NOTICE

The quality of this microform is heavily dependent upon the quality of the original thesis submitted for microfilming. Every effort has been made to ensure the highest quality of reproduction possible.

If pages are missing, contact the university which granted the degree.

Some pages may have indistinct print especially if the original pages were typed with a poor typewriter ribbon or if the university sent us an inferior photocopy.

Previously copyrighted materials (journal articles, published tests, etc.) are not filmed.

Reproduction in full or in part of this microform is governed by the Canadian Copyright Act, R.S.C. 1970, c. C-30.

AVIS

La qualité de cette microforme dépend grandement de la qualité de la thèse soumise au microfilmage. Nous avons tout fait pour assurer une qualité supérieure de reproduction.

S'il manque des pages, veuillez communiquer avec l'université qui a conféré le grade.

La qualité d'impression de certaines pages peut laisser à désirer, surtout si les pages originales ont été dactylographiées à l'aide d'un ruban usé ou si l'université nous a fait parvenir une photocopie de qualité inférieure.

Les documents qui font déjà l'objet d'un droit d'auteur (articles de revue, tests publiés, etc.) ne sont pas microfilmés.

La reproduction, même partielle, de cette microforme est soumise à la Loi canadienne sur le droit d'auteur, SRC 1970, c. C-30.

Canada

THE UNIVERSITY OF ALBERTA

EYEWITNESS LINEUP BEHAVIOR AND IDENTIFICATION-ACCURACY ASSESSMENT:

CAN EYEWITNESS CONFIDENCE CALIBRATION AND JURORS' ACCURACY ASSESSMENT BE IMPROVED BY EXPOSURE TO THE IDENTIFICATION TASK?

BY

JOHN W. TURTLE

A THESIS

SUBMITTED TO THE FACULTY OF GRADUATE STUDIES AND RESEARCH

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE

OF DOCTOR OF PHILOSOPHY

DEPARTMENT OF PSYCHOLOGY

EDMONTON, ALBERTA

FALL, 1988

ISBN 0-315-45658-2

THE UNIVERSITY OF ALBERTA

RELEASE FORM

NAME OF AUTHOR: John Warren Turtle

TITLE OF THESIS: Eyewitness Lineup Behavior and Identification-Accuracy Assessment: Can Eyewitness Confidence Calibration and Jurors' Accuracy Aseesment be Improved by Exposure to the Lineup Task?

DEGREE: Doctor of Philosophy

YEAR THIS DEGREE GRANTED: 1988

.....

(Student's signature)

#3, 1310 Cypress Street
Vancouver, B.C. V6J 3L2
(Student's permanent address)

Date: August 24, 1988

THE UNIVERSITY OF ALBERTA

FACULTY OF GRADUATE STUDIES AND RESEARCH

The undersigned certify that they have read, and recommend to the Faculty of Graduate Studies and Research for acceptance, a thesis entitled "Eyewitness lineup behavior and identification-accuracy assessment: Can witness confidence calibration and jurors' accuracy-assessment be improved by exposure to the lineup task?" submitted by John W. Turtle in partial fulfillment of the requirements for the degree of Doctor of Philosophy.

.....
(Supervisor)

.....

.....

.....

.....

Date: August 18, 1988

ABSTRACT

These studies investigated several hypotheses regarding the relationship between lineup behavior, confidence, and accuracy to see what aspects of lineup behavior may be related to accuracy and to see if the probative value of identification testimony could be improved. Some participants in Study 1 were allowed to see a videotape of their previous identification decision regarding the suspect in a staged, videotaped crime. Some saw either a full-color videotape or only a silhouette, some were instructed to simulate the task, and others received no instuction. Witnesses who saw either a normal or silhouette videotape demonstrated a significant relation between their self-reports of confidence and actual accuracy, whereas witnesses in the other two conditions did not. Some participants in Study 2 were allowed to see witnesses' lineup behaviors from one of the conditions in Study 1. The correlation between these observers' ascribed confidence and actual witnesses accuracy was signficant only for observers who saw the silhouette version. This finding, together with the Study 1 results, was interpreted as evidence for the notion that cues to identification accuracy may be contained in the tone of witnesses' oral utterances. It was predicted in Study 2 that jurors' abilities to distinguish accurate testimony from inaccurate testimony would be better for witnesses who showed a signficant relation between accuracy and confidence and for jurors who were allowed to view videotapes of the witnesses' identification behavior. Accuracy-assessment was not better in either of these conditions compared to control groups. Participants in Study 3 were given false feedback indicating they took a long or short time relative to other eyewitnesses to make a lineup decision. Short time/perpetrator-present lineup condition witnesses were more confident than witnesses in all other conditions; those in the Short time/absent lineup condition were less confident than those in all other conditions. It is concluded that witnesses can infer their confidence from their identification behaviors, but video replays of those behaviors for witnesses and jurors might not be of much forensic use.

ACKNOWLEDGMENT

I wish to thank my supervisor, Gary Wells, for his constant support, for treating me as a friend and colleague, and for the hundreds of hours he spent sharing his wisdom with me. I was fortunate to be under Gary's supervision and I am grateful to him for opening the door to a successful career. In addition, I owe a lot to Mike Enzle, whose advice, instruction, and comments over the past four years added greatly to my graduate training.

I also wish to thank Curt Hoffman, Jim Hackler, and Mike Leippe for their diligent reading of my dissertation and for the provocative questions they brought up during the defense. Finally, Wendy Wintoniuk deserves my appreciation for collecting most of the data in Study 2.

TABLE OF CONTENTS

CHAPTER		PAGE
I.	INTRODUCTION	1
	Nature of the Problem	4
	The Confidence/Accuracy Relation in Eyewitness Research	7
	Confidence/Accuracy Analogues in Other Areas of Research	11
	What Aspects of the Identification Task Might Predict Accuracy?	17
	Preview of the Present Research	20
II.	STUDY 1	27
	Method	30
	Results and Discussion	35
III.	STUDY 2	44
	Method	46
	Results and Discussion	48
IV.	STUDY 3	55
	Method	56
	Results and Discussion	59
V.	GENERAL DISCUSSION	64
REFERENCES		71
APPENDIX 1		92
APPENDIX 2		94
APPENDIX 3		95
APPENDIX 4		96
APPENDIX 5		97
APPENDIX 6		98
APPENDIX 7		99
APPENDIX 8		100
APPENDIX 9		101
APPENDIX 10		102

LIST OF TABLES

TABLE	Description	PAGE
1	Breakdown of Lineup Decision Data for Study 1	81
2	Lineup Type, Gender, Choosing Rates, Accuracy, and Decision Time Data for Study 1 Broken Down by RSA Condition	82
3	Confidence, Probability, Difficulty, and Estimated Decision Time Data for Study 1 Broken Down by RSA Condition	83
4	Witnesses' Self-Reported Confidence and Estimated Probability of Accuracy in Study 1 as a Function of RSA Condition and Actual Identification Accuracy	84
5	Mean Lineup Decision Times for Witnesses in Study 1	85
6	Correlations Between Decision Times and Confidence and Accuracy for Witnesses in Study 1	86
7	Correlations Between Jurors' Ascriptions of Witness Confidence and Actual Witness Accuracy on the Basis of Lineup Behavior Only as a Function of Video Type in Study 2	87
8	Juror Accuracy and RAtings of Witness Confidence as a Function of Juror Condition in Study 2 and Witness RSA Condition in Study 1	88
9	Witnesses' Estimated Decision Time in Study 3 as a Function of Decision-Time Feedback and the Presence or Absence of the Perpetrator in the Lineup	89

LIST OF TABLES (Continued)

TABLE	Description	PAGE
10	Confidence Ratings for Witnesses in Study 3 as a Function of Feedback Condition and the Presence or Absence of the Perpetrator	90
11	Mean Decision Times for Witnesses in Study 3 as a Function of Decision Accuracy and the Presence or Absence or the Perpetrator	91

I. INTRODUCTION

The ability of witnesses to a crime or accident to recall accurately the details of what transpired is an integral part of the judicial system. More generally, however, the investigation of social and cognitive factors that influence the accuracy of eyewitness recollections is important in that it reveals something about the nature of how people encode, store, and retrieve information. Psychological researchers have long been aware of the opportunity the eyewitnessing situation presents as a naturally occurring instance of the complex human information-processing system at work.

At the turn of the century, prominent researchers from both Europe and North America began investigating the accuracy of eyewitness testimony. Foremost among the American investigators was Hugo Munsterberg (1908), whose book *On the Witness Stand* is often cited as the beginning of a literature concerned with the empirical study of people's eyewitnessing capabilities. Subsequent to Munsterberg's book, Guy Montrose Whipple (e.g., 1912, 1918) published a series of *Psychological Bulletin* articles discussing the advancement of eyewitness research. In Europe, Germany's Wilhem Stern (e.g., 1904, 1939) was conducting "realistic" eyewitness experiments. The gist of this early research is often characterized by the conclusion that eyewitness testimony generally is unreliable (e.g., Buckout, 1976; Loftus, 1984). Perhaps it was this simple, somewhat pessimistic view of the eyewitness that led to the relative demise of eyewitness research until the mid 1970s.

Robert Buckout's (1974) *Scientific American* article regarding eyewitness testimony marks the beginning of the modern era of eyewitness research. Since then, hundreds of journal articles and numerous books devoted entirely to eyewitness research have appeared, and the area continues to enjoy popular support among both academic psychologists and

the legal community. The tone of eyewitness research has evolved from the view that eyewitnesses are unreliable to the investigation of factors responsible for inaccuracy when it occurs, and procedures that may reduce error and increase accuracy. Consistent with this view, the present series of studies were designed to assess the potential of several factors that might bolster the relation between eyewitness accuracy and confidence, which in turn was expected to improve the ability of mock-jurors to identify accurate testimony.

The rationale and hypotheses that form the basis of the present studies stem from a social-psychological perspective. Generally, a juror's decision to believe or discount an eyewitness's testimony is viewed as a complex process that is influenced by a number of factors, many of which are not related to eyewitness accuracy and many of which the juror is not aware. Specifically, the present studies were aimed at determining the process(es) by which mock witnesses and some mock jurors in a previous series of studies (Kassin, 1985) were able to improve their judgments of identification confidence after viewing videotapes of the identification task. In some ways, the process by which people judge the accuracy of their own or another person's testimony is an attribution issue. Just as social perceivers in other situations must infer an actor's motive for a particular behavior based on cues like consistency, distinctiveness, and consensus information, a juror must assess a witness's identification accuracy based on some relevant set of cues. There are several plausible candidates for these cues (e.g., facial expression, voice intonation, decision time, apparent "confidence"), each of which is discussed and then incorporated into one of three studies reported.

The present research has the potential to be of significant practical import for the police and the judicial system. The manipulated factors involve procedures that could be implemented by the police during the initial stages of the investigatory process and then utilized by the courts during the judicial phase of a case. Foremost among these procedures is the idea of videotaping witnesses while they make an identification of an alleged

perpetrator and then allowing jurors the opportunity to view this tape either by itself or in addition to the witness's cross-examination testimony. It should be noted that the courts are becoming more influenced by the recommendations of some eyewitness investigators because they are allowed to provide expert testimony regarding eyewitness factors in an increasing number of cases. In turn, the police are becoming more aware of eyewitness-procedure recommendations because they must anticipate possible refutations of impoperly obtained eyewitness testimony in court. Some aspects of the procedures investigated in the present studies, therefore, have been designed with the possibilty of their application to the courtroom and to police procedures in mind.

The potential application of the procedures used in the current studies reflects what Wells (1978) termed a system-variable approach to eyewitness research. Wells distinguished system variables from what he termed estimator variables in eyewitness research. System variables are those factors that are under at least some control of the justice system and that may affect a witness's testimony after he or she has witnessed an event. The construction of suspect lineups, interview techniques, the time between witnessing and providing a statement, and even courtroom procedures are all examples of system variables. Estimator variables are those factors over which the justice system has little, if any, control and of which the system can only estimate the influence. The lighting (and other viewing) conditions at the time of the incident, the duration of the exposure, the stress experienced by the witness during encoding, race differences between the witness and perpetrator are examples of estimator variables. Although of obvious importance to the reliability of testimony in some cases, Wells noted that system-variable research is more likely to be a fruitful pursuit for eyewitness researchers and the course of research over the past decade has borne out that prediction. [Note, for example, the utility of suspect-lineup research regarding such aspects as the number of lineup-members (e.g., Wells, Leippe, & Ostrom, 1979), their similarity to one another (e.g., Lindsay & Wells, 1980), the

diagnosticity of nonidentifications (Wells & Lindsay, 1980), the number of suspects relative to foils (Wells & Turtle, 1986), the phrasing of witness instructions (e.g., Malpass & Devine, 1981), the effect of misleading questions in interviews on subsequent testimony (e.g., Loftus, 1979; Loftus, Miller, & Burns, 1978), and so on.]

