) reported the theft and the experimenter acted distraught, instructing the subjects to remain in the room while he

called the F.S.U. Campus Police. The experimenter returned five minutes later, telling the subjects that an F.S.U. Police Officer was on his way over to question them about the theft. Soon afterward, the uniformed officer entered the laboratory and asked the group of subjects about what had taken place. He then asked the experimenter if he could see the subjects individually in another room to show some photos he had brought with him. He explained that the photos were of suspects already apprehended for similar crimes committed elsewhere on campus. The experimenter gave permission and the subjects were called one at a time to an adjoining room to meet with the police officer.

The officer gave each subject a set of positively-biased, unbiased, or negatively biased verbal instructions (similar to those used by Malpass & Devine, 1981) before calling the student to view the photo lineup. Positive-bias instructions stated, "We believe that the person who stole the video camera is in one of the pictures in this array. Look carefully at each of the pictures in the array. Which of these is the picture of the thief who stole the camera?" The unbiased instructions were, "The person who stole the camera may or may not be in one of the pictures in this array. Look carefully at each of the six pictures in this array and indicate whether or not the thief's picture is present. If so, which one is it?" In the negative-bias condition the policeman said, "We doubt that the person who stole the video camera appears in any of these pictures. But anyway, look at the six pictures in this array and indicate whether or not the thief's picture is present. If so, which one is it?"

Also before viewing the lineup, half of the subjects (those assigned to rate their confidence before making an identification) were asked to rate how

certain they were that they could identify the thief. Certainty ratings were made on a 7-point scale, with one being "completely uncertain" and seven being "certain enough to testify in a court of law." Each subject was then shown the photo lineup. Half of the subjects viewed a lineup including the target photo and the other half viewed a target-absent lineup. Following their choice (or lack of one), subjects in the "after" group were asked to rate the certainty with which their decisions were made, using the same 7-point scale described above.

Care was taken to insure that subjects who had already seen the photo-lineup did not discuss their choices with other subjects. After all subjects in the group had met individually with the officer, the group was thoroughly debriefed and instructed not to tell others about the experiment.

Results

Accuracy

The overall accuracy rate was 61.1% in the target-present condition. The two possible types of errors, choosing someone other than the criminal (a false alarm) and failing to identify anyone (a miss) were about equally prevalent when the criminal was included in the lineup (Table 1). In the target-absent condition, slightly less than half of the subjects (47.2%) made the correct decision that the thief was not in the photo lineup. The overall accuracy rates in the target-present and target-absent conditions were not significantly different at the .05 level, but the false alarm rates differed greatly[1] (all footnotes are at the end of the text). The rate of false alarms (positively identifying someone other than the real thief) was more than twice as great with the criminal-absent lineups than with criminal-present lineups.

Insert Table 1 about here

In order to assess the overall impact of all three independent variables on accuracy, a 3 x 2 x 2 multiple contingency analysis (Winer, 1971) was performed, the three factors being (1) bias in instructions (positive, negative, or none), (2) presence or absence of the thief in the photo lineup, and (3) certainty measured before or after the identification was attempted. For this analysis responses were dichotomized into accurate (hit, correct rejection) and inaccurate (false alarm, miss) responses. This analysis found that people whose confidence was assessed before viewing the lineup were significantly less accurate in identifications than those assessed after the lineup. Forty-five percent of the people in the before conditions compared to 63% percent in the after conditions made a correct response (i.e., they either identified the target in target-present conditions or said he was not in the lineup in target-absent conditions). No other main effects or interactions were significant.

Dichotomized accuracy scores were also submitted to a 3 x 2 x 2 Analysis of Variance (ANOVA). The results paralleled those of the multiple contingency analysis. The only significant effect was a main effect for the before-after confidence assessment manipulation, accounting for 3% of the variance in accuracy. ANOVAs for each officer-thief team showed that this effect was due to the Team 3 subjects. There were no main effects or interactions for Teams 1 and 2, but for Team 3 the before-after confidence assessment main effect was significant, accounting for 7% of the variance.

Certainty

A 3 x 2 x 2 ANOVA was also performed on certainty judgments. (The terms certainty and confidence are used interchangeably in this report.) There was a main effect for predecisional vs. postdecisional confidence, accounting for 5% of the variance in confidence. Confidence judgments assessed after presentation of the lineup were significantly higher than confidence judgments assessed before the lineup was presented. There was also a main effect for target presence, accounting for 6% of the variance in certainty. Subjects in the target-present conditions were significantly more confident than those in the target-absent conditions, replicating one of Malpass and Devine's (1981) findings. ANOVAs for each officer-thief team, however, indicated that the main effect for predecisional vs. postdecisional confidence occurred only with Team 1, accounting for 13% of the variance in certainty. The main effect for target presence was not significant in the individual team analyses. However, the two-way interaction between target presence and lineup bias for Team 3 was significant, accounting for 11% of the variance in confidence.

Certainty-Accuracy Relationship

The overall correlation[2] between certainty and accuracy was .36. The effect was consistently significant and positive across teams: $\underline{r}$ = .43, for Team 1, .41 for Team 2, and .22 for Team 3. The correlation coefficient for Team 3 was somewhat lower than those for Teams 1 and 2 (though not significantly so), possibly because of the significant interaction between instruction bias and target presence on confidence for Team 3. These extraneous influences on confidence for Team 3 may have weakened the relationshp between confidence and accuracy.

Timing of certainty assessment. It has already been noted that ANOVAs on confidence and accuracy revealed no consistent effects of the timing of the confidence assignment (before vs. after). An examination of the self-perception hypothesis, however, requires that we also look at the effect of the before-after manipulation on the confidence-accuracy correlation. These data are presented in Table 2. Overall, the confidence-accuracy relationship

---

Insert Table 2 about here

---

when confidence was assessed before the lineup ($\underline{r}$ = .20) did not differ significantly from the relationship when confidence was assessed after the lineup ($\underline{r}$ = .34). Hence, these data offer no support for the self-perception hypothesis.

It is also important to point out that the lineup procedure apparently had little effect on confidence. For subjects who were assessed on confidence both before and after the lineup, correlational analysis revealed that before and after assessments were strongly correlated ($\underline{r}$ = .66). Furthermore, this relationship was replicated across teams: $\underline{r}$ = .53 for Team 1; $\underline{r}$ = .87 for Team 2; $\underline{r}$ = .72 for Team 3.

Choosing. The overall correlation between certainty and accuracy for choosers was significant ($\underline{r}$ = .50) and was significantly higher than the overall certainty-accuracy correlation for nonchoosers ($\underline{r}$ = .16). This difference was roughly equivalent in all 3 teams (see Table 3). This significant confidence-accuracy correlation for choosers is of considerable forensic relevance since the confidence-accuracy relationship is of greatest legal and practical importance in situations when a witness identifies a suspect (makes

a choice). Police and lawyers are generally not interested in an individual who does not make an identification and persons who do not make an identification are not usually asked to testify. These data revealed a consistent positive relationship between confidence and accuracy for choosers within all conditions.

Insert Table 3 about here

In contrast to the significant positive relationship between choosing and confidence found by Malpass and Devine (1981), the current data revealed a significant negative relationship between choosing and confidence (r = -.23).[3] Nonchoosers were significantly more confident than choosers. Further analyses revealed that this negative relationship was due largely to Team 1 subjects. The differences were not significant in Teams 2 and 3. The data also revealed a significant negative relationship between choosing and accuracy (r = -.23). Nonchoosers were significantly more accurate than choosers were. In contrast to the choosing-confidence analyses, further analyses found that the negative choosing-accuracy relationship was more prevalent in Team 2 and Team 3 than with Team 1.

Discussion

These data offer little support for the hypothesis that witnesses are confident in whatever choice they make. Choosing was not positively related to confidence. Furthermore, there was a significant positive relationship between confidence and accuracy for choosers but not for nonchoosers.

This latter result is of potential forensic relevance because the confidence-accuracy relationship is of greatest legal and practical importance in situations when a witness makes an identification. Other research has also found the relationship between confidence and accuracy to be greater among choosers than nonchoosers (Wolfskeil & Brigham, 1984) but this is not always the case. Some studies have found the confidence-accuracy relationship among choosers to be small or nonsignificant (e.g., Leippe et al., 1978; Lindsay et al., 1981; Wells, Lindsay & Ferguson, 1979; Wells, Ferguson & Lindsay, 1981). Such discrepancies were apparent in a recent meta-analysis (a way of statistically combining the results from many independent studies) of the confidence-accuracy literature (Bothwell, Brigham, & Deffenbacher, 1984). An analysis of the studies reviewed by Wells and Murray (in press) revealed a mean confidence-accuracy correlation coefficient of .23 resulting in a 95% confidence interval for r of -.05 to .51 (a 95% probability that the true correlation is somewhere between -.05 and .51). Hunter, Schmidt and Jackson (1982) point out that in such cases much of the remaining variability might be accounted for by methodological artifacts, such as range restriction and error of measurement. Leippe et al. (1978) found evidence of range restriction in confidence among choosers in their study of crime seriousness. When either confidence or accuracy are of limited variability the relationship between confidence and accuracy will necessarily be small. This was not a problem in the current study: there was no ceiling or floor effect for accuracy and choosers and nonchoosers were equally variable in their confidence. Nevertheless, most eyewitness researchers have not been overly concerned with such methodological artifacts; in fact, overall means and standard deviations for confidence are rarely reported. Eyewitness researchers have instead focused on moderator variables, such as optimality (Deffenbacher, 1980), choosing, and the absence of the offender (Malpass &

Devine, 1981).

The present data suggest that the effects of choosing and target-presence may not have any consistent moderating effects on the confidence-accuracy correlation. The effects that Malpass and Devine (1981) reported regarding these variables were quite different from those found in the current study. Table 4 shows the current results for subjects in positively and unbiased conditions with the target either present or absent. These

---

Insert Table 4 about here

---

results should be compared with Malpass and Devine's (1981, p. 487) Table 2. While Malpass and Devine reported a strong positive correlation between choosing and confidence within conditions and overall, the current data revealed a weak negative correlation between choosing and confidence that varied within conditions. Malpass and Devine reported a strong positive correlation between confidence and accuracy in target-present conditions, a strong negative correlation between confidence and accuracy in target-absent conditions, and an overall correlation between confidence and accuracy close to zero. In contrast, the current data revealed a weak positive correlation between confidence and accuracy in target-present conditions, a moderate positive correlation between confidence and accuracy in target-absent conditions and an overall moderate positive correlation between confidence and accuracy.

Despite the different pattern of findings, these discrepancies give further support for Malpass and Devine's (1981) interpretation of choosing as mediating the confidence-accuracy relationship. When choosing and confidence

are positively correlated, the confidence-accuracy relationship will depend on whether or not the criminal is in the lineup. However, when choosing and confidence are weakly related, the confidence-accuracy relationship is likely to be moderately positive regardless of whether or not the criminal is in the lineup. Furthermore, the current data suggest when choosing and confidence are weakly related, choosers will show a higher confidence-accuracy correlation than will nonchoosers.

The self-perception hypothesis that confidence assessed before a lineup may be more predictive of accuracy than confidence assessed after a lineup also depends upon the assumption that witnesses will be confident in an identification regardless of whether tht identification is accurate or inaccurate. Having found no support for the latter assumption, it was not surprising to find confidence judgments measured before the lineup no more predictive of accuracy than confidence judgments assessed after the lineup. Therefore, these results, in combination with those reported by Murray and Wells (1982), cast doubt on the generalization that witnesses are confident in the accuracy of their identifications regardless of whether the identifications are in fact accurate. Further, the present data indicate little improvement in the confidence-accuracy relationship by assessing confidence before rather than after the lineup. Taken together, these results indicated that, although a moderate relationship was found between confidence and accuracy in the present study (r = .36), this relationship is too small and too variable to use certainty as a reliable and valid indicator of a witnesses' accuracy in identification, contrary to the U. S. Supreme Court's opinion in the Biggers and Manson decisions.