The Nature of the Problem

The impetus for this research is based largely on the gap that exists between assumptions made by the courts, police, and laypeople on the one hand and a fairly consistent series of psychological research findings on the other. The judicial process operates on the assumption that an eyewitness's expressed confidence in his or her testimony and/or suspect-identification is an appropriate cue for jurors to use in their assessments of eyewitness credibility. In a landmark case, Neil vs. Biggers (1972), the U.S. Supreme Court court ruled that the expressed confidence of an eyewitness should be among the five criteria recommended to juries as the bases on which to assess the credibility of eyewitness testimony. Among the other criteria were the witness's opportunity to view the incident, the amount of attention the witness paid to the event, the correspondence between the witness's initial description of the perpetrator and the appearance of the suspect chosen from the lineup, and the length of the interval between the event and the identification (see Wells & Murray, 1983, for comment on these criteria). The police and attorneys also make the assumption that witness confidence is a useful cue to their witnesses' or clients' accuracy (Brigham & Wolfskeil, 1983). And finally, laypeople have reported that they believe there is a positive accuracy/confidence relation (Yarmey & Jones, 1983).

In contrast to the assumed relation between eyewitness accuracy and confidence, psychological researchers have for the most part failed to find a stable accuracy/confidence relation using standard laboratory procedures (e.g., Deffenbacher, 1980; Kassin, 1985;

Leippe, 1980; Wells, Ferguson, & Lindsay, 1981; Wells, Lindsay, & Ferguson, 1979; see Wells & Murray, 1984 for a review of 31 studies). Not only is there often no relation between accuracy and confidence overall, but some studies have demonstrated the lack of a correspondence between the two measures by supporting Leippe's (1980) argument that one can be manipulated independently of the other. For instance, Wells et al. (1981) succeeded in manipulating confidence by "briefing" half of their subject witnesses about what they should expect during their upcoming cross-examination. Consistent with Leippe's independence view, self-reports and mock-jurors' judgments of briefed witnesses' confidence were higher compared to nonbriefed witnesses, but without an accompanying increase in accuracy. In addition to the Wells et al. demonstration that confidence could be increased without an increase in accuracy, Lindsay et al. (1981) demonstrated that accuracy could be manipulated without a corresponding adjustment in people's judgments of confidence. In that study, witness accuracy was increased between subjects by affording some witnesses a "better" view of the to-be-remembered event than others. Despite reliable differences in accuracy between groups, neither witness self-reports nor mock-juror ratings of confidence revealed any effect of the initial-exposure manipulation. The significance of the Wells et al. and Lindsay et al. findings becomes even more apparent in light of research showing that jurors spontaneously utilize eyewitness confidence cues in their assessments of eyewitness credibility (Brigham & Bothwell, 1983; Lindsay, Wells, & Rumpel, 1981; Wells et al., 1979; Wells et al., 1981).

The problem to which the proposed studies are addressed, then, can be summarized as follows: Judges, lawyers, police, and laypeople hold a common belief that eyewitness confidence and eyewitness accuracy are positively related. Consistent with this belief, jurors tend to ascribe more credibility to more-confident witnesses compared to less-confident witnesses. In addition, the U.S. Supreme Court has ruled that confidence should be used by jurors to assess the credibility of eyewitness testimony. Recent

empirical findings, however, indicate that a witness's expressed confidence often is not a useful predictor of his or her testimony- and/or identification-accuracy.

The lack of a significant correlation between eyewitness confidence and actual eyewitness accuracy has sparked a number of research efforts aimed at identifying the factors that influence each of the two components and, in many cases, ways to improve the magnitude of the relation between them. [These studies are discussed in the next section.] Although the present studies are also concerned with the confidence/accuracy relation at the witness level, there is an additional emphasis here on the impact that a signficant confidence/accuracy correlation might have on mock-jurors' subsequent ability to distinguish accurate from inaccurate testimony. In [illegible] manipulation that improves the correlation between witnesses' self-rated confidence and actual accuracy may or may not improve jurors' abilities to detect accurate from inaccurate witnesses. Improving jurors' discrimination abilities is important, because just as the intuitively appealing confidence/accuracy relation has not been supported by research, previous studies that have examined jurors' abilities to discriminate accurate from inaccurate witnesses have shown that jurors did not perform above chance level (Lindsay, Wells, & Rumpel, 1981; Wells, Ferguson, & Lindsay, 1981; Wells, Lindsay, & Ferguson, 1979). From a judicial standpoint, then, the strong form of the argument is that the significance of factors that improve the accuracy/confidence relation within eyewitnesses rests ultimately on the degree to which the improved relation also improves jurors' judgments of eyewitness credibility. From an investigatory perspective, however, an improved confidence/accuracy relation at the witness level would be useful even if police could better allocate very limited investigative resources by feeling more confident themselves that a given witness's identification is more likely to be accurate.

To improve the accuracy of jurors' belief decisions, three links among four elements in the eyewitness process would have to hold. First, there would have to be a significant

correlation between accuracy and confidence at the witness level (r_{AW}). Second, the correlation between witness confidence and jurors' ascriptions of confidence (r_{WJ}) would have to be signficant and large. This value has been calculated in a number of previous eyewitness studies and was found to be signficant, although by no means perfect. Wells et al. (1979) reported a correlation of $r = .549$, $n = 42$, $p < .01$ and Wells et al. (1981) reported a correlation between the two of $r = .53$, $n = 34$, $p < .01$. And third, the correlation between jurors' ascriptions of confidence and their decision to believe the witness (r_{JB}) would also have to hold. This value has also been measured in previous studies and was also found to be signficant. Wells et al. (1979) reported an $r = .706$, $n = 42$, $p < .001$ and Wells et al. (1981) reported an $r = .58$, $n = 152$, $p < .001$. Given this series of significant correlations, it would be expected that jurors would believe accurate witnesses more than inaccurate ones. The correlations calculated in the studies mentioned above, however, are not extremely large and the multiplicative nature of the chain leaves much room for an imperfect transfer of confidence. Therefore, the impact of a manipulation that boosts the confidence of accurate witnesses and/or lowers the confidence of inaccurate witnesses would have to be very robust in order for its benefit to transfer to the accuracy of jurors' decisions.

The Confidence/Accuracy Relation in Eyewitness Research

What are "accuracy" and "confidence"?

The accuracy component of the confidence/accuracy relation in eyewitness research does not require much of an explanation. Most often, it is simply a measure of whether or not a witness's recollection of a briefly seen detail is correct. In the vast majority of eyewitness studies, accuracy refers to a witness's suspect-identification. Similarly, throughout this paper, accuracy refers to whether or not the person a participant-witness identifies from a series of very similar-looking pictures is in fact the perpetrator of a

simulated crime viewed by the witness on videotape. If the actual perpetrator is not among the pictures in the lineup, however, a witness's decision not to choose anyone from the pictures could also be considered accurate, a distinction that will be discussed in subsequent sections.

The confidence part of the relation can be viewed as witnesses's belief in the accuracy of their testimony. This belief is dependent on a number of factors that different witnesses in different situations may or may not use as a basis to judge their own confidence. One of these factors is witnesses' belief in their opportunity to view the to-be-remembered event. Witnesses who believe their opportunity to view was good (e.g., long exposure time, good lighting conditions, allocation of attention to the event, short duration between viewing and recollection) might express more confidence in their recollection compared to those who believe their opportunity was poor. Another factor comes from external sources that boost witness confidence without an accompanying increase in accuracy (e.g., the common practise of attorneys coaching their witnesses on how to behave under cross-examination). Another factor concerns people's beliefs regarding their personal abilities as eyewitnesses. Some people might believe they are especially perceptive and good candidates for an eyewitnessing task (e.g., police), whereas others are convinced of their poor memory for briefly-seen faces. Finally, and most relevant for the present research, people may make their judgments of confidence after reflecting on their own behavior during a recall or recognition task (e.g., suspect lineup).

Theory and research

The most general hypothesis bearing on the relation between eyewitness confidence and accuracy is the "optimality" hypothesis proffered by Deffenbacher (1980). According to Deffenbacher, the better the initial encoding conditions at the time of the to-be-remembered incident, the better the relation between accuracy and confidence. Despite the intuitive appeal of the optimality prediction, it has received little empirical support. Recall, for

instance, the Lindsay et al. (1981) study in which encoding situations designed to result in higher rates of eyewitness accuracy did not yield higher magnitudes of accuracy/confidence correlations.

Jack Brigham and his colleagues have reported findings consistent with their view that the confidence/accuracy relation should be greater in real-world settings as opposed to the laboratory situation (e.g., Brigham, Maass, Snyder, & Spaulding, 1982; see also Krafka & Penrod, 1985). They contrast the anxiety of making a lineup identification in an actual criminal investigation to the laboratory setting in which the participants know there are no meaningful ramifications of an inaccurate lineup choice. The Brigham et al. findings are inconsistent, however, with other research demonstrating that the accuracy/confidence relation is in fact greater when participant-witnesses know that their identification is not related to a real crime (Murray & Wells, 1982). Clearly, the role of laboratory- versus field-experiment factors requires more investigation.

A more recent study by Fleet, Brigham, and Bothwell (1987) suggests that the accuracy/confidence relation may be significant and quite large when measured only for that subset of witnesses who made a choice from the lineup. Fleet et al. reported an $r\ (94) = .50$, $p < .001$ for choosers compared to $r\ (44) = .14$, ns, for nonchoosers. Although the Fleet et al. findings are interesting, the advantage of measuring the accuracy/confidence relation only among choosers has failed to receive support in many other studies (see Bothwell, Deffenbacher, & Brigham, 1986 for a review of 11 studies). In addition, Wells and Lindsay (1985) have noted that there are a number of interesting and forensically relevant ways of subdividing samples for calculating the accuracy/confidence relation, so that reliance on a particular grouping is not always useful.

The research most relevant to the current studies is that by Kassin (1985). In a series of four experiments, Kassin demonstrated that the accuracy/confidence relation could be improved by allowing witnesses to view themselves on videotape making their

identification from a photo lineup before asking them to rate their confidence in that identification. Kassin called this opportunity for the witness to view their own identification behavior "retrospective self-awareness" (RSA). Averaged over the four experiments, Kassin reported an accuracy/confidence correlation of .48 (n = 102) for the RSA participants compared to .04 (n = 126) for control groups. Kassin also calculated accuracy/confidence correlations among the sub-sample of "choosers" (i.e., only those witnesses who identified someone from the lineup). Consistent with the results reported by Fleet et al. (1987) discussed earlier, Kassin found an appreciable increase in the accuracy/confidence relation among RSA witnesses in the chooser-only group compared to the overall sample. This increase was especially apparent for observers; when observers' confidence ratings of only choosers were considered, the accuracy/confidence relation jumped from .24 (n = 30, ns) to .41 (no n reported, $p < .05$). This increase suggests that the behavioral cues provided by witnesses who make a choice from the lineup are interpreted more clearly by observers compared to witnesses who do not make a choice. Kassin's series of studies will be discussed at length in a subsequent section.

At this point, it makes sense to ask why any differences between accurate and inaccurate witnesses should be expected at all, given the discouraging results of standard eyewitness studies aimed at investigating confidence differences as a function of accuracy. Of course the justice system depends on the existence of *some* difference between accurate and inaccurate witnesses, as was mentioned above, but perhaps this difference is characterized by some intangible, unmeasureable "feeling" on the part of witnesses that cannot be observed, described, or controlled even by psychological investigators whose research is dedicated to just these goals. The perspective taken in the present research, however, is at odds with this conclusion; in addition to the social psychological perspective guiding it, another strong influence on the present research is the notion that accurate and inaccurate memories are in some way(s) fundamentally different from one another.

The most directly relevant work regarding distinctions between accurate and inaccurate memories is that of Johnson and Raye (1981) and of Schooler, Gerhard, and Loftus (1986). Both of these groups of researchers have attempted to qualify the differences between people's real and imagined memories, as represented by the titles of their respective articles: "Reality Monitoring" and "Qualities of the Unreal". The basis of their work is that the way in which people describe memories of things that they actually experienced is distinguishable from the way in which they describe imagined, or even externally distorted, inaccurate memories. In addition, there is another vast area of somewhat less relevant research dealing with statements from people who may be purposely trying to provide inaccurate information. For example, Undeutsch (1982) and Yuille, Raskin, and Steller (in press) hold strongly to the idea that truthful statements are qualitatitvely different and discernable from deceptive ones. The paradigm common to all of these researchers, however, has been restricted to the evaluation of transcribed texts of people's memory descriptions and not applied to the evaluation of actual behavior. In that domain, the literature on nonverbal communication (e.g., DePaulo, 1985; Ekman, 1985) is also based on a similar idea that the behavior of deceptive people is different and discernable from that of sincere communicators. The basis of the present research is similar to all of this work, in that the studies described are concerned with identifying differences between the behavior of accurate and inaccuate identifiers during both the actual identification task and under subsequent cross-examination.

Confidence/Accuracy Analogues in Other Areas of Research

What does psychology have to say about how well a person's feeling of knowing is related to actual accuracy in domains outside the eyewitnessing situation? The answer appears to depend on which task-type, research domain, experimental procedure, and unit and level of analysis one chooses to investigate as an instance of the relation between

feeling of knowing, or comprehension, or confidence, and accuracy. Research findings on "feeling of knowing" (FOK), for instance, suggest that people are fairly good at predicting their ability to recognize a word in the future that they cannot recall in the present (e.g., Gruneberg, Monks, & Sykes, 1977; Hart, 1965, 1967; Nelson, 1984). Similarly, "tip-of-the-tongue" studies suggest that people are quite good at predicting their ability to recognize a word in the future based on how many aspects of the word they can recall in the present (e.g., Brown & McNeil, 1966). In contrast, "calibration of comprehension" studies (e.g., Glenberg & Epstein, 1985, 1987) suggest that people are quite poor at predicting their ability to answer a question in the future based on material they have just read. Similarly, "calibration of performance" studies (e.g., Lichtenstein, Fischhoff, & Phillips, 1982) suggest that for many tasks people are not very good at calibrating probability estimates of their own accuracy with actual performance. Each of these areas is discussed below to support the contention that the eyewitness identification task is a special case of calibrating confidence with accuracy.

First, research dealing with feelings of knowing (FOK) in general is indeed varied. Just researching the phrase "feeling of knowing" in the psychological literature may lead a reader to feel as if he or she has a good grasp on the relation between people's expressions of certainty and actual accuracy, when in fact there are numerous other relevant areas of research. The beginning of research on FOK per se can be traced to Hart (1965, 1967). In this early work, Hart used a recall-judgment-recognition procedure to assess people's ability to estimate the likelihood that they would later recognize a word that they could not recall at the moment. Participants were given the defintion of a familiar, though low-frequency, word and asked to recall the word. Those who could not recall it were asked to indicate whether or not they thought they could recognize the word later (i.e., "Do you *feel you know* what the word is?") Finally, recognition accuracy was measured. Hart's results showed that those people who indicated a FOK were significantly more

likely to recognize the word later compared to people who said they were unlikely to recognize it. He concluded that FOK as demonstrated in his study was suggestive of an efficient memory-monitoring process inherent in human judgment. Several noted theorists have discussed the importance of such a process in the overall operation of the human information-processing system (e.g., Tulving & Madigan, 1970; Underwood, 1969). In addition, other researchers have taken up from where Hart left off and developed varied and elaborate techniques for calculating FOK (e.g., Eysenck, 1979; Gruneberg, Monks, & Sykes, 1977; Nelson, 1984; Nelson & Narens, 1980, D.L. Schacter, 1983).

One year after Hart's (1965) FOK paper, Brown and McNeil (1966) published an article dealing with people's often reported feeling that they know a word, but can't recall it exactly at the moment. This investigation of the "tip-of-the-tongue phenomenon" used a slightly more sophisticated methodology than the FOK studies. Brown and McNeil again gave people a definition of a known, but uncommon word. In this case, however, they measured how many aspects of the word (e.g., number of syllables, inflection, first and last letters) people could remember and used that to predict subsequent recognition performance. The results showed that people were quite good at recalling a number of things about the word except the word itself, and that this recall was positively related to subsequent recognition. Brown and McNeil concluded that people have a dictionary-like representation for words. When asked to recall a particular entry, parts of relevant pages of the dictionary can be recalled so that similar looking and sounding words are available, but not necessarily the word itself. In a similar vein, Blake (1973) showed that the more information participants recalled about a previously seen trigram the more likely they were to indicate a strong FOK.

A number of studies have investigated people's ability to predict whether or not they can verify an inference in the future based on material they have recently learned. These "calibration of copmprehension" studies differ from the Hart (1965) and Brown and

McNeil (1966) studies in that a continuous confidence-judgment scale (e.g., 7-point, from "Not at all confident" to "Absolutely confident") was used instead of an FOK/no FOK response (Hart) or number of word attributes recalled (Brown & McNeil). For example, in studies by Glenberg and Epstein (1985, 1987) participants read a one-paragraph text about a topic (e.g., music, physics) and then rated their confidence in their ability to answer an upcoming question for which they would have to make an inference from the information given in the text. In both of the Glenberg and Epstein studies calibration of comprehension was extremely poor overall, but showed slight improvements with increased task familiarity. Despite rigorous attention to methodological protocol, these authors offer little in the way of a theoretical interpretation. Basically, they concluded that people's evaluation of their comprehension in certain domains is often inappropriate for assessing future ability to verfiy inferences based on their knowledge in that domain.

There are some important differences between the FOK and tip-of-the-tongue studies on the one hand and the calibration of comprehension studies on the other. The former were concerned primarily with memory monitoring or metamemory; that is, participants "confidence" ratings were based on some notion of how much they thought they remembered about the meaning of a word they had enountered before. In contrast, the calibration of comprehension studies were concerned with people's predictions of their upcoming performance at answering novel questions. The distinction is clear, but often treated lightly or ignored completely in discussions of the relevant findings. For example, Glenberg & Epstein (1987) included both types of studies in their "sampling of the literature" (p. 85) as if they were equivalent.

There are eyewitness studies that fall into, or are similar to, the FOK and "tongue" studies discussed above, in that all involve some kind of memory monitoring. They differ on two important points, however; (1) eyewitness experiments involve episodic memory whereas the others involve semantic memory, and (2) eyewitness experiments involve

visual, face memory and the others involve verbal memory. First, in a similar vein to FOK studies, some eyewitness experiments have asked witnesses to rate their confidence in their ability to recognize the criminal before the identification task (e.g., Cutler & Penrod, 1986; Fleet, Brigham, & Bothwell, 1987; Murray & Wells, 1982). In contrast to the FOK studies using words, these eyewitness studies did not show a significant pre-identification confidence/identification-accuracy relation. Second, along the lines of the "tip-of-the-tongue" studies, witnesses who provide fairly accurate descriptions of the criminal (i.e., recall) before the identification task do not show an concomitant improvement in subsequent recognition performance (Goldstein, Johnson, & Chance, 1979; Wolfskiel & Brigham, 1985). Again, this result is in contrast to the "tongue" studies, in which better initial recall was related to subsequent recognition performance.

The last area to be discussed, confidence calibration, is the most similar to the eyewitnessing situation. Within this area, the grouping of confidence calibration studies that are most similar to the eyewitness task deal with people's ratings of confidence in the accuracy of a question they have just answered. Findings in this area suggest that people are not randomly uncalibrated in assessing their past performance, but that they tend to be consistently overconfident in their judgments (Adams & Adams, 1961; Fischhoff, Slovic, & Lichtenstein, 1977; Koriat, Lichtenstein, & Fishhoff, 1980; Lichtenstein & Fischhoff, 1977). In these studies, participants were asked to rate their confidence in their answers to a large number of general knowledge questions. As Adams and Adams (1961) first pointed out, a perfectly calibrated judge should assign a particular confidence rating (p_i) to that grouping of answers which have a corresponding frequency of actually being correct (P_i). In other words, for all answers assigned a confidence rating of 7 on a 10-point scale, a perfectly calibrated judge would have answered 70% of the questions correctly. In contrast to this ideal case, people tend to ascribe confidence ratings to questions that exceed the probability of actually answering the questions correctly.

With respect to straightforward confidence calibration studies, witnesses to a staged or otherwise simulated crime have been asked to identify the perpetrator of that crime from a number of very similar photos and then rate their confidence in that choice. The common finding that eyewitness confidence is often unrelated to accuracy is consistent with many calibration of performance studies in that no significant calibration was found (e.g., Fischhoff et al., 1977; Koriat et al., 1980). One difference, however, is that the eyewitness studies have not investigated, nor discussed, whether or not eyewitnesses are generally overconfident in their identifications. Finally, the eyewitness studies stand in stark contrast to the Glenberg and Epstein (1985, 1987) studies; here people's confidence in the accuracy of a recently given answer to a general knowledge question *was* significantly related to the actual accuracy of the response.

Does the episodic-semantic distinction between the eyewitness studies on the one hand and the FOK, "tongue", and confidence calibration studies on the other hand account for these differences? Or that the former involve visual memory codes for faces whereas the latter involve verbal information? In general, both are probably true to some degree, but there are other specific differences between the eyewitnessing situation and each of these other domains that also may account for inconsistencies between them. First, with respect to the type of pre-recognition confidence ratings recorded in FOK studies, eyewitness pre-identification confidence ratings include people's reservations about the consequences of making a lineup choice that may influence the relationship between their ratings and their subsequent identification accuracy. Second, with respect to the amount of pre-recognition information people can provide in "tip of the tongue" studies, verbal descriptions of faces in the eyewitness case include piecemeal, surface judgments of physical appearance that often are not meaningful indicators of a witness's memory for the perpetrator. Several eyewitness studies have failed to find a relationship between witnesses' pre-identification descriptions of the perpetrator and their subsequent identification accuracy (e.g., Goldstein,

Johnson, & Chance, 1979; Wells, 1985; Wolfskiel & Brigham, 1985). Finally, with respect to the multiple-trial nature of confidence calibration studies, eyewitness identifications are a one-shot affair that preclude a traditional, within-subjects analysis of calibration. The current studies, however, do include a probability estimate of accuracy from witnesses that will provide some measure of general under- or overconfidence across subjects.

What Aspects of the Identification-Task Might Predict Accuracy?

The Kassin Studies in Detail

Recall Kassin's (1985) article in which witnesses', and some obervers', confidence assessments were improved by allowing them to view the identification task on videotape (i.e., the RSA manipulation). Kassin's explanation of the RSA effect is mixed. On the one hand, he speculated that the improved relation between confidence and accuracy may derive from inferences the witnesses make while watching their own observable decision-behavior on videotape. Witnesses may, for example, have inferred that the longer it took them to make a decision the less likely they were to be accurate. On the other hand, Kassin suggested that RSA may have worked because it allowed people to relive the private cognitions that they were experiencing during the identification task (a direct-access, retrieval-cue model, e.g., Ericsson & Simon, 1980; Zajonc, 1980). In other words, perhaps the opportunity for witnesses to view themselves on tape making their identification facilitated an introspective process whereby witnesses access information about their decision that they had not retained from the task. Despite several attempts to find support for one interpretation over the other, however, Kassin's studies generated more questions than answers.

The notion that RSA worked because witnesses learned from their observable behavior would be supported by a significant relation between observers' ratings of witness

confidence and actual identification accuracy. Kassin (1985) included observers only in Experiment 2. Although he reported that observers did not reliably rate accurate witnesses as more confident than inaccurate witnesses overall, the correlation between observers' ratings of confidence and actual accuracy was substantially higher than that for witnesses who were not allowed to view themselves on tape at all (i.e., the control group), r (30) = .24 vs. r (17) = .05, respectively. And, as mentioned above, when only choosers were considered, observers' ability to assess confidence was improved significantly compared to their ability with both choosers and nonchoosers combined. This suggests that there may be some useful information for observers to gain from viewing witness identifications. The proposed studies are aimed at maximizing this possibility.

A very plausible observable cue to accuracy is witnesses' decision time to make a choice from a suspect lineup. In order for decision time to aid jurors' judgments of accuracy, both parts of a two-stage relation would have to hold true: First, there should be a significant correlation between decision-time and actual identification accuracy and, second, there should be a significant correlation between decision-time and witnesses' judgments of confidence. In other words, if knowledge of decision-time contributes to the improvement in confidence assessment, then it must be a valid indicator of accuracy and be seen as such by the witnesses watching the tape. Kassin's (1985) studies provide marginal support for the first stage, in that accuracy and decision-time were consistently, although not significantly, related in each of the four experiments. The second stage received more support, in that decision time and judgments of confidence were consistently and signficantly related in each study. People appear, then, to use the decision-time cue in their judgments of confidence even though it is not always a valid predictor of accuracy. Note that these relations between decision time and accuracy are only correlational. In order to assess the causal influence of decision-time on judgments of accuracy, some witnesses in the third study presented in present paper were misled as to how long they took to make an

identification. This manipulation is discussed in the next section.

Kassin (1985) included several measures and manipulations designed to test the retrieval-cue hypothesis to explain the RSA effect. In his Experiment 3, participants were categorized as being either high or low on Fenigstein, Scheier, and Buss's (1975) Self-Consciousness Scale. As Kassin suggests, "People who score high in private self-consciousness tend to be cognizant of thier own inner thoughts, motives, and feelings, and characteristically, even in the absence of [RSA], they attempt to decipher their own cognitive processes" (p. 884). Kassin predicted, therefore, that the retrieval-cue hypothesis would be supported if participants high in private self-consciousness showed higher accuracy/confidence relation compared to participants low on that dimension, even without the RSA feedback. This hypothesis was not supported; in fact, those high in private self-consciousness were among those with the lowest accuracy/confidence correlation (r = .02 in the RSA group, r = -.33, ns, in the control group).

In addition to this individual-difference measure, Kassin (1985) manipulated two between-subjects variables to test the retrieval-cue interpretation of the RSA effect. In Experiment 3, some participants were asked to "reconstruct and describe the thoughts that accompanied their decision making process" (p. 885) prior to assessing their confidence. Among these "narration" participants, some viewed their identification behavior while they described their thoughts. Kassin hypothesized that if RSA enhances the accuracy/confidence relation by facilitating the retrieval of thoughts experienced during the identification, then narration participants should recall more of these thoughts and, hence, the correlation should be improved even further. Again, however, this hypothesis was not supported; in fact, the narration procedure appeared to interfere with the basic RSA manipulation in that neither of the narration groups produced a significant accuracy/confidence correlation (rs = .05 for RSA group, .18 for control, both ns).

Finally, in Experiment 4, Kassin (1985) included conditions in which some participants

spoke aloud during the original identification task. He hypothesized that if the retrieval-cue explanation for the RSA effect is correct, then thinking aloud should produce more process-related thoughts to draw on when witnesses are asked to judge their own confidence. Like the narration procedure in Experiment 3, Kassin predicted that the think-aloud manipulation should improve the accuracy/confidence correlation even among participants who do not watch themselves on tape. This hypothesis was not supported; both think-aloud and control participants in the RSA condition showed a significant confidence/accuracy relation (*r*s = .47 and .58, respectively, both *p*s < .05), but participants who only thought aloud during the identification task did not show any relation at all (*r* = -.03, *ns*).