Study 2: Focus of Attention and "Weapon Focus"

As noted earlier, an eyewitnesses' "degree of attention" was another of the five factors enumerated by the Supreme Court in its Biggers and Manson decisions. The aspect of attention which has received the most interest from social scientists has been the witnesses' degree of arousal and possible focus of attention on a weapon.

In this regard, Loftus (1979) has used the term "weapon focus" to refer to:

> ...the situation in which a crime victim is faced with an assailant who is brandishing a weapon. The weapon appears to capture a good deal of the victim's attention, resulting in, among other things, a reduced ability to recall other details from the environment, to recall details about the assailant, and to recognize the assailant at a later time (1979, p. 35).

This phenomenon appeared to be involved in one study which found that a weapon-present condition resulted in a decrement of target identifications (Johnson & Scott, 1976). The experiment used two levels of arousal in a facial recognition task. In the high arousal condition, subjects overheard an increasingly hostile argument between two males while sitting alone in the waiting room. The dispute culminated with the suspect bursting into the reception room clutching a knife in his blood-covered hands. The low arousal condition consisted of a similar sequence of events, except that it was less dramatic. In the low arousal condition, the suspect appeared after an overheard conversation about equipment failures with grease on his hands and holding a pen. Additionally, in the high arousal condition the subject was interrogated by a confederate who wore a fully equipped police uniform, while low arousal subjects were interrogated by experimenters wearing simple lab

coats. Johnson and Scott (1976) reported that high arousal increased the amount of information the witness could retrieve about circumstances of the encounter, such as physical description of the reception room, content of the conversation overheard, and description of the suspect's exit. However, there was no overall increase in the richness or quality of suspect descriptions or identifications.

There are several intuitively plausible reasons which might account for changes in recognition accuracy when a witness feels threatened with a weapon. First, the potential witness might focus on the weapon rather than attend to cues that could aid in later identification of the assailant. Secondly, the victim's attention would very likely be invested in scanning the surroundings for possible escape routes, at the expense of encoding relevant information for later recall or recognition. A third potential factor that might affect recognition accuracy is the high level of arousal an individual would be expected to experience when threatened with a weapon.

More than 65 years ago Whipple (1915), after reviewing the literature on arousal effects in experimental tasks, concluded that "excitement improves observation and memory of witness up to a given point (variable for different persons) and impairs it beyond that point (p. 233)." This conclusion is essentially identical to the Yerkes-Dodson Law, which states that;

> There is a relation between efficiency and level of arousal such that maximum performance occurs at the most favorable intensities of arousal. Although this relationship varies for different individuals activated by the same motive, and for the same motive in different situations, in general intermediate levels are best (Vinacke, 1968, p. 465).

An attentional explanation for the Yerkes-Dodson Law was offered by Easterbrook (1959). According to his "cue utilization hypothesis," increasing arousal will narrow the attentional field; less important cues will be ignored under conditions of higher arousal, while more attention is paid to central cues. At moderate levels of arousal, this could lead to better memory for central factors. When individuals are extremely aroused, however, their attention field is likely to become so narrow that cue utilization will become increasingly ineffective. This shift in attention becomes particularly relevant in crime situations; the witness in the crime situation may concentrate on stimuli relevant to his/her opportunity to escape rather than the appearance of the perpetrator. Quite naturally, the witness is likely to pay more attention to his/her own well-being and safety than to less essential elements in the environment.

To summarize, Easterbrook's (1959) cue utilization hypothesis postulates that the narrowing and focusing of the attentional field under arousal involves the diminished utilization of peripheral cues and equal if not enhanced use of control, immediately relevant cues. From this postulate, Bacon (1974) has pointed out two related hypotheses that can be empirically tested: (a) There is a differential shift in responsivity under arousal depending upon the initial amount of attention being focused on a cue; i.e., cues which initially attract less attention should show further diminished attention under arousal. (b) There is heightened sensitivity under arousal to those cues on which a high degree of attention was initially focused (p. 82).

Bacon attempted to use an auditory task to assess the focus effect. His subjects were assigned to either of two levels in which the amount of attention paid to auditory cues was manipulated. Results of Bacon's study supported Easterbrook's (1959) hypothesis that the effect of arousal is to diminish the range of cue utilization. However, Bacon cautions that the pursuit rotor task which he used "...is such a compelling task that, despite the manipulations designed to focus attention on the auditory task, it is quite likely that the auditory cue never achieved attentional centrality (p. 86)."

Many crime situations involve very high degrees of arousal (fear, anger, terror, etc.) and often weapons as well. But there are obvious ethical and practical problems involved in simulting such situations for reearch purposes. Clearly, one cannot (and should not) terrorize research subjects to the same degree that a violent crime might cause. However, since the impact of high arousal on eyewitnesses' accuracy is a factor of central importance in many crime situations, researchers have attempted to create milder levels of arousal to see what impact they have on accuracy.

A number of different approaches have been used to manipulate levels of arousal. These methods range from the use of drugs, sleep deprivation (Malmo, 1965; Malmo & Surwillo, 1960; Wilkinson, 1962, 1965), incentives (Bahrick, Fitts, & Rankin, 1952), threat of electric shock (Brigham, Maass, Martinez, & Whittenberger, 1983; Rohn, 1954), and the use of "white noise" (like loud radio static) (Boggs & Simon, 1968; Broadbent, 1957, 1958; Davies, 1968; Hockey, 1969, 1970; Houston, 1968). White noise has been shown to produce physiological changes (elevated EEG readings and Galvanic Skin Response (GSR) readings) which are related to arousal (Berlyne & Lewis, 1963).

M. Eysenck (1975) used white noise at 80 db to investigate the effect of arousal on recall on a semantic memory task. Other researchers (e.g., Berlyne, Borsa, Craw, Gelman, & Mandell, 1965) found that memory for visual

patterns was impaired when subjects were presented with 72 db of white noise during the acquisition stage, while recall appeared to improve at the 58 db level. Schwartz (1974) studied the effects of arousal on retrieval strategies at three different noise levels, no noise, medium noise (65 db), and high noise (85 db). The results of the above studies suggest that high levels of white noise introduce interference or arousal similar to what might be expected in a threatening situation (e.g., Berlyne et al., 1965; Eysenck, 1975; Schwartz, 1974).

The present study combined two of these factors, white noise and the threat of electric shock, to induce arousal in subjects. We also artificially manipulated the focus of our subjects' attention, leading them to focus on the faces of the people viewed, on their hands (creating an orientation like "weapon focus"), or on background features. Another set of subjects were given no instructions on where to focus their attention.

In this study we attempted to recreate the perspective of a convenience-store clerk (a position where robbery is always a possibility) by projecting slides to like-size full-body views of black and white males in a convenience store. Half of the slides suggested that a robbery might be in progress, because they depicted males holding weapons in their hands. The other half of the slides depicted males holding items to be purchased, instead of weapons. Half of our subjects were in a high-arousal state (through the use of the white noise and expectation that shock may be received) while the other half of the subjects were in a relatively low-arousal state (no white noise, no expectation of possible shock). In addition, the focus of subjects' attention was manipulated. Some subjects focused their attention primarily on the faces of the males depicted in the slides, some focused on

Malpass, 1984, for reviews). In general, both laboratory studies and field studies have found that witnesses tend to do better at identifying members of their own race than members of another race. This phenomenon was expected to occur in the present experiment as well, leading the white subjects to be more accurate in their identification of whites than of blacks.

Method

Overview of Design

A 2 x 4 x 2 x 2 factorial design was employed in the experiment. There were two between-subjects variables (arousal and focus) and two within-subjects variables (race of target person and type of object held in hand). Within each arousal/focus cell, each subject saw a total of 24 target photos: six black targets with weapons in their hands, six white targets with weapons in their hands, six black targets with objects other than weapons in their hands, and six white targets with objects other than weapons in their hands.

Subjects

Subjects were 96 white female students enrolled in introductory psychology courses at Florida State University. They received course credits for their participation in the study. Exclusive use of female subjects was based on a finding by Brigham et al. (1983) which suggested that the arousal manipulation might have a more measurable effect on females than on males.

Apparatus

Stimuli for the study included 24 color slides of males standing in a convenience store. They were upper-body photos taken with a Pentax 35mm camera and using 400 ASA color slide film. The 24 color slides were counterbalanced for race of the target person and the type of object held in his hand.

the objects held in the persons' hand (analogous to "weapon focus"), some focused on the background, and some were looking at the entire side as a whole.

As stated above, Easterbrook's (1959) cue utilization hypothesis postulates that the narrowing of the attentional field under arousal involves the diminished utilization of peripheral cues, and sustained if not argumented use of the central, directly focused-upon cues. In general, high arousal subjects would be expected to perform relatively superior at the recognition task as to the extent that their focus condition has direct relevance to the dependent measure. The cue utilization hypothesis strongly implies that there should be an interaction between focus and arousal. Therefore, the face focus-high arousal condition would be expected to rank highest among groups in terms of recognition accuracy, and the hand focus-high arousal condition was predicted to perform most poorly. Theoretically, in the face focus-high arousal condition, subjects would already be focused on relevant cues for the recognition task, so that arousal should further heighten their sensitivity to those cues. That heightened sensitivity to facial cues should consequently facilitate recall to a higher degree than all of the other conditions. In contrast, the hand focus-high arousal group would be focused on cues which are least likely to facilitate recognition.

The "weapon focus" manipulation was expected to create a result similar to that found by Johnson and Scott (1976), that is, weapon-present conditions would lead to diversion of attention from the facial area; thus there would be a decrement in recognition accuracy. To date, the evidence from previous research has generally been supportive of own-race bias in facial recognition (see Barkowitz & Brigham, 1982; Brigham, in press, and Bothwell, Brigham, &

Malpass, 1984, for reviews). In general, both laboratory studies and field studies have found that witnesses tend to do better at identifying members of their own race than members of another race. This phenomenon was expected to occur in the present experiment as well, leading the white subjects to be more accurate in their identification of whites than of blacks.

Method

Overview of Design

A 2 x 4 x 2 x 2 factorial design was employed in the experiment. There were two between-subjects variables (arousal and focus) and two within-subjects variables (race of target person and type of object held in hand). Within each arousal/focus cell, each subject saw a total of 24 target photos: six black targets with weapons in their hands, six white targets with weapons in their hands, six black targets with objects other than weapons in their hands, and six white targets with objects other than weapons in their hands.

Subjects

Subjects were 96 white female students enrolled in introductory psychology courses at Florida State University. They received course credits for their participation in the study. Exclusive use of female subjects was based on a finding by Brigham et al. (1983) which suggested that the arousal manipulation might have a more measurable effect on females than on males.

Apparatus

Stimuli for the study included 24 color slides of males standing in a convenience store. They were upper-body photos taken with a Pentax 35mm camera and using 400 ASA color slide film. The 24 color slides were counterbalanced for race of the target person and the type of object held in his hand.

Objects held were of two distinct types: a weapon (an automatic pistol or a machete), and non-weapon (a can of soup or a small bag of potato chips). For each of the 24 target persons there was also a matching head and shoulder (bust) photograph which was taken in a different setting (standard, plain background) and with different clothing. There were an additional 48 "bust" photographs of other males (again racially counterbalanced) with the same background used for the target persons. The additional 48 photographs served as distractors in the recognition phase of the experiment, making a total of 72 facial photographs.

Other instrumentation included a specially constructed laboratory box equipped with a visual feedback mechanism which utilized two red LED lights to signal "correct" or "incorrect" randomly when one of four buttons (labeled A, B, C, and D) was pressed. Also used was an electroshock unit made by Grass Medical Instruments, model S4K; a Grason-Stadler Model 455C white noise generator which was calibrated to 90 db with a General Radio Company Sound Level Meter Type 1551-C; a Sonic-30 stereo/mono headphone set made by Sonic International Corporation, and a Kodak carousel projector equipped with a zoom lens.

Procedure

A bogus game was employed in order to provide a rationale for the subjects' exposure to the focus and arousal manipulations. For high arousal subjects, contingencies were ostensibly set such that finding a solution to the problem and making a correct response to the stimulus enabled the subject to avoid undesirable consequences. If the subject failed to make a correct response, she was exposed to aversive levels of white noise and was told that she would become more likely to receive electric shocks, provided that the

failures continued. Low arousal subjects were not advised of any negative consequences related to their performance.

Centrality/peripherality of focus consisted of the following four treatment levels: 1) Face focus, in which subjects were instructed to search the target faces for cues in a bogus visual discrimination task; 2) Hand focus, in which the relevant cues for the task were reportedly to be found in the object which the target person held in his hand; 3) Background focus, in which the subjects were told to scan the background for critical features; and 4) Free focus, in which no particular search strategy was suggested. Subjects simply were told that the cues were "somewhere in the picture."

The high arousal treatment consisted of the threat of electric shock and the use of white noise to manipulate the subject's level of activation. M. W. Eysenck (1975) used white noise at 80 db to investigate the effect of arousal on recall in a semantic memory task. Other researchers (e.g., Berlyne, Borsa, Craw, Gelman, & Mandell, 1965) found that memory for visual patterns was impaired when subjects were presented with 72 db of white noise during the acquisition stage, while recall appeared to improve at the 58 db level. Schwartz (1974) studied the effects of arousal on retrieval strategies while using three different noise levels; no noise, medium noise (65 db), and high noise (85 db). The results of the above studies suggest that high levels of white noise introduce interference or arousal similar to what might be expected in a threatening situation (e.g., Berlyne et al., 1965; Eysenck, 1975; Schwartz, 1974). A noise level of 90 db was used in the present study, which seemed appropriate for inducing high arousal without risk of injury to the subject. The use of noise and possible shock was described to the subjects before the study began.

Before proceeding with the experiment, all subjects were screened for visual problems. High arousal subjects were additionally screened for hearing deficiencies and/or any other medical concerns which might have contraindicated their participation in the experiment (e.g., heart trouble, history of epileptic seizures). Subjects were asked to sign a consent form noting that their participation was voluntary. All subjects met screening requirements.

For each experimental run the participant sat at a table upon which was the laboratory box equipped with the 4-button array and the two LED lights which provided the random "correct" or "incorrect" feedback. In the high arousal conditions, electrodes were applied to the index finger and the middle finger of the subjects' nondominant hand. First, the fingers were wiped clean with alcohol-soaked cotton balls, then the electrode paste was applied and the electrodes were secured with white surgical adhesive tape. High arousal subjects also wore the headphones for receiving white noise and were met by experimenters dressed in white lab coats. Following the instructions, each subject was presented with slides of the 24 target persons (in the convenience store setting) arranged in random order. The inspection phase consisted of a 10-second exposure to each individual slide followed by a 5-second interval in which the subject decided on the "correct" button based on (supposed) cues in the slide (cues which were actually nonexistent). Phony feedback was available immediately following each trial, with an approximate frequency of 45% correct and 55% incorrect. "Incorrect" responses by low arousal subjects held no consequences, but high arousal subjects received an immediate 5-second blast of white noise at 90 db and increased probability that they would receive painful electric shock (no shock was

actually administered) with each "incorrect" response.

Prior to each six-slide sequence, high arousal subjects predicted the probability that they would be shocked during the interval on a scale from 0% (certain would not be shocked) to 100% (certain would be shocked). That is, they predicted the probablity that they would be shocked at four points: before the first slide, seventh slide, thirteenth slide, and nineteenth slide. In addition, both high arousal and low arousal subjects rated their level of arousal on a scale from 0 (extremely relaxed) to 100 (extremely tense) at four different points: following the sixth slide, the twelfth slide, the eighteenth slide, and the twenty-fourth slide. After the inspection phase was completed, subjects moved directly to the recognition phase of the experiment. Transition time between the inspection phase and the recognition phase was approximately 60 seconds.

In the recognition phase, subjects were shown 70 facial photographs (busts) which included the 24 target persons shuffled in among the 46 distractors. Photographs were presented individually to each subject at a rate of 10-seconds each. The subject's task was to indicate whether or not the face was a target person by checking yes or no on the identification form. Subjects were routinely debriefed after the recognition task, sworn to secrecy, and invited to make further inquiries about the findings following completion of the experiment.

Results

Manipulation Check

Subjects' self-reports of arousal showed a statistically significant difference between high (Mean = 72.38) and low (Mean = 56.08) arousal groups. Across all trials, high arousal subjects reported a mean probability expec-

tation of 59.74% that they would receive shock. These expectations of receiving shock did not change across time.

Scoring Procedures

Accuracy of recognition was scored in three ways, using the raw number of hits (correct identifications) as the criterion measure and using two types of d' scores which take into account false alarms as well as hits.[4] The d' score is a statistic utilized in Signal Detection Theory which provides a means of dealing with the problem of guessing effects in recognition testing. Guesses (or false alarms) are penalized and there are tables (Elliot, 1964) that list values of d' for each pair of values for the hit rate and false alarm rate. All subjects had four different sets of hits and d' scores, one for each race/type of object combination.

Hit Scores

Hit scores for all conditions are presented in Table 5. A 2 x 4 x 2 x 2 repeated measures ANOVA yielded significant main effects for all four independent variables. Race of the target was the variable which showed the

---

Insert Table 5 about here

---

greatest magnitude of effect, with black targets being identified significantly more often than white targets. The overall hit rate was 19.0% for white targets and 26.2% for black targets. The facial recognition rates were higher in weapon-absent conditions than in weapon-present conditions. High arousal subjects performed significantly above the level of low arousal subjects. Face focus was significantly more accurate overall than the remaining conditions, surpassing hand focus, background focus, and free focus. No

significant differences were found among any of the remaining combinations of hand focus, background focus, and free focus.

Finally, there was a significant interaction between race of the target person and the type of object held. Black targets were identified significantly more often when holding weapons than without weapons. In contrast, white target persons were identified significantly more often in weapon-absent conditions than they were in weapon-present conditions.

D-Prime Scores

A 2 x 4 x 2 x 2 repeated measures ANOVA on D-prime scores (see Table 6) revealed main effects for three of the four independent variables. In direct contrast to the analysis of hits, analysis of the d' scores indicated that white targets were more accurately identified than black target persons

---

Insert Table 6 about here

This was because, although blacks had been identified more often than whites (hits), they were misidentified much more often than whites. The false alarm rate was 40.9% for black photos and 28.8% for white photos. Recognition in weapon-absent conditions was superior for both races of target in d' scores. Finally, there was a significant main effect for focus. Face focus groups remained superior to hand focus groups, background focus, and free focus groups. As in the hits-only analysis, no other combination of hand focus, background focus, and free focus conditions approached a significant level of difference.

Discussion

This study attempted to increase mundane realism in a facial recognition

task by utilizing as inspection stimuli life-sized projections of slides taken in a convenience store. Half of the slides suggested that a robbery might be in progress because they depicted males holding weapons in their hands. In this analog simulation of an eyewitness experience, the main questions were: 1) How would "weapon focus" affect eyewitness identification? 2) What effect would increased arousal have on facial recognition? and 3) How would the race of the target person affect later recognition?

The data revealed that weapon-absent conditions generally obtained higher rates of accuracy than did weapon-present conditions. Arousal did not appear to significantly affect accuracy except when using hits only, where high arousal subjects were statistically superior to low arousal subjects. However, high arousal subjects also made more incorrect identifications (false alarms) than did low arousal subjects. The race of the target person was a significant factor using all methods of analysis, although not in a uniform direction. Blacks were more accurately positively identified than whites only when guessing effects were unpenalized. Where false alarms were accounted for, in contrast, whites were more accurately identified. One manipulation which did show consistent effects was the centrality/peripherality of focus dimension. Face focus produced significantly higher accuracy than all other focus conditions and, further, none of the remaining combinations of hand focus, background focus, and free focus differed significantly from each other. In addition, there was a fairly robust finding of a race by type of objects held interaction, with black targets being more accurately identified in weapon-present conditions than in weapon-absent conditions, while white targets were more accurately identified without weapons than with weapons.

The expected interaction between arousal and focus failed to materalize.

All analises yielded main effects for focus, but a significant arousal by focus interaction did not occur under any of the analyses. In fact, a main effect for arousal was found when using hits as the sole criterion. Failure of the arousal manipulation to induce the magnitude of stress necessary for truly heightened sensitivity appears to be a possible explanation for the lack of interaction between arousal and focus. Although subjects reported increased arousal levels with the white noise/threat of shock manipulation (high arousal M = 72.38, low arousal M = 56.08), they probably perceived the laboratory setting as relatively safe. As one subject reported, introductory psychology instructors had briefed them on what they might expect as reseach participants and assured them that there would no subject abuse in any experiment.

The "weapon focus" phenomenon appeared to be operative: generally, targets without weapons were identified more often than targets with weapons. In addition, the effects of centrality/peripherality of focus was robust across measurement modalities. These findings support the notion that an eyewitness must first encode the suspect's facial features before an effective identification can be made. But although the face focus groups yielded higher identification rates than other groups, only 75% of these subjects reported that they actually focused primarily on facial cues, while 20% said they focused primarily on the hand area. This compared with 87.5% reportedly focusing on the hand area when instructed to do so, and 83.3% of the background focus subjects reporting primary attention to the background. In the free focus condition, more than 98% of the subjects reported a nearly equal split between focusing on the hand area and in the background. When not explicitly instructed to do so, only about 4% of the subjects reported

focusing primarily on facial cues. These self-report data suggest that a substantially smaller percentage of eyewitnesses might focus on the perpetrator's facial features than is commonly believed.

An own-race bias was found for the d' measures but not for hits-only scoring. The own-race bias occurred with the d' measures because there was a differential false alarm rate for black and white faces. Those results are similar to findings reported by Cross et al. (1971) and Goldstein et al. (1977), that black faces were misidentified more often than white faces and that there was a nonrandom distribution of false alarms. Comments made by some of the participants following the experiment might help to explain these findings. Several subjects reported during debriefing that there seemed to be more blacks than whites in the slide presentation, and one subject actually claimed not to know that there were white target persons in the slide presentation. During some of these informal conversations following debriefing, a few subjects were asked which race of target persons they could identify most often and why. Most thought that own-race identification would be easier and cited greater experience and familiarity with white faces and greater ability to distinguish facial features of other whites. Those subjects who thought blacks were easier to recognize presented reasons which seemed more vague, i.e., "Thought there were more blacks than whites;" "Don't know why;" "Some were darker;" and "Some had meaner expressions."