Preview of the Present Research

Recall the earlier discussion regarding research concerned with differences between (a) statements from people with accurate versus inaccurate memories (e.g., Johnson & Raye, 1981; Schooler, Gerhard, & Loftus, 1986), (b) statements from people who are telling the truth versus those who are lying (e.g., Undeutsch, 1982; Yuille, Raskin, & Steller, in press), and (c) the behavior of people who are telling the truth versus those who are lying (e.g., de Paulo, 1985; Ekman, 1985). Consistent with all of this research, the notion that the RSA effect works because witnesses learn from their behavior implies that witnesses display behavioral indicators of confidence and/or accuracy that aid both themselves and others in calibrating their assessments of confidence with actual accuracy. What might these indicators be? And are they useful for both witnesses themselves and observers? These are the questions to which the present studies are addressed.

Despite the apparent complexity of the issue as it addressed here, the main idea is that people *are* asked to make decisions in real court settings regarding eyewitness identification accuracy and that these decisions to date have been made mostly on the basis of the

witness's behavior under cross-examination, and certainly without access to the witness's behavior during the identification task. Kassin's (1985) studies suggest that viewing the identification behavior aids both witnesses' and jurors' accuracy assessments, but he does not provide a supported argument as to how this improvement is achieved. Furthermore, it remains to be seen whether or not viewing identification behavior in addition to the cross-examination improves accuracy assessments over and above that achieved by viewing cross-examination behavior alone. Further understanding of how identification behavior relates to accuracy should also be of interest to police, since they often have to make decisions regarding whether or not a given witness's hesitant or suspicious identification should be followed up by further costly investigation.

Kassin (1985) suggested that perhaps observers, and witnesses themselves, use facial cues displayed by the witness during the identification task to aid their assessments of confidence and/or accuracy. It may be the case, however, that both types of raters use cues that are only indirectly related to the actual facial expressions of the identifier. Longer decision times, hesitation on a particular choice-alternative, and verbal intonation are examples of cues that are not expressed by facial expressions; this information could be made available to the rater without allowing him or her to see the actual facial expressions of the identifier. If these variables drive raters' assessments of identification accuracy, then no differences would be expected between conditions in which raters view the complete videotape of the identification or a modified tape in which only a silhouette of the witness is observable. A silhouette-only RSA condition is thus included in Study 1.

The addition of a silhouette-RSA condition, and the inclusion of observers, allows for the test of another hypothesis in Study 2: If witnesses' ratings of their own accuracy are correlated with actual accuracy in both the full RSA and silhouette-only RSA conditions, but observers are accurate only in the full RSA condition, then this would support the notion that witnesses have some sort of personal, privileged access to information about

their decision that is made available to them even without seeing their own facial expressions. In other words, if viewing a mere silhouette of themselves making their lineup decision improves witnesses' accuracy/confidence relation, but viewing the same tape does not improve jurors' judgments, this would suggest that the tape is somehow cuing witnesses' memory for their thoughts, feelings of indecision, judgments of similarity, and so on during the lineup task. Conversely, if both witnesses and observers are alike in the accuracy of their ratings in both the Full RSA and Silhouette RSA conditions, then this would support the notion that information unrelated to facial expressions is generally useful in assessing identification accuracy. Decision-time and verbal intonation are likely candidates for such information, and will be discussed at length later on. Of course there more other possible outcomes than are listed here; these will be discussed more fully in Study 2.

In addition to these manipulations of the type of RSA feedback, Study 1 investigates the effect of witnesses' self-monitoring scores (e.g., Snyder, 1979) on accuracy assessment. Self-monitoring is an individual-difference variable not measured by Kassin, but included in a number of other eyewitness experiments (e.g., Brigham & Cairns, in press; Hosch, Leippe, Marchoni, & Cooper, 1984; Hosch & Platz, 1984; Pigott & Brigham, 1985). Snyder's self-monitoring questionnaire "is a reliable self-report instrument designed to measure an individual's ability to observe and control his [or] her own behavior by reading subtle situational and physical cues which assist the person in behaving in a socially appropriate manner " (Brigham & Cairns, p. 10). Although no consistent relation between self-monitoring and eyewitness accuracy has been reported, it is possible that self-monitoring is an important moderating variable that influences how much a given individual benefits from viewing his own identification behavior on tape. Self-monitoring scores were therefore measured for all witnesses in Study 1.

Study 2 is also designed to test two hypotheses concerning the effect of (a) RSA on

witnesses' subsequent behavior under a simulated cross-examination and (b) exposing jurors to videotapes of witnesses' lineup tasks, to see if either manipulation improves jurors' ability to distinguish accurate from inaccurate identifiers. Accordingly, some mock-jurors will view and rate RSA witnesses from Study 1 on videotape as they undergo cross-examination while other jurors will view and rate non-RSA witnesses. Perhaps the RSA manipulation from Study 1 will affect witnesses' cross-exam behavior in a way that improves jurors' accuracy assessment. In addition, some jurors will see both a witness's cross-exam and lineup task on videotape while others will see only the cross-exam. Perhaps the witness's lineup behavior provides useful information to jurors beyond that available from the standard cross-exam that can improve jurors' accuracy assessment.

The role of actual identification decision time and people's impressions of decision time are the concern of Study 3. Decision-time has been mentioned throughout the current paper as a plausible, publicly available indicator of identification accuracy. In an early *Psychological Review* article, Henmon (1911) concluded that inaccurate judgments were associated with longer and more variable decision-times and that while many inaccurate judgments took a long time, many took especially little time. In addition, Henmon noted that the time of judgment increased monotonically as confidence decreased. All of this is in most ways consistent with the more recent data reported by Kassin.

Regarding actual accuracy and decision time, Kassin (1985) reported a nonsignificant correlation overall, but noted that when a sequential lineup was used (Experiment 4) a significant accuracy/decision-time correlation was found (r_s = -.62 and -.50 for RSA and control groups, respectively, both n_s = 18, both p_s < .05). [The sequential method of lineup adminstration differs from the more common simultaneous procedure in that each suspect-photo is presented individually to the witness without informing him or her of the total number of photos to be viewed. At the presentation of each photo, the witness is asked to make a yes/no decision as to whether or not that photo is of the criminal in the

original event. Lindsay and Wells (1985) demonstrated that the sequential procedure resulted in fewer false identifications when compared to the simultaneous procedure.] In addition, RSA witnesses in Experiment 4 showed the highest confidence/decision-time correlation in the series of studies ($r = .-75$, $n = 18$, $p < .01$). In other words, witnesses who were administered a sequential lineup and allowed to view themselves making the decision showed a strong tendency to associate longer decision times with less confidence.

What is the causal influence of decision-time on judgments of accuracy? Do less confident and/or accurate witnesses take longer to make a identification or do those who take longer to make an identification appear less confident and/or accurate? In order to investigate the causal role of decision-time on judgments of accuracy, the design for the proposed Study 3 includes conditions in which witnesses are given false feedback regarding how long they took to make their lineup decison compared to some bogus "average value observed in the study so far".

In addition to these main manipulations, the presence or absence of the actual perpetrator from the crime video is an important variable throughout all three studies presented below. Approximately half of the witnesses in Studies 1 and 3 saw pictures that contained the perpetrator and the other half saw pictures in which a predetermined look-alike was substituted in his place. This presence/absence manipulation was done for a number of related reasons.

Eyewitness identification accuracy can be likened to a signal-detection task, common in psychophysical research dealing with people's abilities to detect the presence of extremely faint physical stimuli against a background of interfering "noise". For example, researchers have investigated people's ability to detect faint, intermittent dots of light on an already-lit screen or on-and-off tones from a constant din of sounds. Participants are asked to respond "Yes" when they believe the stimulus is present. In order to measure the sensitivity of people's sense mechanisms, a measure that compares the number of times

they say "Yes" when the stimulus is present (hits) to the number of times they say "Yes" when it is not (false alarms) is calculated. In other words, it is not sufficient to respond "Yes" all the time, even though such a stategy would result in a high hit rate, because the overall accuracy would be lowered by an equally high false-alarm rate (assuming the stimulus is present only half the time, or less). Similarly, the inclusion of perpetrator-absent lineups is essential, so that a measure of witensses' accuracy can be calculated that includes their false-alarm errors (i.e., identifying an innocent suspect from an absent lineup).

The notion of people's ability to identify a suspect more often when he is actually the criminal (hits) compared to the case in he just happens to look like the criminal but is in fact innocent (false alarms) is clearly relevant to the eyewitness context. Accordingly, Wells (1984) introduced the idea of a "relative judgment" process, whereby witnesses tend to choose the member of a suspect lineup who looks most like the actual perpetrator of the crime. When the lineup contains an innocent suspect, then, even one that bears only a surface resemblance to the actual perpetrator but is most similar to him compared to the other members, witnesses show a tendency to choose him some proportion of the time. Lindsay and Wells (1980) reported a rate of from 31% to a surprising 70%, depending on the similarity of the innocent suspect to the actual criminal, relative to the other members of the lineup. This is true even though lineup instructions inform witnesses that the criminal may or may not be among the pictures although Malpass and Devine (1981) reported a reduction in the false alarm rate from 78% to 33% when the warning was given.

The relative-judgment strategy is clearly inferior to one by which witnesses withold their decision to choose someone from the lineup unless their belief that he is the criminal exceeds some very high threshold of probability, what Wells (1984) termed an absolute-judgment strategy. The presence of the perpetrator is therefore manipulated so that a measure of people's ability to choose the criminal when he is present can be

compared to their tendency to choose a similar-looking substitute from an absent lineup. A diagnosticity ratio (defined later and discussed at length by Wells & Lindsay, 1980) is therefore reported in the appropriate places in Studies 1 and 3 of the present research.

The last, but certainly not the least important, reason for including absent lineups has to do with the ecology of witness identifications in actual police settings. Although police usually conduct a lineup only when they have reason to believe the person they have apprehended is the perptetrator of the crime, there are numerous occasions on which the wrong person is suspected and subsequently identified. Enlightening accounts of such "false alarms" have been written describing life-destroying situations in which people have been arrrested and subsequently identified and convicted as perpetrators of crimes they did not commit (see, for example, Borchard, 1932; Brandon & Davies, 1973; Devlin, 1976; Frankfurter, 1927; O'Connor, 1974; Wells & Loftus, 1984; Williams, 1958). Similarly, the popular media have taken up the cause of several such cases. Obviously, it is important to simulate the real-world situation in which suspect lineups do not contain the actual criminal in the kind of experimental research described in the present paper.

II. STUDY 1

The objective of Study 1 was to investigate the roles of various aspects of videotaped lineup behavior both on people's estimates of their own lineup accuracy and confidence. In the only previous, similar studies to date (Kassin, 1985), mock witnesses viewed either a complete, full-color videotape of their own lineup behavior (Full RSA) or they received no feedback at all about their lineup task. In those studies, confidence ratings of witnesses who viewed the videotape of their own lineup behavior were significantly correlated with their actual identification accuracy, whereas control-group witnesses' ratings were not related to their actual accuracy. These results suggest that something about watching themselves on videotape serves to help witnesses calibrate their self-reported confidence with their actual accuracy. But what is it that people are using to improve their confidence calibration?

The literature on nonverbal behavior is an area of research relevant to the accuracy-assessment issue, although the area is dominated by research on cues that aid observers' detection of actors' *deceptive* communications. It is important to keep in mind that the present studies are not concerned with the case in which a witness is purposely trying to mislead the police or the court regarding an identification. Instead, the focus in on how to improve witnesses' and jurors' abilities to assess the accuracy of sincere attempts to make an accurate identification. The distinction might not be this clear, however, if expressed confidence is considered an intervening variable that influences accuracy detection in judgments of both deceitful and sincere communicators. In other words, it may be that what makes a liar believable is his or her ability to communicate in a confident manner. Therefore, the process by which judgments of accuracy are made in both the deceitful and sincere cases may be very similar. With this in mind, the most relevant

variables from the nonverbal literature appear to involve the type, or how much, of the nonverbal behavior is made available to the observer.

Overall, observers are only moderately accurate in their ability to detect deceptive communications (see Miller & Burgoon, 1982, for a review), but some manipulations have significant effects on relative accuracy between conditions. In a series of articles, Ekman (1965) and Ekman and Friesen (1967, 1969, 1974) demonstrated that observers were better able to detect whether actors were experiencing positive or negative emotions when the actors' bodies, hands, and feet were observable compared to when only faces and heads were shown. Their interpretation of this finding (the "leakage" hypothesis) is that actors are better at controlling their facial expressions compared to body movements, which apparently betray attempts to deceive. The leakage hypothesis has received some support from studies by Hocking, Bauchner, Kaminski, & Miller (1979) and by Littlepage and Pineault (1979). In these studies, however, the face-body difference was limited to cases in which the actor was being deceitful. With respect to the identification task, Kassin (1985) suggested that facial expressions may have been especially useful indicators of identification accuracy in his RSA studies. Study 1 of the present research was designed to investigate this possibility, as discussed later in this section.