White female introductory psychology students might be an atypical subgroup of the population at large, and therefore there are obvious limits to generalizations from the data. Consequently, extrapolation of these findings to other groups must be advanced with a high degree of caution. However, some of the current findings have strong precedents in other

investigations. Own-race bias, one of the most robust phenomena in the literature (e.,g., see Bothwell, Brigham, & Deffenbacher, 1984; Brigham, in press; Brigham & Barkowitz, 1978; Brigham & Wiliamson, 1979; Cross et al., 1971; Malpass & Kravitz, 1969), was replicated in this study. The consistency of this phenomenon across studies suggests that the courts should be aware of the possible unreliability of cross-racial identification when eyewitness testimony is the primary evidence for the prosecution. Further, a strong case can be made from the available research evidence in this area to include this factor in the standards for judging eyewitness reliability.

Although "weapon focus" has not been as extensively researched as cross-racial identification, there has been at least one study (Johnson & Scott, 1976) in which it appeared to affect subsequent accuracy of memory. "Weapon focus" appeared to be replicated in this present study in that higher recognition rates occurred in weapon-absent conditions than in weapon-present conditions. However, where a race by type of object interaction was found (for hit scores), black targets were identified significantly more often in weapon-present conditions than in weapon-absent conditions. Hence, although "weapon focus" appears to be a factor in recognition accuracy, it remains for future research to clarify this relationship further. It would be interesting to compare the findings in this study to results obtained from a pool of black subjects performing the same task. Of particular interest in this regard is the question of whether or not the pattern extablished in this study would persist, i.e., other-race target persons being more accurately identified in weapon-present conditions and own-race target persons more accurately identified in weapon-absent conditions.

Study 3a The Relationship Between Accuracy of Prior Descriptions and Facial Identification

Although the U. S. Supreme Court included "accuracy of prior descriptions" as one of the five factors to be taken into account in evaluating an eyewitnesses' identification of a suspect, there have been no empirical investigations of this factor. Hence, the purported importance of this factor has been based on pure speculation; Study 3 is designed to gather empirical evidence on this issue. The lack of empirical evidence related to this factor is particularly troublesome since in most criminal cases involving an eyewitness, the witness is asked to provide the police with a verbal description of the perpetrator immediately following the crime. From the witness' description, a sketch or a composite of the suspect is constructed. These drawings and composites are extremely valuable to the police in most cases because mugshots are not available for persons who are not criminals with prior records (Yarmey, 1979). After a suspect is apprehended by the police, the witness is required to view a live (corporeal) lineup or a photograph lineup. Since the witness may assume that the guilty person is in the lineup (Doob & Kirshenbaum, 1973), his/her prior description becomes a crucial factor in the identification process. It has been demonstrated (Clifford & Bull, 1978; Doob & Kirshenbaum, 1973) that in a lineup situation, the witness will often select the person who best fits his/her prior description. This result is further enhanced by the public nature of the witness' description of the suspect. A great deal of psychological research (e.g., see Keisler, 1971) has shown that statements of a public nature can result in commiting the witness to his/her initial description.

Although verbal descriptions of suspects are an important part of police

work, only a few empirical studies have attempted to investigate the influence of verbal descriptions on facial recognition. The results of the studies which have been conducted, however, are consistent. Goldstein and Chance (1971), in a study involving verbal encoding of faces and later facial recognition, found that while faces almost always evoked verbal responses (word associates), the ease with which a face was labeled was not related to accuracy of recognition. This finding led Goldstein and Chance to conclude that there is a "possibility that pictures almost always elicit verbal responses from humans but that verbal responses are not (or need not be) facilitating for recognition." A further experiment (Chance & Goldstein, 1976) was conducted to discover whether face-specific verbalizations, as opposed to simple word associations, facilitated facial recognition. The results of this study showed that subjects who were instructed to describe something about a particular face which would help them to later recognize it were only slightly more accurate in recognition performance than subjects who either generated word associates or only looked at the faces. No significant difference was found between the "describe" and the "associate" group in their ability to recall the verbal labels they had previously generated for each face. Since verbalization only slightly improved recognition accuracy, these authors concluded that there is only a weak involvement of verbal coding in recognition memory for faces. A study conducted by Malpass, Lavigueur, and Weldon (1973) also failed to demonstrate a strong relationship between facial recognition and verbal encoding.

Even though the studies discussed above have been concerned with the witnesses' prior description and later recognition of a suspect, no attempt has been made to measure the accuracy of these prior descriptions. The limi-

tations of this previous research are increased by the fact that in two of the above studies (Chance & Goldstein, 1976; Goldstein & Chance, 1971), subjects were instructed to use only one verbal label to describe a face, and in the third study (Malpass et al., 1973) subjects used composite descriptions that were generated by others to recognize facial stimuli. The present study, in contrast, is an attempt to parallel the actual police procedures involved when a witness describes and later identifies an offender. Groups of people were exposed briefly to a target person and then asked to write out physical descriptions of that person; later they were asked to identify the person from a photograph lineup in which he might or might not be present. Their degree of certainty about their identification was also assessed.

Research findings from other paradigms also suggest that people's ability to recall something seen previously will not be strongly related to their ability to identify the same object on a recognition test. A large body of research (see Flexser & Tulving, 1978; 1982, for reviews of these studies) demonstrates that persons who can recall a word when provided a retrieval cue often cannot recognize the word on a recognition test. This general finding, based upon several variations in paradigm (e.g., see Tulving & Thomson, 1973), has been termed retrieval independence, meaning that cues which are present on a recognition task are uncorrelated with those on a recall test. These findings would suggest that, counter to the Supreme Court's assertion, accuracy of prior description (recall) is not likely to be related to ability to identify someone from a lineup (recognition). However, this has never been assessed in a forensically-relevant situation.

To see whether the relationship between accuracy of prior description and accuracy of identification differed when levels of attention were different, two levels of attention were employed. Depth of processing was utilized as an attention manipulation because several studies (e.g., Bower & Karlin, 1974; Chance, Turner, & Goldstein, 1982; Winograd, 1976; 1978; 1981) have found that deeper processing (the focusing of attention on a personality trait of an individual) leads to greater recognition accuracy than shallow processing (the focusing of attention on a physical trait of an individual). These findings led to the prediction that both identification accuracy and description accuracy would be greater in the high attention/deep processing conditions than in the low attention/shallow processing conditions. Identification accuracy and description accuracy were not expected to be related, however, within either of the two attentional levels.

The type of lineup used was also varied. The target person was either present or absent from the photo lineup that subjects viewed. This was done to determine whether subjects would choose the person who best fit their description even when the actual target person was not present in the lineup. Subjects' scores on a self-monitoring scale (Snyder, 1974; 1979) were also measured since some recent research (Hosch, Leippe, Marchioni, & Cooper, 1984) suggests that degree of self-monitoring may be related to accuracy as an eyewitness. This scale consists of 25 true-false items which are designed to measure social interaction styles related to the observation and control of one's expressive behavior and self presentation. Persons scoring high on the self-monitoring (SM) scale are more concerned with the social appropriateness of their self-presentation to others while persons who score low on the SM scale are less concerned with this self-presentation. It was hypothesized, in line with the Hosch et al. (1984) findings, that subjects high in self-monitoring would be more accurate in their descriptions and identifications

than subjects low in self-monitoring. A fifth main analysis investigated the relationship of subjects' certainty about their identification and their actual accuracy. Due to the mixed nature of results from previous studies (e.g., Study 1; also see Brigham, 1983; Deffenbacher, 1980; Leippe, 1980; Malpass & Devine, 1981), no specific prediction about this relationship was made.

Although the U. S. Supreme Court used the word "accuracy" when writing of the criterion to which a witness' description should be compared, some have argued (Wells & Murray, 1983) that the court was really addressing the congruence between the witness' description and the characteristics of the person identified from the lineup. This interpretation would suggest a different type of analysis, focusing on the correspondence between the subject's description and the characteristics of the lineup member selected, whether target or foil. This congruency analysis was carried out in addition to the accuracy analysis.

Method

Overview of Design

The present study involved a 2 (attention/depth of processing: deep or shallow) by 2 (type of photo lineup: target-present or target-absent) by 2 (target person) factorial design, with all factors serving as between-subjects factors. Fifteen subjects were randomly assigned to each cell. Two target persons were employed to increase the generalizability of the findings. Individual subjects were randomly assigned to one of the attention/depth of processing conditions and the type of lineup viewed was randomly determined for each subject. Dependent measures were subjects' description accuracies, identification accuracy, Self-Monitoring scores, and identification certainty

scores.

Subjects

Subjects were 120 male and female undergraduates enrolled in introductory psychology classes at Florida State University. All subjects received course research credit for their participation.

Instruments

A checklist of the type frequently used by police was administered to the subjects after the departure of the target person. This checklist required subjects to describe the confederate's physical characteristics.[5] A Self-Monitoring questionnaire (Snyder, 1974; 1979) was administered to all subjects. As noted, this scale consists of 25 true-false items which are designed to measure social interaction styles related to the observation and control of one's expressive behavior and self-presentation. Persons scoring high on the scale are those who are concerned with social appropriateness, that is, they are more concerned than others about behaving correctly in social situations.

Also utilized were four sets of 6-photograph lineups which were constructed with the aid of the F.S.U. Campus Police Department. All photographs were Polaroid color prints (7.25 cm. x 9.5 cm.) of lineup members' head and shoulders. Two lineups, one target-present and one target-absent, were constructed for each white male confederate. The foil photos were chosen by the F.S.U. Police from a group of 25 photos of college-age white males.

Development of Lineups

In addition to adopting the standards of the F.S.U. Police for lineup construction, mean similarity ratings were compiled for each potential foil

photo. The similarity ratings, based on the ratings of 60 pilot subjects (30 for each target person), measured each photo's degree of similarity to the target person on a 7-point scale. All of the foil photos chosen for the lineups had mean similarity ratings between 5 ("somewhat similar") and 7 ("very similar"). In the target-absent lineups, the photo receiving the highest similarity rating was substituted for the target person (highest mean similarity ratings were 6.7 and 6.4), with the other foils remaining the same as in the target-present lineup.

Two other pilot studies were performed to obtain "accurate" descriptions of the targets and foils with which to compare the subjects' descriptions. Descriptions of the target persons were obtained from 50 college-student raters (25 for each target person) who viewed them in person for as long as they wished while completing the checklist. Pilot subjects' responses were tabulated into means and standard deviations for the age, height, and weight variables. Modes were computed for the discrete variables. "Accurate" criterion descriptions for each foil photo in Target A's lineup and 1 additional lineup (described below) were obtained from 96 pilot subjects (16 per photo) in a similar manner.

Procedure

Groups of 3 to 6 subjects were told that they were participating in a study concerning expressive behavior and self presentation. Subjects were seated the same distance (4.5 M) from the front of the experimental room and given a booklet containing instructions and the dependent measures. The instructions stated that in a few minutes a person would enter the room. Subjects were told that they would later be asked to estimate how honest he was (for subjects in the high attention/deep processing condition) or how

tall he was (for subjects in the low attention/shallow processing condition). One of the target persons then walked into the room, stood facing the subjects for 15 seconds, and turned around and left the room.

After the target person's departure, subjects completed the description checklist and the Self-Monitoring questionnaire. Following this, subjects individually viewed one of the 6-photograph lineups (target-present or target-absent) for the confederate they had seen. Instructions for the lineup phase of the experiment were: "A photo of the person you just saw and described may or may not be included in this set of photos. Please look carefully at all of the pictures and indicate whether you see a photo of the person you just saw and described." Whether or not they chose a photo, all subjects rated their confidence in the accuracy of their choice on a scale ranging from 1 (uncertain that one's choice was correct) to 7 (certain that one's choice was correct). After this, subjects were completely debriefed and sworn to secrecy.