In a similar vein to the manipulation of face vs. body cues, other researchers have manipulated the presentation of the audio and video portions of actors' communications. Although most studies show no difference between live, audio-only, and written-transcript presentations of communications (e.g., Hocking et al., 1979; Bauchner, Kaplan, & Miller, 1980), those that do show a difference often favor presentation modes that eliminate much of the available nonverbal behavior (e.g., Hocking et al. (nonsignificant trend); Maier & Thurber, 1968). Again, all of these studies were concerned with deceptive communications. These manipulations would not necessarily be expected to have the same impact on accuracy assessments in the proposed studies, in which actors are communicating as truthfully as possible. These studies do, however, suggest a

manipulation that is potentially useful for an eyewitness identification task.

The present study manipulated components of the Full RSA feedback to see which of them on their own may be sufficient to result in witnesses' improved calibration. First, though, it was important to see if the Kassin (1985) results could be replicated with the current materials, so a Full RSA condition based on Kassin's procedure was included. In a second condition, witnesses also viewed a videotape of their lineup behavior, but they saw only a silhouette, or shadow, of themselves on the TV screen (Silhouette RSA). Witnesses in both group then, could see how long it took them to make a decision and could hear any verbal hedges, hesitiation, intonation changes that may have accompanied it. Witnesses in the latter group, however, could not see their own facial expressions and therefore could not benefit from any facial-cue information that may have helped witnesses in the previous studies to improve their calibration. Witnesses in a third group did not receive any videotaped feedback, but were told me to, "Imagine what you might look like if you were to see yourself on videotape" (Imag ed RSA). Perhaps merely thinking about the task, recreating their thoughts and feelings of indecision, their judgments of the pictures' similarity, and so on, is sufficient to produce a significant relationship between wintesses' self-reported confidence and actual accuracy. Finally, witnesses in another group received no feedback or instructions regarding their lineup task, but were given a brief explanation of the research-credit system for about the same amount of time that they spent looking at the pictures (Control). This condition represents the standard procedure in actual police lineups and in other eyewitness identification studies, in that witnesses received no manipulation before rating their confidence in their identification decision.

Method

Participants and design

Participants were 89 male and female students from the introductory psychology research pool at the University of Alberta. Each participant was given one credit for partial fulfillment of introductory psychology course requirements. There were four conditions in the design: Full RSA, Silhouette RSA, Imagined RSA, and Control. Participants were assigned to condition randomly, with 23, 19, 22, and 25 participants in each of the four groups, respectively. Within each condition, approximately half of the participants saw a series of pictures that contained the actual perpetrator from the video.

Materials

Each participant viewed a videotaped, simulated crime projected onto a 150 cm screen by a Sony projection TV. The crime video opened with a suspicious-looking man prowling the hallway of a large apartment complex. The scene then cut to a woman preparing her clothes for the laundry, leaving her suite, and encountering the man in the hallway. There, she saw him knocking on a door and he engaged her in very polite conversation. He explained that his room-mate had locked him out of their new suite and that he had not yet received a set of keys to the doors. He asked if he could use her phone to call the manager so that he or she could let him in. She replied that she had better not let him in her own apartment because her husband was not home, but that she could call the manager on his behalf. He agreed, and followed her back to her suite. The woman asked the man to remain in the hallway while she made the call and again he agreed. After a few moments, the man burst in by opening the unlocked door, quickly surveyed the apartment for valuables, grabbed a purse that was on a desk next to the door, and fled. The criminal was in full view of the camera for approximately 45 seconds, but was in and out of shadows throughout the exposure.

After the crime video, but prior to viewing the suspect lineup, all participants completed the Revised Self-Monitoring Scale (Gangestad & Snyder, 1985a; see Appendix 1). This

instrument consists of 18 items, a subset of the 25 items from the original Self-Monitoring Scale (Snyder, 1974). Each of these 18 items was show to load above .15 on the first unrotated, general self-monitoring factor in an oblique-rotation factor analysis conducted by Gangestead and Snyder (1985b).

The suspect lineup consisted of eight color slides projected on the same 150 cm screen, this time by a Kodak Carousel remote-contol projector. The pictures were highly similar, each easily matching even a very accurate verbal description of the criminal (e.g., caucasian, mid-20s, medium build, medium length brown hair combed back, full moustache, dark eyebrows, dark eyes). Each picture presented a front pose, from the mid-chest upward, and was approximately 35 cm by 50 cm when projected on the screen. In the perpetrator-absent lineups, a prerated look-alike for the real criminal was substituted in his place. The slides were presented in the same order for all witnesses, just as the pictures in a simultaneous lineup are often in the same configuration for all witnesses in a given study.

While they looked through the lineup, all participants were videotaped with a Panasonic VHS "Reporter" camera from their left at about a 45 degree angle. All participants were seated at a table during the lineup, and the camera recorded a full upper-body shot. In the Silhouette condition, a translucent screen was placed between the participant and the camera and a small light was used to project a shadow of the witness onto the screen. In all conditions, sound was recorded on the videotape via a voice-quality, Radio Shack PZM reflecting-base, electric microphone. After making a lineup decision, participants in the Full and Silhouette RSA conditions were allowed to watch the videotape of their lineup behavior on a 50 cm, RCA color TV.

Subsequent to receiving the appropriate post-lineup manipulation, all participants were asked to answer five questions regarding their impressions of the lineup task (see Appendix 2). These questions asked participants to (1) rate their confidence in their lineup decision, from 1 (not at all confident) to 7 (very confident), (2) estimate the probability that it was

correct, from 0 to 100%, (3) indicate whether or not they were willing to testify under oath, (4) rate the difficulty of the task, from 1 (not at all difficult) to 7 (very difficult), and (5) estimate how long they spent looking through the pictures, in seconds or minutes.

Procedure

Participants arrived at the laboratory individually. They were told at the outset of the session that the project had to do with people's impressions of videotaped versus slide material, that they would see some samples of both media, and then be asked some simple and straightforward questions about them afterward. Participants were told to sit back and to play close attention to the first videotape. The first tape showed a 75 sec portion of an amateur tennis match between two males, one with dark hair and the other blond. After the tape, participants were asked to complete four questions regarding their "impressions" of the two players (see Appendix 3). Answering these questions required participants to make personal judgments of the players' relative abilities, their experience with the game, the nature of their relationship with each other, and who would likely make the better tennis instructor. The presentation of this first tape, and the accompanying questions, was intended to keep participants from spontaneously treating the subsequent crime film as a memory task. This procedure has been used successfully in previous similar studies (e.g., Turtle & Wells, 1988), in that several participants remarked that they thought they would have to make judgments about the people in the crime tape that were similar to those made about the tennis match. In addition, postexperimental interviews suggest that very few participants guess the memory nature of the study when this procedure is used.

After completing the initial questions, participants were asked to watch another tape. They were told that this second tape would be about the same length as the first, and that they would be doing something very similar to the first task when the second tape was over. Participants then watched the crime film, which lasted about 2 min and 30 sec. Afterward, they were asked to complete the Revised Self-Monitoring Scale (Gangestead & Snyder, 1985a), which took about 3 to 5 min to complete.

At this point, participants were told the real purpose of the study; namely, that their memory for the crime depicted in the second videotape would be looked at in a manner similar to what might happen to real witness who had seen similar event take place. Participants were told that their first task was to see if they could recognize, or identify, the criminal in the tape from a series of very similar photographs. Intructions for the lineup task were as follows:

> What we're actually interested in in this study is people's eyewitness memory, or eyewitness testimony, abilities. What I'd like you to do is to imagine that you are an eyewitness to the crime that was depicted in that second videotape. I'm going to take you through some of the steps that a real witness to a real crime might go through if you were to come forward and say that you had seen something like that take place. The first thing I'd like you to do is to look through a number of pictures of very similar-looking people to see if you can recognize, or identify, the criminal from that second videotape. Now, there are a number of things to keep in mind as you look through the pictures. First, the man from the tape may or may not be among pictures, but if he is there, he's there only once. Second, you're going to be seeing these pictures one at a time, and for each picture I'd like you to decide, yes or no, whether or not *that's* the man from the tape. Third, as you look through the pictures, you can't go back to a picture after you've passed it. And, finally, if you do choose someone you don't get to see the rest of the pictures, if there are any -- you won't know, of course, because you may have chosen the last one. For each picture, I'd like you to say outloud the number of the picture ("Number 1", "Number 2", etc.) and "Yes" or "No", in answer to the question, "Is that the man from the second videotape?"

Prior to viewing the pictures, participants were asked for their permission to videotape them while they looked through the lineup. Participants were then instructed on how to use the remote control for the slide projector, the video camera was turned on, and participants went through the pictures at their own pace. Witnesses' decision times were not recorded at this time, but were measured from the videotape at a later date.

After making a lineup decision (either choosing someone or going through the pictures

without making an identification) participants received one of the postlineup manipulations. Those in the Full RSA condition were given the opportunity to watch the videotape of themselves looking through the pictures. Those in the Silhouette RSA condition also saw the videotape of themselves, but only a shadow of themselves was visible. In these two conditions, participants were told that, "There is nothing specific about your behavior that you should be looking for, but you should watch whatever it is about your own behavior that attracts your attention". Those in the Imagined RSA condition were asked merely to, "Imagine what you might look like if you were to see yourself on videotape. Imagine that I've rewound the tape of you looking at the pictures and that you get to see yourself on the TV." Participants in the Control group received a brief explanation of the research-credit system, the duration of which was matched as closely as possible to the actual time taken to look through the pictures. After the RSA manipulation, participants answered five questions regarding their confidence in, and the likely accuracy of, their lineup decision (see Appendix 2).

Witnesses who had made an identification were then asked 11 questions about the film and the person they identified (see Appendix 4). These questions were designed to simulate a cross-examination; some were misleading or suggestive (e.g., "Isn't it true, though, that the man [in the film] was facing *away* from you for most of the time?") and some were designed to undermine the witnesses' identification confidence if they had any doubt in it to begin wtih (e.g., "Do you think it's possible that you've only seen the man you chose from the pictures somewhere esle, and that you just *think* he's the man from the film?"). Finally, participants were fully debriefed, given credit, and thanked for their participation.

Results and Discussion

Definition of scoring terms. There are a number of ways to score witnesses' photo lineup decisions, depending on a study's emphasis on forensic relevance and ecological validity, the assumptions made about the lineup model (see Wells & Turtle,

1986), the presence or absence of the actual perpetrator in the lineup, the presentation format of the photos (simultaneous vs. sequential), and other factors. In Study 1, the relevant scoring terms were defined as follows:

1) Accurate identification: A witness chooses the actual perpetrator of the crime from a perpetrator-present lineup.
2) Inaccurate, or false, identification: A witness chooses an innocent suspect from either a perpetrator-present or perpetrator-absent lineup.
3) Accurate rejection: A witness chooses no one from a perpetrator-absent lineup.
4) Inaccurate rejection: A witness chooses no one from a perpetrator-present lineup.

Lineup decisions. The lineup-relevant data are presented in Table 1. Overall, the rate of identifcations (accurate or inaccuate) was low. Only 36 of the 89 witnesses made an identification from the lineup (40%); 20 of these 36 were accurate identifications (56%). It is important to point out, however, that of the 89 lineups administered, 39 did not contain the actual perpetrator from the videotape (44%) and that 10 of the 16 inaccurate IDs were from these perpetrator-absent lineups (63%). If only identifications from perpetrator-present lineups are considered, then the accuracy rate increases to 20 out of 26, or 77%. When all lineup decisions in the perpetrator-present cases are considered, however, the accuracy rate drops to 20 out of 50 (40%). In the absent-lineup cases there were 10 IDs out of a total of 39 presentations, which is a false ID rate of 26% for those lineups. Thus, the diagnosticity ratio for identifications (i.e., the ratio of percent accurate IDs with present lineups over the percent inaccurate IDs with absent lineups; see Lindsay and Wells, 1980) is equal to .400 / .256, or 1.563. The diagnosticity ratio for nonidentifications (i.e., the ratio of percent accurate rejections with absent lineups over the percent inaccurate rejections with present lineups) is equal to .744 / .600, or 1.239. A diagnosticity ratio of 1.0 indicates a nondiagnostic or chance level of performance.

Recall the earlier discussion regarding the common finding in t[illegible]tional confidence-calibration studies that people are generally overconfi[illegible]when asked to estimate the likelihood of their accuracy. Witnesses' probability estimates of decision accuracy in Study 1 are consistent with that finding, in that their average estimate was 69.2% compared to an actual overall accuracy of only 55.1%.

Overall effects of the RSA manipulation. The RSA manipulation was administered after witnesses had made their lineup decisions. Therefore, no differences in relative rates of choosing, accuracy, or decision time were expected because these variables were measured before witnesses experienced any of the manipulations regarding RSA levels. Accordingly, no differences on these variables were observed in a one-way analysis of variance (ANOVA) with type of RSA as the factor, all Fs (3, 85) < 1. The means for these measures plus the gender breakdown and lineup-type as a function of the RSA manipulation are presented in Table 2.

The RSA manipulation could have affected the dependent measures that were recorded after the treatment was administered. Witnesses exposed to their own videotapted lineup behavior, for example, may have felt more confident overall compared to other conditions, without regard to their actual accuracy. Table 3 presents the confidence, probability, difficulty, and estimated decision-time measures as a function of RSA condition. Again, these data were submitted to a one-way ANOVA with type of RSA as the factor. Most important is the fact that confidence was not affected overall by the type of RSA feedback received, F (3, 85) < 1. Similarly, there were no significant differences among the means for any other dependent variable as a function of RSA, all Fs < 1.

The confidence/accuracy relation.. The point-biserial correlation between confidence and accuracy overall was significant, r = .179, n = 89, p (one-tailed) < .05. This figure is close to the population estimate of r = .252 calculated in a recent meta-analysis of 35 studies by Bothwell, Deffenbacher, and Brigham (1987), and falls well within their 95% confidence interval that ranges from r = .08 to .42. Apparently, accurate

witnesses were more likely overall to report higher confidence in their decisions than inaccurate witnesses.

Broken down by condition, however, the impact of the RSA manipulation shows clearly in the differences among the correlations. In the Imagined RSA and Control conditions, the confidence/accuracy correlations were not significant, rs = .076 and -.071, ns = 22 and 25, respectively, both ps > .05. In the Full and Silhouette RSA conditions, however, the correlations were highly significant, rs = .446 and .540, ns = 23 and 19, ps < .025 and .005, respectively, and did not differ from each other, z = .373, ns. The significant correlation in the Full RSA condition replicates the Kassin (1985) results, in that witnesses who saw a complete video picture of themselves were better at calibrating their confidence with their actual accuracy. And, the significant correlation in the Silhouette RSA condition suggets that those witnesses did not need to see their own facial expressions in order for their confidence calibration to benefit from exposure to a videotape of their lineup task. Furthermore, when the data were recoded as Video (Full and Silhouette RSA) or No Video (Imagined RSA and Control), the correlation in the former case was r = .484, n = 42, p < .005 and in the latter case r = -.014, n = 47, p > .05. These two correlations are significantly different from each other, z = 2.47, p < .01. In other words, if no distinction is made between witnesses in the two video-feedback groups and they are compared to all other witnesses, then the advantage of *any* video-feedback, even without facial-cue information, is clear.

Further supporting the confidence/accuracy correlational data are the correlations between witnesses' probability estimates of having made an accurate ID and their actual ID accuracy. Although self-reported confidence and estimated probability were highly correlated overall (r = .934, n = 89, p < .001), rendering the measures virtually redundant with each other, it is encouraging that the corresponding confidence/accuracy and probability/accuracy correlations are consistent as a function of RSA feedback. For instance, the probability/accuracy correlation for witnesses in the Video groups was r =

.479, $n = 42$, $p < .005$, whereas for witnesses in the No Video groups it was only $r = -.070$, $n = 47$, $p > .05$.

Although the point-biserial correlation between accuracy and self-rated confidence is the most common indicator of the relation between the two measures, the data can also be analyzed from an analysis of variance perspective. Table 4 presents the mean confidence and probability ratings as a function of RSA condition and lineup accuracy. Accurate witnesses were more confident overall and reported higher probabilities that they were correct compared to inaccurate witnesses. Two-way ANOVAs with RSA type and accuracy as the factors revealed that the difference in probability was not significant as a function of accuracy, $F(1, 81) = 1.652$, $p > .05$, but that the difference in confidence was marginally significant, $F(1, 81) = 3.380$, $p = .061$. The benefit of RSA feedback for confidence calibration, however, could be revealed by a significant two-way interaction between Video/No Video and identification accuracy on witnesses' mean confidence ratings and probabilities; if RSA gives witnesses an advantage, then accurate witnesses who viewed their own lineup behavior (Video) should have reported higher confidence and higher probabilities in their lineup decisions compared to inaccurate witnesses. In contrast, no difference would be expected between accurate and inaccurate witnesses in the Imagined RSA or Control conditions (No Video).

Examination of Table 4 shows that the difference in mean confidence between accurate and inaccurate witnesses as a function of RSA is consistent with the idea that both Full and Silhouette RSA improve witnesses' confidence calibration. Accurate versus inaccurate witnesses in these two conditions showed about a 1-point average difference in confidence, whereas witnesses in the other two groups differed by less than half of a point (the difference in the Control group was actually in the wrong direction). A two-way ANOVA with Video (Full and Silhouette RSA combined vs. Imagined RSA and Control combined) and accuracy as the factors revealed a significant interaction for both the confidence and probability measures, $Fs(1, 85) = 3.988$ and 4.802, respetively, $ps < .05$. In other words,

accurate witnesses who received either type of video feedback were more confident, and thought there was a higher probability that they were correct, compared to their inaccurate counterparts. There were no differences, however, between accurate and inaccurate witnesses who did not receive any type of video feedback.

The self-monitoring scores (SMS). The 18-item, Revised Self-Monitoring Scale (Gangestead & Snyder, 1985a) was administered to all participants. The average score was 9.180, the median score was 9, and the lower and upper quartiles were 6 and 12, respectively. Self-monitoring is considered a class variable; a high score idicates a higher *probability* that an individual belongs to the class of high self-monitors (HSMs), rather than suggesting that he or she has a greater *amount* of self-monitoring (Pullyblank, 1987). Therefore, a median split at a score of 9 was used to divide the participants into High and Low self-monitors (LSMs). Accordingly, there were 37 participants classified as LSMs and 37 as HSMs. Fifteen participants had scores of exactly 9 and were excluded from the analyses.

The main question of interest was whether or not HSMs were better at decoding aspects of their own behavior that reflect accurate vs. inaccurate lineup decisions. In other words, did HSMs who viewed their own lineup behavior on videotape benefit more in their confidence calibration compared to LSMs? It is logical to hypothesize that HSMs are more practiced than LSMs at evaluating their own behavior, although Snyder's (1979) theory postulates that these evaluations are usually made on the basis of information observed in other people's reactions to the actor's behavior. Still, it is likely that HSMs are more apt than LSMs to seek out opportunities to see themselves on videotape, or in mirrors, and that this greater experience with their own appearance and behavior may lead to more accurate evaluations of lineup accuracy.

Overall, HSMs did not differ from LSMs in their ability to calibrate their confidence estimates with their actual lineup accuracy. HSMs showed an $r = .117$, $n = 37$, $p > .05$, and LSMs showed an $r = .196$, $n = 37$, $p > .05$. Within the Full RSA condition, however,

there appeared to be a dramatic difference between the accuracy/confidence correlation for LSMs and HSMs. LSMs showed a negligible correlation, $r = .061$, $n = 8$, $p > .05$, whereas HSMs showed a highly significant correlation, $r = .847$, $n = 10$, $p < .005$. It is tempting to conclude that the results are consistent with the original hypothesis; namely, that HSMs have an advantage over LSMs when it comes to evaluating their own observable behavior. A closer look at the data, however, revealed that the correlations for HSMs and LSMs within the Full RSA condition are not interpretable because of insufficient variance in, and a disproportionate distribution of, the accuracy scores between the SMS groups. Only 1 of the 8 LSMs made an accurate lineup choice, compared to 9 of the 10 HSMs, Chi-Square (1) = 5.27, $p < .05$. This vast difference in accuracy in the Full RSA condition as a function of the SMS was not significant across the entire study, however. When all participants in all conditions are considered, there is no difference in accuracy as a function of SMS, F (2, 84) < 1. Similarly, HSMs did not differ from LSMs on most of the other measures, all Fs (2, 84) < 1. In fact, the only the signficant difference as a function of the SMS split was for gender; 76% of the LSMs were female, compared to only 46% of the HSMs (60% of the median SMSs were female), F (2, 86) = 3.93, $p < .05$.

Decision time data. Overall, witnesses spent more time looking at the target picture compared to any of the other slides. Like many of the other measures, however, witnesses' decision times can be grouped in a number of different ways, some of which are presented in Table 5. Most interesting is the question of whether or not witnesses who inaccurately rejected the perpetrator-present lineup spent more time looking at the target picture than the other slides. If so, this would provide support for the idea that witnesses had some sort of memory for the criminal, but that for one reason or another they did not choose him from the lineup. The bottom portion of Table 5 shows the relevant times for the 24 witnesses who "missed" identifying the target from a target-present lineup. Note that these witnesses did spend more time looking at the target compared to the average time they spent on the other slides. A within-subjects ANOVA with slide number as the factor

revealed that there were significant differences among the decision times, $F(1, 161) = 5.032$, $p < .001$, and an orthogonal contrast pitting the time spent on the target picture against that spent on each of the other slides indicated that witnesses did spend more time on the target compared to the other pictures when they rejected the perpetrator-present lineup, $F(1, 161) = 21.107$, $p < .001$. Perhaps even more informative is that 14 of these 24 (58%) inaccurate rejectors spent more time on the target's picture than on any of the other seven slides (chance = 12.5%).

Another interesting group is composed of the 20 witnesses who accurately identified the target. The decison times for this group are presented in the top portion of Table 5. Another within-subjects ANOVA indicated that there were differences among the decision times, $F(4, 76) = 5.894$, $p < .001$, and a subsequent contrast indicated that accurate identifiers did spend more time looking at the target compared to the other slides, $F(1, 76) = 24.260$, $p < .001$. Nine of the 20 (45%) accurate identifiers spent the most time on the target picture and many of them spent almost 10 times as long looking at him, compared to the other slides. Similarly, inaccurate identifiers spent somewhat more time on the picture they chose compared to the other slides, but this difference was not significant. The times for accurate rejectors indicate that this group also spent somewhat more time on the innocent suspect compared to the other pictures in the perpetrator-absent lineup.

The correlations between actual total decision time, estimated decision time, confidence, and accuracy are presented in Table 6. Overall, total decison time and accuracy were positively related ($r = .236$, $n = 89$, p two-tailed $< .05$), whereas decision time and self-reported confidence were negatively related ($r = -.259$, p two-tailed $< .02$). In other words, longer decision times were associated with greater accuracy, and with lower confidence. Although these results appear to be at odds with Kassin's (1985) decision time data and with intuition, it is important to keep in mind the nature of the identification task as it was administered in this study. It is understandable that many witnesses spent so much time making identifications because they were told that they would not be allowed to

see any of the remaining pictures, if they were any. Therefore, although they may have made the decision sooner in other circumstances, many witnesses waited a long time before identifying a given picture.

Perhaps more informative than the relations regarding total decision time are the data concerning only the decision time for the individual picture chosen by both accurate and inaccurate identifiers. With a sequential lineup, it is intuitively appealing that both identifiers' accuracy and confidence ratings would be most related to how long they spent looking at the picture they eventually chose, rather than the amount of time spent on all the pictures in total. Even though accurate identifiers spent almost twice as long as inaccurate identifiers on their chosen picture, however, the values were not signficantly different from each other, t (35) = 1.50, $p > .05$ (corresponding r between accuracy and chosen-picture decision time = .249, n =36). It is interesting to note that accurate identifiers were over three times as variable in their chosen-picture decision times compared to their inaccurate counterparts. The correlation between chosen-picture decision time and self-reported confidence was also not signficant, r = -.238, n = 36.

Finally, there are the data for witnesses' estimates of how long they spent looking at the lineup. Overall, estimated decision time was significantly correlated with actual decision time (r = .45, n = 89, $p < .01$), suggesting that witnesses had a good idea of their relative decision time, but overall their estimates tended to be much higher than their actual values (Ms = 30.0 sec and 69.3 sec for actual and estimated times, respectively). Such overestimation of time, especially in a cogntively-engaging task like a lineup identification, is common (e.g., Shiffman & Bobko, 1974). Estimated decision time was not correlated with either accuracy or confidence, rs = .190 and -.101, respectively, ns =89, both two-tailed $ps > .05$.

It appears, then, that enough information was contained in the Silhouette RSA condition for witnesses' confidence ratings to benefit from the viewing themselves on tape, even without seeing their facial expressions. This fact eliminates explanations of the RSA

effect that rely on facial cues as the basis for the improved confidence/accuracy relation. Depending on the results of the observers' ratings in Study 2, however, it is not clear whether witnesses in the Full and Silhouette RSA conditions based their calibrated confidence ratings on their overt behavior recorded on tape (e.g., verbal hedges and/or intonation, decision time) or if reviewing the tape reminded them in some private way of their own certainty or indecision during the task.

The decision-time results have important implications for Study 3, in which alleged decision time was manipulated in a categorical manner (i.e., less than or greater than the "average" witness). Perhaps people think that accurate decisions should take more or less time than inaccurate ones, but only by manipulating people's impressions of how long they took to make a decision can the relationship be understood more fully.

III. STUDY 2

Study 2 extends Kassin's (1985) work by including participant-jurors' ratings of witnesses' lineup behavior in either normal or silhouette form and witnesses' cross-examination behavior, with three objectives in mind. First, will jurors who see witnesses' identification behavior on videotape ascribe more confidence to accurate compared to inaccurate identifiers? More importantly, will there be any difference between jurors who see only a witness' silhouette compared to those who see a normal viedeotape?