Description Accuracy and Description Congruency Scores

The accuracy of subjects' descriptions was calculated by comparing their descriptions with those obtained from the raters. As no precedent has been set for the scoring of person's physical descriptions, four different measures of description accuracy were employed. Subsequent analyses determined these measures to be related (r's from .55 to .65, all p's < .001). Since overlapping data were used in the calculation of the description accuracy measures, this lack of independence was expected. In addition to the description accuracy scores, a description congruency score was calculated for all subjects who chose someone from the photo lineup. The congruency analysis involved a comparison of a subjects' completed descrip-

tion checklist with a description of the person (target or foil) chosen from the lineup.

Results

Identification Accuracy Rates

Identification accuracy scores were obtained by calculating the proportion of hits for subjects viewing the target-present lineups or the proportion of correct rejections for subjects viewing the target-absent lineups. Identification accuracy rates are presented in Table 7. Overall

---

Insert Table 7 about here

---

identification accuracy was 70.83%. This accuracy rate was not affected by type of lineup, target-present or target-absent (70.0% versus 71.7% accuracy). Similar accuracy rates were also obtained regardless of the attention manipulation (75.0% and 66.8% for high and low attention). However, there were significant differences in identification accuracies between target persons (80% and 62%).

Analyses of Identification Accuracy

A 2 x 2 x 2 multiple contingency analysis (Winer, 1971) was performed on subjects' identification accuracy scores. Of the seven calculations that were performed, two were statistically significant. The effect of target person on accuracy was significant; Target A was identified more accurately overall than Target B (Table 7). The interactive effect of target person and type of lineup on accuracy was also significant. For Target A, accuracy was greater for target-present lineups (90.0%) than for target-absent lineups (70.0%), while accuracy for Target B was less in target-present lineups

(50.0%) than for target-absent lineups (73.3%). A 2 x 2 x 2 analysis of variance replicated the above findings.

Analyses of Description Accuracy

A 2 x 2 x 2 factorial analysis of variance (attention x type of lineup x target person) was performed on each of the four description accuracy scores. There was a consistent significant main effect for target person across all four measures of description accuracy, with Target A described more accurately than Target B. There was no interactive effect, however, of target person and type of lineup on accuracy. Significance was not obtained in any of the four ANOVAs for the attention or lineup variables or for the interactions among variables.

Identification Accuracy-Description Accuracy

Legal and criminal justice personnel have been most concerned with the situation when a witness identifies a suspect (makes a choice). The witness' expressed level of confidence may then be used in the courtroom to buttress the identification testimony. The most relevant comparison regarding this situation in the present study is the confidence-accuracy relationship for all witnesses who made a choice in both target-absent and target-present conditions (since, in an actual eyewitness situation, it is a matter of dispute whether the lineup shown the witness was target-absent or target-present). In the current study, 75.0% (N = 45) of witnesses exposed to the target-present lineup and 28.3% (N = 17) of the witnesses exposed to the target-absent lineup picked someone from the lineup.

Point-biserial correlations were calculated between each of the four measures of description accuracy and the identification accuracy scores, and between congruence and identification accuracy, within each condition and

across all conditions. Correlations between the four description accuracy scores (and congruence scores) and identification accuracy were separately assessed for persons making and not making a lineup choice. Only six of the 135 correlations (all combinations of target, target presence/absence, and attention, analyzed separately and in all combinations) for choosers reached significance and none of the five correlations across all conditions were significant. None of the 135 correlations for nonchoosers reached significance. Correlations were also calculated between the mean identification accuracy rates in each condition (see Table 8) and the four corresponding description accuracy scores to see whether conditions which produced greater identification accuracy also yielded better description accuracy. None of the four correlations reached significance.

Point-biserial correlations were also calculated between identification accuracy and description congruence scores for the conditions involving Target A. Again, there were no significant results. These results replicate those discussed above, thus providing additional support for the finding of independence between description accuracy and identification accuracy.

Confidence-Identification Accuracy

For witnesses who picked somebody from the lineup, the accuracy-confidence correlation was highly significant, $r = .51$. For purposes of comparison, within the target-present conditions the confidence-accuracy correlation for choosers was $r = .41$. Table 8 presents the confidence-accuracy correlations for all witnesses who made a choice, according to which target person they observed and level of attention. The confidence-accuracy correlation for the target for whom the accuracy rate was significantly lower (Target B) was significantly higher than the confidence-accuracy correlation

for Target A. Confidence-accuracy correlations were also calculated for subjects who did not make a lineup choice. The overall correlation for nonchoosers was not significant.

Insert Table 8 about here

Self-Monitoring

Correlations between Self-Monitoring and the two accuracy measures (description and identification) were also calculated both within and across conditions. Since statistical significance was obtained in only two of these 45 comparisons, it seems most appropriate to treat the two significant findings as chance occurrences. Correlations between self-monitoring and description congruence likewise yielded no significant results.

Discussion

The present study is the first empirical test in a forensically-relevant situation of the U. S. Supreme Court's guideline regarding accuracy of prior descriptions. Results indicate that the guideline is inappropriate; there was no relationship between subjects' accuracy in describing the target person and the accuracy with which they recognized him in a photograph lineup. Similarly, there was no relationship between description congruence and identification accuracy. This means that the congruence between a witness' description of the suspect and his/her lineup choice provides no means by which to judge the accuracy of the lineup decision. Thus, contrary to the Supreme Court's guideline, it cannot be assumed that persons who are accurate in describing another person will also be accurate in recognizing that person.

The observed results are therefore in agreement with those of earlier

laboratory face-recognition studies (e.g., Chance & Goldstein, 1976; Goldstein & Chance, 1971; Goldstein et al., 1979; Malpass et al., 1973) in finding no significant relationship between facial recognition and description accuracy. The results are also similar to those of experiments showing recognition/recall independence (e.g., see Flexser & Tulving, 1978; Tulving & Thomson, 1973). Applying results from these memory experiments to the present study, one could argue that the retrieval cues which were available to subjects during completion of the description checklist did not assist them in identifying the target from the photo lineup. Thus, a subject who accurately described the target person would not, as a consequence, be expected to accurately recognize him later.

The finding that identification accuracy rates were equally high for target-present and target-absent lineups seems to indicate subjects' unwillingness to be pressured into making an incorrect choice. While the multiple contingency analysis indicated a significant interaction between target person and type of lineup on accuracy, observation of individual cells indicated that this effect was attributable to the low identification accuracy of one target person in the target-present lineup.

The substantial correlation between confidence and accuracy for subjects who made a choice is consistent with some recent research (e.g., Study 1; Brigham, Maass, Snyder & Spaulding, 1982; Maass & Brigham, 1982; Malpass & Devine, 1981) which has found confidence-accuracy correlations of this magnitude. Other researchers, however, have found little or no relationship between these variables (e.g., Brown, Deffenbacher & Sturgill, 1977; Leippe, Wells & Ostrom, 1978; Wells, Ferguson & Lindsay, 1981). As noted earlier, a recent meta-analysis of 26 studies investigating the confidence-accuracy

relationship (Bothwell, Brigham, & Deffenbacher, 1984) has shown the results to be extremely variable. The present confidence-accuracy correlation (.51) is at the uper limit of the confidence limits across all 26 previous studies.

The present findings can be interpreted as giving some support for Deffenbacher's (1980) "optimality hypothesis" which asserts that substantial confidence-accuracy relationships will occur only when viewing is done under relatively optimal conditions. The use of a 15-second viewing time, short retention interval, and low stress would suggest that the current situation was of relatively high optimality. In this vein, it is of interest that the confidence-accuracy correlation was somewhat higher in the high-attention conditions ($\underline{r}$ = .64) than in the low-attention conditions ($\underline{r}$ = .41), a difference which would be predicted by the optimality hypothesis. This difference between correlations was not significant, however.

In conclusion, there is no empirical support for the assumption that an eyewitness who accurately describes a perpetrator of a crime will be more accurate in identifying him/her than a witness whose initial description is less accurate. Therefore, the validity of the guideline imposed by the Supreme Court is highly questionable. Since the courts have judicial reason for the continuing use of these guidelines, the communication to the judicial system of research results concerning their validity would seem to be of considerable importance.

Study 3b: Rationale

The central hypothesis of Study 3a, that description accuracy would not be significantly related to identification accuracy, was supported by the data. However, the absence of any effect for level of attention was unex-

pected. To see whether attention would have an impact on recognition or description accuracy we carried out another study, employing a less subtle manipulation of attention. This time the target person knocked loudly on the door, opened it, and either looked at the subjects for 5 seconds (low attention) or said "Pay attention to me" and stood there for 5 seconds (high attention). In most other respects the procedures and analyses paralleled those in Study 3a.

Method

Overview of Design

A 2 (attention: high or low) x 2 (type of photo lineup: target-present or target-absent) x 2 (target person) factorial design was utilized in the study. Fifteen subjects were in each cell, with all factors serving as between-subjects factors. Subjects were randomly assigned to the attention, lineup, and target person conditions. Dependent measures were description accuracy, identification accuracy, self-monitoring scores, and identification certainty scores.

Subjects

One hundred sixteen male and female undergraduates participated in the study to fulfill a course requirement. All subjects received research credit for their participation.

Instruments

The same 25-item self-monitoring questionnaire used in Study 3a (Snyder, 1974; 1979) was completed by all subjects at the beginning of the experimental session. A checklist which required subjects to describe the confederate's physical characteristics was administered to the subjects after his departure.

Four sets of 6-photograph lineups were utilized in the study. Two lineups, one target-present and one target-absent, were constructed for each white male confederate by the F.S.U. Police Department. Foil photos were chosen by the F.S.U. Police from a group of 25 photos of college-age white males. The photographs were Polaroid color prints (7.25 cm x 9.5 cm) of lineup members' head and shoulders.

Development of Lineups

Mean similarity ratings were calculated for each potential foil photo. The similarity ratings were based on the ratings of 55 pilot subjects and measured each photo's degree of similarity to the target person, with the other foils remaining the same as in the target-present lineup. The functional size (Wells, Leippe, & Ostrom, 1979) and effective size (Malpass, 1981) of the lineups were assessed based on the responses of 66 additional subjects. A shortened version of the description checklist (describing each confederate's race, age, height, weight, hair color, hairstyle, facial hair, and eye color) was administered to the subjects to allow computation of the lineup fairness measures. The functional size of target A's lineup was 8.00, the effective size was 4.63. Target B's lineup had a functional size of 13.00 and an effective size of 4.43.[6]

Additional pilot studies were conducted to obtain "accurate" descriptions of the targets and foils. Subjects' descriptions were compared with these pilot descriptions. Fifty college-student raters (24 for each target person) viewed the target persons while completing the description checklist. Means and standard deviations were computed for the age, height, and weight variables; modes were computed for the discrete variables. "Accurate" descriptions for each foil photo were obtained from 96 pilot subjects (16 per

photo) in a similar manner.

Procedure

Groups of 3 to 6 subjects were told that they were participating in a study concerning expressive behavior. After subjects were seated (all at approximately the same distance (4.5 m) from the front of the experimental room), they began to complete the self-monitoring questionnaire. When all subjects appeared to be working on the questionnaire, the experimenter said "I'll be right back" and then left the room.

Several minutes after the experimenter's departure one of the target persons knocked loudly on the door, opened it, and either looked at the subjects for 5 seconds or said "Pay attention to me." He then left the room. A few minutes later the experimenter re-entered and acted as if nothing unusual had happened.

When all subjects had finished completing the SM questionnaire, they were asked to describe the target person they had seen earlier by using the description checklist. Following this, subjects individually viewed one of the 6-photograph lineups (target-present or target-absent) for the confederate they had seen. Instructions for the lineup viewing were: "A photo of the person you just saw and described may or may not be included in this set of photos. Please look carefully at all of the pictures and indicate whether you see a photo of the person you just saw and described." All subjects they rated their confidence in the accuracy of their decision on a scale ranging from 1 (very uncertain) to 7 (very certain). After all subjects completed the lineup phase, they were completely debriefed as a group.

Description Accuracy and Description Congruency Scores

Subjects' description accuracy scores were calculated by comparing their

descriptions with those obtained from the raters. As in Study 3a, four different measures of description accuracy were employed (including a separate analysis for combined age, height, and weight judgements). A description congruency score was calculated for all subjects who chose someone from the photo lineup. The congruency analysis involved comparing subjects' completed description checklist with the description of the person (target or foil) they chose from the lineup.

Results

Identification Accuracy Rates

Identification accuracy scores were obtained by calculating the proportion of hits for subjects viewing the target-present lineups or the proportion of correct rejections for subjects viewing the target-absent lineups. Overall identification accuracy was 56%. A 2 x 2 x 2 ANOVA was performed on identification accuracy scores. There were no significant main effects or interactions obtained for any of the variables.

Analyses of Description Accuracy

A 2 x 2 x 2 ANOVA was performed on each of the four description accuracy scores. There was a significant main effect for target person for the second description accuracy measure. Significance was not obtained in the ANOVA's for any other variable or for the interactions among variables. A 2 x 2 x 2 ANOVA was also conducted on the description congruency scores. The only significant main effect occurred for target person. There were no significant interactions.

Identification Accuracy-Description Accuracy

Point-biserial correlations were calculated between each of the four description accuracy measures and the identification accuracy scores within

each condition and across all conditions. Since only seven of the 108 correlations reached statistical significance, these can best be treated as change occurrences. Point-biserial correlations were also calculated between identification accuracy and description congruence scores. Only four of these 27 correlations reached statistical significance, all in the negative direction. This finding of independence between identification accuracy and description accuracy and congruence supports the results of Study 3a.

Confidence-Identification Accuracy

Correlations were calculated between subjects' identification accuracy and their confidence in the accuracy of their decision. In the present study 67 subjects chose someone from the lineup (N = 42 for TP lineups, N = 25 for TA lineups). For these subjects, the identification accuracy-confidence correlation was near zero, r = -.05. Within the target-present lineup conditions, the correlation was also nonsignificant, r = .07. This finding differs considerably from Study 3a where a substantial relationship (r = .51) was found between certainty and accuracy for subjects who chose someone from the lineup. The contrast between these two sets of findings underscores the difficulty of drawing any strong conclusions about the general relationship between identification accuracy and certainty.

Self-Monitoring

Correlations between self-monitoring and the two accuracy measures (description and identification) were calculated both within and across conditions. None of the overall correlations were significant, echoing the results of Study 3a.

Discussion

The results of Studies 3a and 3b found no relationship between description accuracy and identification accuracy. These results are in clear agreement with all of the past research in experimental social psychology and verbal memory which provide evidence for the independence of recall and recognition process (see WolfsKeil, 1983, for a review of this research). Taken together, these studies have shown that the U. S. Supreme Court's guideline regarding accuracy of prior descriptions is not appropriate.

The failure of the attention manipulation in both studies to produce between-group differences in description and recognition accuracy is problematical. The best explanation for this result is probably that the appearance of someone other than the experimenter was unusual and led subjects in the low attention condition to pay attention to him without being told to do so. The more direct type of attention manipulation in Study 3b was thought to be superior to the more subtle depth of processing manipulation which was utilized in Study 3a, but neither method led to significant differences in recognition rates. It remains to be seen, then, if the pattern of results would be different when attention is manipulated successfully.

The fact that identification accuracy and certainty were not related for persons who chose someone from the lineup in Study 3b is also puzzling. The results of a recent meta-analysis on the confidence-accuracy relationship (Bothwell, Brigham, & Deffenbacher, 1984) has indicated that those who are made confident in their decision are generally more accurate than are less confident persons. The only study which has reported a confidence-accuracy relationship similar to that found in the present study was by Malpass and Devine (1981) who reported an accuracy-confidence correlation of -.06.

Self-monitoring was not related to either description accuracy or identification accuracy in Study 3a and Study 3b. This finding may be due to the

time at which self-monitoring was measured or a restriction in the range of self-monitoring scores. More research is needed on this question.

The results of the present studies are direct contradiction to the U. S. Supreme Court's evaluation guidance concerning the relationship of prior description and identification. The next phase in this research endeavor should be the communication of these empirical findings to our legal system so that eyewitness evidence may be evaluated more effectively.

The Studies: General Discussion

The results of the four studies above provide considerable additional evidence concerning the points enumerated by the U. S. Supreme Court in the Neil vs. Biggers and Manson vs. Brathwaite cases in the 1970s. This series of studies yields information of considerable theoretical and practical importance to our law enforcement and judicial systems which often depend, however reluctantly, on eyewitness evidence. Our statewide survey of attorneys (Brigham, 1981) as well as the second statewide survey of Police and Sheriff's Department personnel (Brigham & WolfsKeil, 1983) indicate that among those who work with eyewitness evidence on a day-to-day basis there is considerable disagreement about the impact of several factors suggested by the U. S. Supreme Court: witness confidence, arousal and "weapon focus," and the accuracy of prior descriptions. The present studies provide data on each of these issues. While the legal and criminal justice systems have not been notably eager to attempt to apply the results of relevant psychological research, such applications can have value in indirect as well as direct ways. As Yarmey (1979, p. 227) has pointed out, "What one generation of lawyers prefer to understand as 'common sense' often depends upon the theory

and findings of the previous generation of investigators."

In brief, the most central finding of this series of studies are as follows. (1) Contrary to previous research, making a choice from a lineup does not significantly increase a person's certainty about the decision he or she has made (Study 1). (2) The certainty-accuracy correlation was most often positive and statistically significant (in Studies 1 and 3a, but not in Study 3b). (3) The certainty-accuracy correlation was usually significantly stronger for witnesses who chose someone from the lineup than for witnesses who chose no one (in Studies 1 and 3a, but not in Study 3b). (4) Contrary to others' predictions, the timing of the confidence assessment (i.e., before or after an identification was attempted) did not affect overall accuracy or the confidence-accuracy relationship (Study 1). (5) Identification accuracy was generally higher for targets initially seen without weapons than for targets initially seen with weapons in their hands (Study 2). (6) When guessing was unpenalized, white subjects more accurately identified black target persons than white target persons, because they were more likely to say "seen before" for black target persons (Study 2). (7) When guessing was controlled for, the white subjects did significantly better at identifying whites than identifying blacks (Study 2). (8) Black targets tended to be more accurately identified in weapon-present than in weapon-absent conditions, while white targets were more accurately identified without weapons than with weapons (Study 2). (9) People focusing on someone's face make more accurate subsequent identifications than people focusing on the target person's hands, or the background, or the entire photo (Study 2). (10) Contrary to the U. S. Supreme Court's assertions, there is no relationship between accuracy of identification and the ability to describe someone accurately (Studies 3a and

3b). (11) No relationship was found between self-monitoring and the ability to identify or describe someone accurately (Studies 3a and 3b).

What can we make of this welter of findings? First, the complexity of the certainty-accuracy relationship has been further demonstrated. The straightforward notion put forth by the U. S. Supreme Court, that more certain witnesses are more accurate, is too simplistic to accurately describe the world. Certainty and accuracy are sometimes moderately related (e.g., Studies 1 and 3a) and sometimes not related at all (e.g., Study 3b). Further research is needed to untangle this complex relationship and identify those conditions under which an eyewitness' degree of certainty provides a valid indication of his or her probable accuracy (Bothwell, Brigham, & Deffenbacher, 1984; Wells & Murray, 1984).

A second major finding is the absence of any relationship between description accuracy and identification accuracy. This guideline provided by the U. S. Supreme Court is clearly wrong--witnesses who are able to accurately describe a target person are <u>not</u> better able to identify a target person than witnesses who cannot describe the target accurately. Recall memory and recognition memory are independent of one another. It seems particularly important that this information be transmitted into the legal/criminal justice system so that authorities do not continue to utilize this invalid criterion in evaluating eyewitness evidence.

The concept of "weapon focus" receives indirect support from this research. Although overall degree of arousal did not significantly affect identification accuracy in Study 2, attentional focus did. People who focused their attention on the target persons' faces were better able to later identify the target persons that were people who focused on what the

target persons were holding in their hands (the condition representing weapon focus), or on the background, or on the entire scene. As noted, the experimental situation created was quite distant, for both ethical and practical reasons, from a crime where a weapon is used. Nevertheless, the study provides the first systematic, empirical evidence supporting the impact of "weapon focus." The interactions of these factors with the race of the target persons (points 6, 7, and 8 above) are intriguing and deserve further investigation.

At an applied level the present research results can be of value to those who are involved in the gathering of eyewitness evidence (e.g., police), the utilization of such evidence (attorneys), and the evaluation of such evidence (judges and, eventually, jurors). The legal/criminal justice system must currently function on the basis of "educated guesses" about the overall validity of eyewitness evidence and the factors which may affect its accuracy. Our surveys of persons directly involved with eyewitness evidence - prosecutors, criminal defense attorneys, and Sheriff's and Police Department personnel - demonstrate that there are wide differences of opinion between and within these groups concerning those issues. A parallel survey of 90 randomly-selected prospective jurors (registered voters) indicates that they too show wide differences in their assessment of which factors should be taken into account in evaluating eyewitness evidence. This latter survey also indicated that prospective jurors are not able to accurately estimate what the findings of research in this area have been (Brigham & Bothwell, 1983).

The results of the research are of relevance to law enforcement personnel who must evaluate the aspects of an eyewitness' behavior, such as con-

fidence in his/her identification, amount of arousal present during the crime, presence or absence of a weapon, and the degree of correspondence between the witnesses' initial description and the actual characteristics of the suspect. Law enforcement personnel cannot make efficient use of their time and resources by gathering and going forth with evidence which turns out to be ineffectual or inaccurate. Hence, having better guidelines for the evaluation of eyewitness evidence, i.e., guidelines based upon results of well-conducted research instead of "educated guesses," can result in considerable savings in time and expenditures within law enforcement agencies. Such research results can be utilized in the training of law enforcement personnel to focus upon eyewitness evidence gathered from persons and in situations wherein accuracy is most likely to occur, yielding eyewitness evidence which is likely to be valid and useful within the legal system.

In a related vein, prosecuting and defense attorneys can utilize the research results in the conduct of cases which involve disputed eyewitness evidence. Again, reliance on (or at least awareness of) the research results can lead to more appropriate usage of eyewitness evidence than do the "educated guesses" of attorneys in our adversarial system. Our attorney survey (Brigham, 1981) indicated that the educated guesses of prosecuting and criminal defense attorneys differ radically from each other in ways which could be predicted from their adversarial roles. The results of research can provide attorneys on both sides with "reality estimates" about the overall accuracy of eyewitness evidence and about the factors which are the most important determinants of this accuracy.