Recall from Study 1 that witnesses' confidence ratings in both the Full and Silhouette RSA conditions were signficantly related to their actual identification accuracy, suggesting that they did not need to see their own facial expressions in order for their confidence calibration to benefit from viewing their lineup task. There are several possible outcomes of interest regarding jurors' ascriptions of witness confidence and actual witness accuracy. (1) If jurors' ascribed confidence ratings are also related to witness accuracy after viewing either type of lineup tape, then this would suggest that something about the witness' lineup behavior, with or without facial cues, provides publicly available information regarding the witness' identification accuracy. (2) If jurors' ascribed confidence ratings are related to witness accuracy only after viewing the normal lineup tape, then this would suggest that something about the witness' lineup behavior, with facial cues included, provides publicly available information regarding the witness' identification accuracy. (3) If jurors' ascribed confidence ratings are related to witness accuracy only after viewing the silhouette tape, then this would suggest that something about the witness' lineup behavior, facial cues excluded, provides publicly available information regarding the witness' identification accuracy. It would also suggest that jurors' viewing of the normal tape actually interfered with their confidence ascriptions and that they were better off seeing only the witnesses'

silhouette. (4) Finally, if jurors' ascribed confidence ratings are not related to witness accuracy after viewing either type of lineup tape, then this would suggest that the benefit enjoyed by witnesses who viewed their lineup tape in Study 1 was a function of some kind of private, retrieval-cue process.

The second major question has to with the potential transfer of the improved confidence/accuracy relation for RSA witnesses to their cross-examination behavior. Because witnesses in the Full and Silhouette RSA conditions in Study 1 demonstrated improved confidence calibration over their Imagined RSA and Control counterparts, it is expected that this improvement could transfer to the witnesses' cross-examination testimony and thereby give jurors who see only the cross-examination a better chance to distinguish accurate from inaccurate identifiers. If so, then videotaping witness identifications could benefit credibility judgments in court just through their effect on witnesses' cross-exam behavior and without being shown to jurors at all. Therefore, the cross-examination rating data will be analyzed to see if accurate witnesses who saw their identification behavior on videotape are believed by jurors more than inaccurate witnesses who also saw themselves. Conversely, no difference in believability is expected between accurate and inaccurate witnesses who did not receive any RSA feedback. This lack of a difference in jurors' belief rates for accurate versus inaccurate witnesses would be consistent with previous research in which jurors' ability to discriminate accuracy on the basis of witnesses' cross-examination did not exceed chance performance (e.g., Wells, Lindsay, & Ferguson, 1979; Wells, Ferguson, & Lindsay,1981).

Finally, it is expected that jurors who are allowed to see both the cross-examination and lineup behavior should have the best chance to evaluate the accuracy of the witnesses' identification compared to jurors who see only the cross-examination or only the lineup behavior. If the opportunity to view witnesses' lineup identifications proves to be benefical for jurors, then this would provide the basis for a recommendation to the justice system that

such videotapes be made available to jurors in eyewitness identification cases. In addition to these main hypotheses, the relationship between witnesses' decision times and participant-jurors' ratings of their behavior will be analyzed and discussed.

Method

Participants and design

The participants were 192 male and female students from the introductory psychology research pool at the University of Alberta. Each participant was given one credit for partial fulfillment of introductory psychology course requirements. Each participant watched a videotape of one of 32 of the 36 witnesses who made an identification in Study 1. Participants were assigned randomly to one of three conditions: Lineup Only, Cross-Examination Only, or Both (lineup and cross-examination). Each of the witnesses was seen by a total of six jurors, one in each of the three condiditons with a replication.

Materials

The stimuli consisted of videotapes of 32 witnesses who made a suspect identification from the lineup in Study 1. There were four accurate and four inaccurate witnesses from each of the four RSA conditions (the three extra accurate identifiers from the Full RSA condition and the one extra accurate identifier from the Imagined RSA condition were deleted randomly). Participant-jurors viewed the witnesses on 50 cm RCA color TVs while the audio was broadcast through an intercom system to participants' headphones.

Participants in the Lineup Only condition responded to five questions on a single sheet of paper (see Appendix 5). They were asked (1) whether or not they believed the witness had made an accurate ID, either Yes or No, (2) how confident they were in that belief, from 1 = not at all confident to 7 = extremely confident, (3) to estimate the probability that the witness was correct, from 0% to 49% if they answered No to Number 1 or 51% to 100% if they answered Yes to Number 1, (4) how confident they thought the witness was, from 1

to 7, and (5) how long they thought the witness spent looking at the pictures, in seconds or minutes. Participants in the Cross-Examination Only condition responded to the same first four questions as those in the Lineup Only group. Instead of the time-estimation question, however, they were asked how believable they thought the witness was, from 1 = not at all believable to 7 = extremely believable (see Appendix 6). Finally, participants in the Both (lineup and cross-examination) condition answered all six of the questions described so far (see Appendix 7).

Participants were given typewritten instuctions tailored to their particular condition (see Appendices 8, 9, and 10 for the Lineup Only, Cross-Examination Only, and Both instructions, respectively).

Procedure

Participants arrived at the laboratory in groups of three and were immediately directed to one of three identical cubicles, each of which contained a TV and an instruction sheet. The TV and headphones for participants in the Cross-Exam Only condition were not initially operable, because they were not allowed to see or hear the witness' lineup tape. Participants read that the study was concerned with people's impressions of eyewitness testimony and eyewitness identification and that they would be seeing and evaluating one witness on videotape. After waiting approximately 2 min, the experimenter ensured that the Lineup Only and Both participants who had headphones were receiving the audio broadcast and that they were ready to watch the witness on the TV screen. Participants could confirm that they were receiving the audio message by pressing a button in their cubicle. At that point, the experimenter played the tape of a witness' lineup identification to two of the three participants.

When the tape was finished, the experimenter delivered a response sheet to the Lineup Only participant and showed him or her that the questions were self-explanatory. The TV and headphones in that cubicle were disconnected at this point. Then the experimenter gave

headphones to the Cross-Examination Only participant and turned on the TV in that cubicle. The experimenter then played the tape of the same witness' cross-examination to the participants in the Cross-Examination Only and Both conditions. Lineup Only participants were free to leave their cubicle and receive credit and a written debriefing whenever they were finished, which was always while the cross-examination tape was playing. When the tape was finished, the experimenter delivered the appropriate response sheet to the two remaining participants, briefly reviewed the questions, and waited for them to finish. When they had completed the questions, participants received credit and a written debreifing. The whole session lasted approximately 20 min.

Results and Discussion

A videotape of each of the 32 identifiers from Study 1 was seen by three participant-jurors at a time in Study 2. One juror saw only the witness' lineup behavior, one saw only the cross-examination, and the other saw both the lineup behavior and the cross-exam. Once all 32 witnesses were rated in each of these three conditions, a replication was conducted so that each witness was eventually seen by a total of six jurors. The data were first analyzed in a two-way ANOVA with juror condition and the replication as factors to see if the two variables interacted for any of the measures. Univariate F-tests for each of the dependent variables indicated that the two factors did not interact, suggesting that the responses from the first 96 jurors could be combined with those from the 96 jurors in the replication group.

Overall, participant-jurors were only 54.7% accurate in their ability to distinguish accurate from inaccurate witnesses. In other words, 105 of the 192 raters who saw either a witness' lineup behavior, cross-examination behavior, or both, made the appropriate decision either to believe an accurate identifier or not believe an inaccurate one. A z-test for proportions indicated that this level of discrimination does not exceed chance performance, $z = 1.299$, $p > .05$.

Did jurors ascribe more confidence to accurate compared to inaccurate witnesses on the basis of witnesses' lineup behavior alone? The first question of interest concerns the relation between jurors' ratings of witness confidence on the basis of viewing only the witnesses' lineup behavior and actual witness accuracy. Of further interest are the differences (if any) between jurors' ascriptions of confidence when they see only a silhouette of witnesses compared to when they see a normal tape. Table 7 presents the relevant correlations. Overall, the correlation between jurors' ascriptions of witness confidence when they saw any type of lineup and actual witness accuracy was not significant, $r = .108$, $n = 64$, $p > .05$. Similarly, the subset of juors who saw only a normal lineup tape (i.e., of witnesses in the Full and Imagined RSA conditions and the control group) also showed a nonsignificant correlation, $r = .013$, $n = 48$, $p > .05$. Somewhat surprisingly, however, the subset of juors who saw only the silhouette lineup tape did show a signficant correlation between their ascriptions of witness confidence and actual witness accuracy, $r = .444$, $n = 16$, $p < .05$. In sum, then, participant-jurors who saw a normal tape of the witnesses' lineup behavior did not demonstrate any tendency to ascribe more confidence to accurate compared to inaccurate witnesses, but those who saw only the silhouette tape did reliably rate accurate witnesses as more confident than inaccurate ones.

With respect to the public versus private nature of the RSA effect on confidence/accuracy calibration, the fact that jurors who saw a normal videotape of witnesses' identification behavior did not ascribe more confidence to accurate compared to inaccurate witnesses suggests that the benefit in calibration enjoyed by witnesses was due to some information on the videotape useful only to them. On the other hand, the fact that jurors who saw only a silhouette tape did ascribe more confidence to accurate over inaccurate witnesses suggests that some kind of public, nonfacial information on the tape was available to jurors on which they could base their confidence judgments. Although

these results are at odds with Kassin's (1985) findings [although he did not use observers for witnesses who saw a sequential lineup], they are consistent in many ways with the findings of the nonverbal research discussed earlier.

It is important to point out here that the Silhouette RSA tape is virtually an audio-only recording of the witnesses' lineup task. Although the witnesses' silhouettes are easily distinguished, there is no body information (e.g., leg movements, hand gestures) like that contained in previous nonverbal research. It may be that the slight postural shifts observable on the tape are of importance, but nearly all of the witnesses were seated upright and remained so throughout the task. In addition, postural cues have figured very little in any of the nonverbal research, except as they relate to cues of friendliness or interest in a conversation partner. The discussion below is predicated on the assumption that the Silhouette tape did not contain anything meaningful beyond the verbal information.

Recall that the work on nonverbal cues showed observers to be more accurate at detecting deceitful communications when they saw videotapes that excluded actors' faces from view (e.g., Ekman, 1965; Ekman & Friesen, 1967, 1969, 1974). Observers' failure to detect deception given the full body exposure was attributed to observers' tendency to pay an undue amount of attention to the actors' face, which provided little deception relevant information. Ekman (e.g., 1988) has elaorated on this finding by showing that the decisions of observers exposed to full audiovisual presentations of deceptive and nondeceptive actors appear to be interfered with, or diluted by, irrelevant information, but that observers can detect deceptive cues when their attention is focused on the communicative channel that contains detection-relevant information. This interpretation fits nicely with the present results, if it is the verbal channel that contains some useful accuracy-relevant information from witnesses. Perhaps either the content or the character of the witensses' verbal statements are in some way(s) indicative of their actual identfiction accuracy. Nonverbal research has shown that speech errors, pauses, and changes of pitch

are reliable indicators of changes in emotion[illegible]e, although not always of deception (Ekman, 1985). The fact that witnesses themselves benefited from exposure to the Full RSA videotape in Study 1 might be due to their ability to override their own irrelevant facial information, since they are more familiar with their own idiosyncratic expressions. These ideas will be discussed further in the General Discussion.

It is important to point out here a possible alternative explanation for the fact that only observers who saw the silhouette version showed a significant relation between their ascriptions of witness confidence and actual witness accuracy. Because witnesses in the Silhouette RSA knew that only a shadow of themselves was being videotaped, it is possible that they somehow enhanced the way they verbalized their identifications decisions so as to overcome the lack of facial cues and convey as much information as possible. In retrospect, it would have been better to have videotaped witnesses with two cameras, one that recorded a normal view and another that recorded only a silhouette so that both Full and Silhouette RSA witnesses had identical experiences.

Was jurors' accuracy discrimination better for RSA than non-RSA witnesses under cross-examination? The second question of interest concerns the effect of the RSA experience on witnesses' subsequent cross-examination testimony: Did the cross-examination behavior of witnesses who experienced the RSA manipulation in Study 1 differ somehow so that jurors who saw only their cross-examination could discriminate accurate from inaccurate identifiers better than for control witnesses? Most important, did the benefit of the Full and Silhouette RSA on witnesses' confidence calibration transfer to their cross-examination behavior so that jurors' accuracy discrimination for those witnesses was improved?

Examination of Table 8 reveals that jurors' overall accuracy score for their cross-examination belief decisions (46.9%) was actually, although not signficantly, below chance performance. Similarly, when broken down by the witnesses' RSA experience in

Study 1, juror accuracy for their belief decisions of witnesses in any of the RSA conditions was not significantly better than chance. Finally, the combined accuracy score (50.0 %) for jurors' belief decisions of Full and Silhouette RSA witnesses (who showed a significant confidence/accuracy relation in Study 1) was no better than the accuracy score (43.8%) for jurors who rated Imagined RSA and control witnesses, $z = 0.709$, $p > .05$. This lack of a difference provides no support for the idea that witnesses who saw their own lineup behavior presented themsleves in a way that allowed to jurors to discriminate accurate from inaccurate identifiers any better than they could for witnesses who did not view themselves.