The research results can be of considerable value to judges also, since they must evaluate eyewitness evidence in many situations, including pretrial

evidentiary hearings, deciding whether to admit expert testimony regarding eyewitness evidence, and judicial instructions to the jury in cases involving disputed eyewitness evidence. Currently there is a great deal of disagreement within the legal system about the appropriate judicial utilization of research evidence (see Brigham, 1983; Grano, 1984; Loftus, 1979; 1980; 1984; Sobel, 1972; Wells, 1984; Wells & Murray, 1982; Woocher, 1977; and Yarmey, 1979, for discussions of the issues involved). For example, although Judge Nathan Sobel (1972, p. vi) has asserted in his widely-respected book that incorrect eyewitness identifications account for more mis-carriages of justice than all other factors combined, many judges have declined to admit expert testimony on this issue, generally asserting either that research findings on eyewitness identifications do not constitute a "generally recognized, scientifically acceptable" research area within psychology or that knowledge about the factors affecting the accuracy of eyewitness evidence is already part of the "common knowledge" of most jury members. The research reviewed in this report strongly suggests that neither of those points reflects the actual state of affairs. Research on factors affecting the accuracy of eyewitness identifications contributes to the growing body of empirical knowledge in this area. This knowledge can be utilized by judges in deciding on the admissibility of disputed eyewitness evidence and instructing the jury on the evaluation of such evidence; it also can be communicated directly to jury members by researchers serving as expert witnesses in cases where the judge decides that such testimony is relevant.

References

Bacon, S. J. (1974). Arousal and the range of cue utilization. Journal of Experimental Psychology, 102(1), 81-87.

Bahrick, H. P., Fitts, P. M., & Rankin, R. E. (1952). Effects of incentives upon reactions to peripheral stimuli. Journal of Experimental Psychology, 44, 404-406.

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, March). Eyewilness news. Psychology Today, p. 102 ff.

Bem, D. J. (1972). Self-perception theory. In L. Berkowitz (Ed.), Advances in experimental social psychology (Vol. 6). New York: Academic Press.

Berlyne, D. E., Borsa, D. M., Craw, M. A., Gelman, R. S., & Mandell, E. E. (1965). Effects of stimulus complexity and induced arousal on paired-associate learning. Journal of Verbal Learning and Verbal Behavior, 4, 291-299.

Berlyne, D. E., & Lewis, J. L. (1963). Effects of heightened arousal on human exploratory behavior. Canadian Journal of Psychology, 17, 398-410.

Boggs, D. H., & Simon, J. R. (1968). Differential effect of noise on tasks of varying complexity. Journal of Applied Psychology, 52, 148-153.

Bothwell, R. S., Brigham, J. C., & Deffenbacher, K. (1984). A meta-analysis of the impact of arousal and confidence on eyewitness accuracy. Unpublished manuscript, Florida State University.

Bothwell, R. K., Brigham, J. C., & Malpass, R. S. (1984). A meta-analysis of eyewitness research: Cross-racial identifications. Unpublished manuscript, Florida State University.

Bower, G. M., & Karlin, M. B. (1974). Depth of processing pictures of faces and recognition memory. Journal of Experimental Psychology, 103, 751-575.

Brehm, J. W., & Cohen, A. R. (1962). Explorations in cognitive dissonance. New York: Wiley.

Brigham, J. C. (1983). Psychological factors in eyewitness identifications. Journal of Criminal Justice, 11, 47-56.

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

Brigham, J. C., & Barkowitz, P (1978). Do "they all look alike?" The effect of race, sex, experience, and attitudes on the ability to recognize faces. Journal of Applied Social Psychology, 8, 306-318.

Brigham, J. C., & Bothwell, R. K. (1983). The ability of prospective jurors to estimate the accuracy of eyewitness identifications. Law and Human Behavior, 7, 19-30.

Brigham, J. C., Maass, A., Martinez, D., & Whittenberger, G. (1983). The effect of arousal on facial recognition. Basic and Aplied Social Psychology, 4, 279-293.

Brigham, J. C., Maass, A., Snyder, L. D., & Spaulding, K. (1982). Accuracy of eyewitness identifications in a field setting. Journal of Personality and Social Psycholohgy, 42, 673-681.

Brigham, J. C., & Williamson, N. L. (1979). Cross-racial recognition and age: When you're over 60, do they still "all look alike?" Personality and Social Psychology Bulletin, 5(2), 218-222.

Brigham, J. C., & WolfsKeil, M. P. (1983). Comparison of description and

identification accuracy under different attentional levels. Unpublished manuscript, Florida State Universtiy.

Broadbent, D. E. (1957). Effects of noises of high and low frequency on behavior. Ergonomics, 1, 21-29.

Broadbent, D. E. (1958). Perception and communication. London: Pergamon Press.

Brown, E., Deffenbacher, K., & Sturgill, W. (1977). Memory for faces and the circumstances of encounter. Journal of Applied Psychology, 62, 311-318.

Carlsmith, J. M., Ellsworth, P. C., & Aronson, E. (1976). Methods of research in social psychology. Reading, MA: Addison-Wesley.

Chance, J., & Goldstein, A. (1976). Recognition of faces and verbal labels. Bulletin of the Psychonomic Society, 7, 384-387.

Chance, J., Turner, A. L., & Goldstein, A. (1982). Development of differential recognition for own- and other-race faces. Journal of Psychology, 112, 29-37.

Clifford, B. R., & Bull, R. (1978). The psychology of person identification. Boston: Routledge and Kegan Paul.

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

Cross, J. F., Cross, J., & Daley, J. (1971). Sex, race, age, and beauty as factors in recognition of faces. Perception and Psychophysics, 10, 393-396.

Davies, D. R. (1968). Physiological and psychological effects of exposure to high intensity noise. Applied Acoustics, 1, 215-233.

Deffenbacher, K. (1980). Eyewitness accuracy and confidence. Law and Human Behavior, 4, 243-260.

Deffenbacher, K. (1983). The influence of arousal on reliability of testimony. In S. M. A. Lloyd-Bostock & B. R. Clifford (Eds.), Evaluating reearch evidence: Recent psychological research and new perspectives (pp. 235-254). New York: Wiley.

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

Elliott, P. B. (1964). Tables of d'. In J. A. Swets (Ed.), Signal detection and recognition in human observers. New York: Wiley.

Einhorn, H. J., & Hogarth, R. M. (1978). Confidence in judgment: Persistence of the illusion of validity. Psychological Review, 85, 395-416.

Eysenck, M. W. (1975). Effects of noise, activation level, and response dominance on retrieval from semantic memory. Journal of Experimental Psychology: Human Learning and Memory, 104, 143-148.

Flexser, A. J., & Tulving, E. (1978). Retrieval independence in recognition and recall. Psychological Review, 85(3), 153-171.

Frank, J., & Frank, B. (1957). Not guilty. New York: Doubleday.

Gilbert vs. California (388 U.S. 263, 87 S. Ct. 1951, 18 L. Ed. 2d 1178, 1967).

Goldstein, A. G., & Chance, J. (1971). Visual recognition memory for complex configurations. Perception and Psychophysics, 9, 237-241.

Goldstein, A. g., Johnson, K. S., & Chance, J. (1979). Does fluency of face description imply superior face recognition? Bulletin of the Psychonomic Society, 13(1), 15-18.

Goldstein, A. G., Stephenson, B., & Chance, J. (1977). Face recognition memory: Distribution of false alarms. Bulletin of the Psychonomic

Society, 9(6), 416-418.

Grano, J. D. (1984). A legal response to the inherent dangers of eyewitness identification testimony. In G. L. Wells & E. F. Loftus (Eds.), Eyewitness testimony: Psychological Perspectives (pp. 315-335). Cambridge: Cambridge University Press.

Hockey, G. R. J. (1969). Noise and efficiency: The visual task. New Scientist, 42, 244-246.

Hockey, G. R. J. (1970). Effect of loud noise on attentional selectivity. Quarterly Journal of Experimental Psychology, 22, 28-36.

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

Houston, B. K. 1968). Inhibition and the facilitating effect of noise on interference tasks. Perceptual and Motor Skills, 27, 947-950.

Houts, M. (1956). From evidence to proof. Springfield, Illinois: Charles C. Thomas.

Hunter, J. E., Schmidt, F. L., & Jackson, G. B. (1982). Meta-analysis: Cumulating research across studies. Beverely Hills, California: Sage.

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 American Psychological Association meetings.

Keisler, C. A. (1971). The psychology of commitment: Experiments linking behavior to belief. New York: Academic Press.

Kirby vs. Illinois (406 U.S. 682, 925, Ct. 1877, 32 L. Ed. 2d 411, 1982).

Leippe, M. R. (1980). Effects of integrative memorial and cognitive processes on the correspondence of eyewitness accuracy and confidence. Law and Human Behavior, 4, 261-274.

Leippe, M. R., Wells, G. L., & Ostrom, T. M. (1978). Crime seriousness as a determinant of accuracy in eyewitness identification. Journal of Applied Psychology, 63, 345-351.

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. (1979). Eyewitness testimony. Cambridge, Massachusetts: Harvard University Press.

Loftus, E. F. (1980). Impact of expert psychological testimony on the unreliability of eyewitness identification. Journal of Applied Psychology, 65, 9-15.

Loftus, E. F. (1984). Expert testimony on the eyewitness. In G. L. Wells & E. F. Loftus (Eds.), Eyewitness testimony: Psychological perspectives (pp. 273-282). Cambridge: Cambridge University Press.

Maass, A., & Brigham, J. C. (1982). Eyewitness identifications: The role of encoding specificity and attention. Personality and Social Psychology Bulletin, 8, 54-59.

Malmo, R. B. (1965). Physiological gradients and behavior. Psychological Bulletin, 64, 225-234.

Malmo, R. B., & Surwillo, W. W. (1960). Sleep deprivation: Changes in performance and physiological indicants of activation. Psychological Monographs, 74, 1-24.

Malpass, R. S. (1981). Effective size and defendant bias in eyewitness identification lineups. Law and Human Behavior, 5, 229-309.

Malpass, R. S., & Devine, P. G. (1980). Realism and eyewitness identification research. Law and Human Behavior, 4, 347-358.

Malpass, R. S., & Devine, P. G. (1981a). Eyewitness identification: Lineup instructions and the absence of the offender. Journal of Applied Psychology, 86, 343-350.

Malpass, R. S., & Devine, P. G. (1981b). Guided memory in eyewitness identification. Journal of Applied Psychology, 66, 343-350.

Malpass, R. S., & Kravitz, J. (1969). Recognition for faces of own and other race. Journal of Personality and Social Psychology, 13, 330-334.

Malpass, R. S., Lavigueur, H., & Weldon, D. E. (1973). Verbal and visual training in face recognition. Perception and Psychophysics, 14, 285-292.

Manson vs. Brathwaite (432 U.S. 98, 97 S. Ct. 2243, 53 L. Ed. 2d 140, 1977).

McNemar, Q. (1969). Psychological statistics (Fourth ed.). New York: Wiley.

Murray, D. M., & Wells, G. L. (1982). Does knowledge that a crime was staged effect eyewitness performance? Journal of Applied Social Psychology, 12, 42-53.

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

Schwartz, S. (1974). Arousal and recall: Effects of noise on two retrieval strategies. Journal of Experimental Psychology, 102(5), 896-898.

Shepherd, J. W. (1983). Identification after long delays. In S. M. A. Lloyd-Bostock & B. R. Clifford (Eds.), Evaluating witness evidence: Recent psychological research and new prospectives (pp. 173-188). New York: Wiley.

Snyder, M. (1979). Self-monitoring processes. In L. Berkowitz (Ed.), Advances in experimental social psychology (Vol. 12). New York:

Academic Press, Inc.

Snyder, M. (1974). The self-monitoring of expressive behavior. Journal of Personality and Social Psychology, 30, 56-537.

Sobel, N. R. (1979). 1980 supplement to Eyewitness identification: Legal and practical problems. New York: Clark Boardman.

Sobel, N. R. (1972). Eyewitness identification: Legal and practical problems. New York: Clark Boardman.

Stovall vs. Denno (388 U.S. 293, 87 S. Ct. 1967, 18 L. Ed. 2d 1199, 1967).

Tulving, E., & Thomson, D. M. (1973). Encoding specificity and retrieval processes in episodic memory. Psychological Review, 80, 352-373.

United States vs. Ash (413 U.S. 300, 93 S. Ct. 2568, 37 L. Ed. 2d 619, 1973).

United States vs. Wade (388 U.S. 218, 87 S. Ct. 1926, 18 L. Ed. 2d 1149, 1967).

Vinacke, W. E. (1968). Foundations of psychology. New York: Van Nostrand Reinhold.

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

Wells, G. L. (1984). How adequate is human intuition for judging eyewitness estimony? In G. L. Wells & E. F. Loftus (Eds.), Eyewitness testimony: Psychological perspectives (pp. 256-272). Cambridge: Cambridge University Press.

Wells, G. L., Ferguson, T. J., & Lindsay, R. C. L., (1981). The tractability of eyewitness confidence and its implications for triers of fact. Journal of Applied Psychology, 66, 668-696.

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

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., & 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 (pp. 155-170). Cambridge: Cambridge University Press.

Whipple, G. M. (1909). The observer as reporter: A survey of the psychology of testimony. Psychological Bulletin, 6, 153-170.

Whipple, G. M. (1912). Psychology of testimony and report. Psychological Bulletin, 9, 264-269.

Whipple, G. M. (1915). Psychology of testimony. Psychological Bulletin, 12, 221-224.

Whipple, G. M. (1918). The obtaining of information: Psychology of observation and report. Psychological Bulletin, 15, 217-248.

Wilkinson, R. T. (1961). Interaction of lack of sleep with knowledge of results, repeated testing, and individual differences. Journal of Experimental Psychology, 62, 263-271.

Winer, B. J. (1971). Statistical principles in experimental design. New York: McGraw-Hill.

Winograd, E. (1981). Elaboration and distinctiveness in memory for faces. Journal of Experimental Psychology: Human Learning and Memory, 7(3), 181-190.

Winograd, E. (1978). Ecoding operations which facilitate memory for faces

across the life span. In M. M. Gruenberg, P. E. Morris, & R. N. Sikes (Eds.), Practical aspects of memory. New York: Academic Press.

Winograd, E. (1976). Recognition memory for faces following nine different judgements. Bulletin of the Psychonomic Society, 8(6), 419-421.

WolfsKeil, M. P. (1983). The role of encoding specificity in eyewitness identifications. Unpublished manuscript, Florida State University.

WolfsKeil, M. P., & Brigham, J. C. (1984). The relationship between accuracy of prior description and facial recognition. Unpublished manuscript, Florida State University.

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

Yarmey, A. D. (1979). The psychology of eyewitness testimony. New York: Free Press.

Footnotes

[1]Results here will be reported in general terms. Detailed descriptions of the statistical analyses and outcomes are available from the author on request. For all statistical tests, the .05 level of significance, traditional in psychological research, was used. This means that for all analyses reported "significant" in this paper, there is a less than 5% probability that a difference this great, or a relationship this strong, could have occurred by chance alone.

[2]The correlation coefficient measures the degree of relationship between two variables. Correlations can only range from -1.00 to (+) 1.00; correlations near 1.00 denote strong positive relations, correlations from about .30 to .60 describe moderate positive relationships, and correlations near zero mean that there was no relationship between the two variables.

[3]Negative correlations describe inverse relationships, that is, high scores on one variable are associated with low scores on the other variable.

[4]Since the people in the distractor photos had not been seen before, with or without weapons, there was no logical basis for deciding whether a given distractor photo should be placed in the "weapon" or "nonweapon" category for the purpose of calculating d'scores. To address this problem, we used all of the same-race distractor photos for each race, so that the possible ranges were 0 to 6 hits (correctly identifying a person seen before) and 0 to 24 false alarms (incorrectly identifying someone who had not been seen before). A second, alternative method of calculating d' was also developed. Results of analyses of this second measure are available on request from the author.

[5]Copies of the checklist and the 4 scoring procedures are available on

request from the author.

[6]Functional (FS) and effective size (ES) calculations are ways to estimate the fairness or bias inherent in a particular lineup. In brief, ideally a 6-person lineup should have a FS or ES of near 6.00. FS or ES much smaller indicate other lineup members are not very similar in general appearance to the target (suspect) and hence the lineup is biased toward the target. The analyses in this study indicated that both lineups were relatively fair. Details on the procedures for calculation of FS and ES can be found in Wells, Leippe, and Ostrom (1978) and Malpass (1981).

Table 1

Precentages of Eyewitnesses Making Correct and Incorrect Choices from Criminal-Present and Criminal-Absent Photograph Lineups

		Criminal-Present Lineup				Criminal-Absent Lineup		
		N	Chooses criminal (hit)	Chooses foil (false alarm)	No choice (miss)	N	No Choice (correct rejection)	Chooses foil (false alarm)
conf. meas. before	neg. bias	12	33%	42%	25%	11	36%	64%
	no bias	11	73%	27%	0%	11	36%	64%
	pos. bias	13	46%	23%	23%	13	38%	62%
conf. meas. after	neg. bias	11	82%	8%	15%	12	50%	50%
	no bias	12	58%	25%	17%	13	46%	54%
	pos. bias	13	77%	8%	15%	12	75%	25%
across conditions		72	61.1%	20.8%	16.7%	72	47.2%	52.8%

Table 2

The Effects of the Timing of the Confidence Assessment on Confidence and Accuracy

	Confidence Assessed Before	Confidence Assessed After
Team 1		
N	28	27
Accuracy	$\overline{X}$ = .61, s = .49	$\overline{X}$ = .63, s = .49
Confidence	$\overline{X}$ = 4.5, s = 1.26	$\overline{X}$ = 5.52, s = 1.34
Correlation	r (26) = .38*	r (25) = .36+
Team 2		
N	13	19
Accuracy	$\overline{X}$ = .46, s = .51	$\overline{X}$ = .74, s = .45
Confidence	$\overline{X}$ = 3.62, s = 1.8	$\overline{X}$ = 4.63, s = 1.73
Correlation	r (11) = .38	r (17) = .36
Team 3		
N	29	28
Accuracy	$\overline{X}$ = .28, s = .45	$\overline{X}$ = .54, s = .5
Confidence	$\overline{X}$ = 3.4, s = 1.42	$\overline{X}$ = 4.39, s = 1.57
Correlation	r (27) = .18	r (26) = .33+

Note: Significance tests on point-biserial correlations were carried out via *t*-tests as recommended by McNemar (1969).

+p < .10
*p < .05

$\overline{X}$ = Mean
s = Standard Deviation

Table 3

The Effect of Choosing on Confidence and Accuracy

Team 1

	Choosers	Nonchoosers
N	36	19
Accuracy	$\overline{X}$ = .58, s = .5.	$\overline{X}$ = .68, s = .48
Confidence	$\overline{X}$ = 4.58, s = 1.48	$\overline{X}$ = 5.95, s = 1.07
Correlation	r (34) = .65***	r (17) = .14

Team 2

	Choosers	Nonchoosers
N	20	12
Accuracy	$\overline{X}$ = .5, s = .51	$\overline{X}$ = .83, s = .39
Confidence	$\overline{X}$ = 3.75, s = 1.77	$\overline{X}$ = 4.75, s = 1.86
Correlation	r (18) = .67**	r (10) = .31

Team 3

	Choosers	Nonchoosers
N	41	16
Accuracy	$\overline{X}$ = .32, s = .47	$\overline{X}$ = .62, s = .5
Confidence	$\overline{X}$ = 4.02, s = 1.62	$\overline{X}$ = 4.06, s = 1.98
Correlation	r (39) = .35*	r (14) = .04

Note: Significance tests on point-biserial correlations were carried out via t-tests as recommended by McNemar (1969).

*p < .05
**p < .01
***p < .001

$\overline{X}$ = Mean
s = Standard Deviation

Table 4

Point-Biserial Correlations Between Confidence and Accuracy for Subjects Who Picked Someone From the Photograph Lineup

Confidence Assessment	Instruction Bias: Neg. Bias	Unbiased	Pos. Bias	Across All 3 Conditions
Before	.57*	.54*	.34	.48***
	n = 16	n = 18	n = 17	n = 51
After	.74**	.35	.48+	.48***
	n = 12	n = 17	n = 17	n = 46
Before and After Combined	.68***	.43**	.39*	.50***
	n = 28	n = 35	n = 34	n = 97

Note: Significance tests on point-biserial correlations were carried out via t-tests as recommended by McNemar (1969)

+p < .10
*p < .05
**p < .01
***p < .001

Table 5

Accuracy of Recognition According to Level of Arousal, Race of Target, and Focus of Attention: Mean Hit Stores Within Each Condition

	Arousal							
	High				Low			
	Race of Target				Race of Target			
Focus	Black		White		Black		White	
	Wp	NWp	Wp	NWp	Wp	NWp	Wp	NWp
Face	4.25	4.00	1.83	3.43	3.58	3.25	2.83	3.08
Hand	3.58	3.33	2.00	2.25	2.75	2.92	1.25	1.67
Background	2.67	3.33	2.17	2.33	2.58	2.83	1.83	2.00
Free	3.25	3.17	1.92	3.33	2.33	2.42	2.25	2.25
Mean	3.44	3.46	1.98	2.83	2.81	2.86	2.04	2.25

Wp = weapon

NWp = nonweapon

Possible range of 0-6 hits per cell.

Table 6

Accuracy of Recognition According to Level of Arousal, Race of Target, and Focus of Attention: Mean D-Prime Scores Within Each Condition

	Arousal							
	High				Low			
	Race of Target				Race of Target			
Focus	Black		White		Black		White	
	Wp	NWp	Wp	NWp	Wp	NWp	Wp	NWp
Face	0.82	0.71	0.22	0.11	0.60	0.60	0.61	0.80
Hand	0.50	0.38	0.28	0.47	0.02	0.01	-0.09*	0.17
Background	-0.13*	0.25	0.47	0.46	-0.04*	0.22	0.46	0.51
Free	0.22	0.18	0.38	0.10	0.04	0.07	0.54	0.55
Mean	0.35	0.38	0.34	0.28	0.16	0.22	0.38	0.51

*These cells are slightly below chance levels of responding.

Table 7

Identification Accuracy Rates

	Target A Attention		Target B Attention		Mean Accuracy for Lineups
	Low	High	Low	High	
Target-Present Lineup	86.7%	93.3%	53.3%	46.7%	70.0%
Target-Absent Lineup	66.7%	73.3%	60.0%	86.7%	71.7%
Mean Accuracy for Targets	80.0%		61.7%		
Mean Accuracy for Attention	Low: 66.8%		High: 75.0%		
Overall Accuracy:	70.8%				

N = 15 per cell (Total N = 120)

Table 8

Confidence-Identification Accuracy Correlations for All Witnesses Choosing Someone From the Lineup, Target-Present and Target-Absent Conditions Combined

	Attention		
	Low	High	Combined Attention
Target A	$r = .233_a$	$r = 468*_b$	$r = .329*$
Target B	$r = .603**_c$	$r = .893***_d$	$r = .716***$
Combined Targets	$r = .406**$	$r = .637***$	

Overall $r = .512**$

$^a N = 19$

$^b N = 18$

$^c N = 16$

$^d N = 9$

* $p < .05$

** $p < .01$

*** $p < .001$