Recall the earlier dicussion that laid out the necessary requirements for the accuracy of jurors' belief decisions to benefit from a signficant confidence/accuracy relation at the witness level. If after some manipulation (e.g., RSA) accurate witnesses rate their confidence in their identification higher than that of inaccurate witnesses, an additional two correlational links must be obtained subsequently in order for jurors to believe accurate witnesses more than inaccurate ones. First, confident (and accurate) witnesses would have to be perceived by jurors as more confident than unconfident witnesses (r_{WJ}). Second, jurors would have to believe confident witnesses more than unconfident ones (r_{JB}). Overall in Study 2, there was a significant relation between witness confidence and witness accuracy for the 32 identifiers selected from Study 1, $r_{WA} = .490$, $n = 32$, $p < .05$. The first link, the relation between jurors' ascriptions of witness confidence under cross-examination and witness self-reported confidence, however, was not signficant, $r_{JW} = .202$, $n = 64$, $p > .05$. In other words, confident witnesses were not rated by jurors as more confident than their unconfident counterparts. The lack of signficance for this value is at odds with previous studies (Kassin, 1985; Wells et al., 1979; Wells et al., 1981) and no interpretation for the low value is attempted here. The second link, the relation between

jurors' ascriptions of confidence and their belief decisions, was signficant, $r_{JB} = .456$, $n = 64$, $p < .005$. This is consistent with previous eyewitness research, which demonstrated that jurors are much more likely to believe witnesses they rate as more confident compared to those they rate as less confident (Kassin, 1985; Wells et al., 1979; Wells, et al., 1981). Given the failure of the first link, then, perhaps it is not surprising that jurors did not reliably believe accurate witnesses more often than inaccurate ones. Despite demonstrating a significant relation between their self-reports of confidence and actual accuracy in Study 1, Full and Silhouette witnesses' calibrated confidence apparently was not robust enough to translate into reliably accurate belief decisions on the part of jurors.

Did viewing both the cross-exam and lineup behavior improve accuracy discrimination compared to cross-exam alone? The third, and perhaps most important question in Study 2 concerns whether or not allowing jurors to view witnesses' lineup identifications in addition to their cross-examination testimony improves jurors' discrimination ability. To answer this question, juror accuracy in the Cross-Exam Only condition was compared to juror accuracy in the Both condition. Jurors in the former group were only 46.9% accurate, but those in the latter were not much better at 56.2%, and neither was significantly better than chance. A z-test between the juror accuracy rates for the two groups indicated that there was no difference, $z = 1.05$, $p > .05$. It appears that despite being provided with additional information, which by itself enabled other jurors to discriminate accurate from inaccurate identifiers, jurors in the Both condition were no better at discriminating accuracy than were those who saw only the standard cross-exam. Although jurors who saw both tapes were somewhat (though not signficantly) better than those who saw only the cross-exam, it was expected that the benefit of viewing both tapes would be additive; in other words, those in the Both condition were expected to be at least as good as those in either of the other two groups, but with more information to use if they found it useful. Obviously, this was not the case. .

The role of decision time on judgments of accuracy. Jurors rated as more confident witnesses who took relatively short amounts of time on the picture they chose from the lineup, $r = -.218$, $n = 64$, one-tail $p < .05$. Perhaps this tendency for jurors to ascribe more confidence to short decision-time witnesses accounts for the low rate of juror accuracy because many of the accurate identifiers from Study 1 took a very long time to make their identification decision. It makes sense, but this idea is not supported because the correlation between jurors' belief decisions and witnesses' decision time was not significantly negative to account for the low rate of juror accuracy, $r = -.018$, $n = 64$, $p > .05$. So, jurors used decison times to ascribe more confidence to witnesses who took less time to make a decision, but something besides decision time influenced their belief decisions.

IV. STUDY 3

Study 3 was designed to investigate the causal relation between witnesses' impressions of their relative identification-decision time and their self-reported confidence in that decision. Witnesses in Study 1 demonstrated a significant inverse relation between their actual decision time and their reported confidence, as did witnesses in Kassin's (1985) studies. Similarly, participant-jurors in Study 2 rated as more confident witnesses who took relativley less time to make a decision. In none of these cases, however, was decision time inversely related to actual identification accuracy, so both witnesses and jurors were relying on a misleading cue to judge confidence. Is it the case that witnesses who take relatively little time to make an identification feel they are more accurate, or that witnesses who make faster decisions are more confident to begin with and that is why they took less time?

Study 3 investigates this question by providing witnesses with false feedback regarding their relative decision time; they were told that they took either a relatively short time, long time, about average, or were given no decision-time feedback. Actual decision-times were also recorded, as well as witnesses' confidence ratings, their estimated probability of accuracy, their judgment of task difficulty, and thier estimates of how long they took to make a decision, all after receiving the false decision-time manipulation. If witnesses are using decision time as a cue to accuracy, it is expected that those who believe they took relatively little time will indicate higher confidence compared to those who believe they took relatively longer to make a decision. In addition to the decision-time manipulation, Study 3 also addresses the issue of actual decision time and accuracy as a function of the presence or absence of the perpetrator in the lineup.

Method

Participants and design

Participants were 70 male and female students from the introductory psychology research pool at the University of Alberta. Each participant was given one credit for partial fulfillment of introductory psychology course requirements. There were four decision-time feedback conditions in the design: Short, Long, Average, and a control group which did not receive any feedback. In each condition, approximately half the witnesses viewed a perpetrator-present lineup and the other half saw a perpetrator-absent lineup. Participants were assigned to condition randomly in groups of two or three, with 18 (9 present lineups), 18 (9 present lineups), 17 (9 present lineups), and 17 (8 present lineups) participants in each of the Short, Long, Average, and control conditions, respectively.

Materials

The same crime video from Study 1 was used in Study 3. Again, participants viewed the approximately 2 min drama projected on a 150 cm screen from a Sony projection TV. The six-member suspect lineups were administered this time on videotape to 50 cm RCA color TV's. Photo prints of the slides used in Study 1 were videotaped under photographic-style lighting using the same Panasonic "Reporter" camera used to tape the crime video. The videos were about 10 min in length so that it would be highly unlikely for the tape to end before witnesses made their decisions, and in fact none of the witness took this long. The photos were numbered 1 through 6, in two rows of three pictures each. In the perpetrator-present case, the actual criminal from the video was in Position 3. In the absent case, the picture most similar to the actual ciminal was substituted in Position 3.

Lineup-decision times were recorded on digital clocks that were electrically triggered when the lineup came on the screen and electrically stopped when a witness made a lineup choice by pressing a button on a 7-button panel in another room. Decisions, either a button

from 1 through 6 for an identification or Button 7 for a rejection of the lineup, were indicated to the experimenter on a light panel. Decision-time feedback was delivered through an intercom system to participants' headphones. Witness responses of confidence, probability, willingness to testify, task difficulty, and estimated decision time were recorded on a response sheet (see Appendix 2).

Procedure

The procedure for Study 3 is basically the same as for Study 1, except that participants arrived at the laboratory in groups of two to three and watched the crime video together, and that a simultaneous rather than sequential lineup presentation was used. They were instructed not to communicate with each other in any way and none of them did. After watching the preliminary tennis tape and answering the associated value-laden questions about it, participants viewed the crime video and then completed a Self-Monitoring Questionnaire. Afterward, they were told that the study was actually about people's eyewitness testimony, or eyewitness identification, abilities exactly as the witnesses had been told in Study 1. They were told that they would be separated and then shown number of still-pictures of very similar-looking people on videotpae to see it they could recognize, or identify, the man from the crime tape. Participants were told how to indicate their lineup decision from their isolated cubicle as follows:

> Now, you're not going to see the pictures in here, but in just a minute I'm going to take down the hallway to your own separate cubicle with your own TV set, a pair of headphones, and a button panel with seven buttons on it. Once you're seated comfortably and your headphones are in place, six pictures numbered 1 through 6 will appear on your TV set. If you think you recognize one of the pictures as the man from the second videotape, then press the appropriate button on the button panel. If you think that the man from the tape is not among the pictures, then press button number 7. Sometime after you have made your decision, I will talk to you individually over your headphones. Please remain in your cubicle until I tell you what I would like you to do next.

Participants were then escorted to the cubicles. When they had their headphones in place and the intercom system was checked, the videotape containing either the present or absent lineup was played to the TV sets. Witnesses' lineup decisions were recorded as they were indicated on the experimenter's light panel connected to the participants' button panel and their decison times were frozen on a separate digital clock for each cubicle to be recorded later.

After everyone had made a decision, the lineup tape was stopped and the decision-time feedback was delivered over the intercom system. In all three experimental conditions, the feedback was delivered so that witnesses thought they were being spoken to individually, when in fact everyone heard the same message. In the Short condition, witnesses were told:

> O.K., now that *you've* made a decision from the pictures, I can tell you a bit about your choice. I can't tell you at this point if you are right or wrong in your decision, but I can tell you that of the hundred or so people who've seen the same set of pictures, you've taken about half as long as the average person in the study so far. For now I'd like you to stay in your cubicle until I come to see you in just a minute. Thank you.

Those in the Long condition heard exactly the same message except that they were told they had taken about twice as long as the average person in the study so far. Those in the average condition were told that they took about the same time as the average person in the study so far. And those in the control group were told merely that now that they had made a decision, the experimenter would be by shortly with the next task. Then the experimenter delivered a response sheet to each cubicle, explained the questions briefly, told the participant to take as much time as he or she needed, and to meet again in the main room when the response sheet was completed. There, witnesses were fully debriefed, questions were answered, credit was given, and participants were thanked for their cooperation.

Results and Discussion

Witnesses' estimates of decision time as a function of the feedback manipulation and the presence/absence variable are presented in Table 9. A two-way ANOVA with feedback and lineup type as the factors indicated that there were no differences in witnesses' time estimates as a function of the feedback manipulation, $F(3, 62) = 1.614$, $p > .05$. This is not too surprising, however, because the feedback manipulation was not expected to have an effect on time estimates, given that witnesses had nothing against which to compare their time estimates. It was for this reason that the feedback was manipulated in a qualitative (e.g., "you took about half/twice as long as the average person"), rather than a quantitative (e.g., "you took 30/60 seconds") manner in the first place. Of course, if the manipulation as administered had its expected impact, it would make sense for, say, witnesses in the Long condition to provide longer time estimates overall, but it is not essential for the purpose of the study; the effect on confidence is of interest.

The two-way ANOVA also did not indicate an interaction between the feedback and lineup-type variables, but there was a main effect for the latter factor, $F(1, 62) = 8.707$, $p = .004$. A subsequent analysis of covariance (ANCOVA), however, revealed that the actual amount of decision time accounted for the vast majority of the lineup-type effect. An initial two-way ANOVA with lineup-type and actual decision time as factors indicated that these two variables did not interact, $F(1, 66) < 1$, thereby satisfying the homogeneity of slopes assumption for the ANCOVA. The ANCOVA indicated that the there was no difference in estimated time as a function of lineup-type, $F(1, 67) < 1$, when the two groups were equated on the actual time dimension, but that actual time had a very large effect, $F(1, 67) = 96.520$, $p < .0001$. Put simply, witnesses who saw an absent lineup took much longer to make their decisions compared to those who saw a present lineup (Ms = 43.9 versus 22.6 secs) and that is why their time estimates were also much greater (Ms = 69.2 versus 32.8 secs).

Lineup decisions and confidence overall. Overall accuracy was only 51.4%, but most of the witnesses' errors were made with the perpetrator-absent lineup. When the perpetrator was among the pictures, 68.6% of the witnesses correctly identified him, but only 34.3% of the witnesses who viewed the blank lineup made the appropriate decision to not choose anyone from the pictures. Forty-seven percent of the inaccurate identifications from the blank lineup were of the substitute suspect. Therefore, when the diagnosticity ratio for identifications was calculated with only these substitute-identifications in the equation, the resulting value was .686 / .314 = 2.183. The ratio for nonidentifications was .343 / .200 = 1.715.

Similar to Study 1 and the majority of previous eyewitness studies, the point-biserial correlation bewtween confidence and accuracy was significant, albeit modest, $r = .227$, $n = 70$, $p < .05$. The coresponding t-test between the confidence means for accurate and inaccuarate witnesses (Ms = 4.9 and 4.3) was therefore also significant, $t(69) = 1.92$, $p < .05$.

Decision-time feedback and self-reported confidence. Witnesses' confidence ratings as function of feedback condition and lineup type are presented in Table 10. A two-way ANOVA on confidence with feedback condition and lineup-type as the factors revealed that there was no main effect for feedback, but that there was a significant main effect for lineup-type, $F(1, 62) = 5.212$, $p = .026$, indicating that witnesses who saw a present lineup were more confident compared to those who saw an absent lineup (Ms = 5.0 vs. 4.2). The rate of choosing did not differ as a function of lineup type, so it is not the case that choosers were merely more confident than nonchoosers. It is true, however, that some of the difference in confidence as a fuction of lineup type is due to the fact that many more people who saw a present lineup made an accurate decision compared to those who saw an absent lineup (69% vs. 34%) and, since accurate witnesses were more confident than innacurate ones (Ms = 4.9 vs. 4.3), it makes sense that confidence was

higher in the perpetrator-present condition than in the perpetrator-absent condition. The accuracy difference as a function of lineup type, however, does not account completely for the confidence difference as a function of lineup type because accurate witnesses who saw an absent lineup (accurate rejectors) were not more confident than inaccurate ones (Ms = 4.2 vs. 4.3); most of the accuracy effect on confidence is due to the large difference in confidence in the perpetrator-present condition between accurate and inaccurate witnesses (Ms = 5.3 vs. 4.3). Although a two-way ANOVA on confidence with accuracy and lineup type as the factors did not indicate an interaction of this type, $F(1, 66) = 2.478$, $p > .05$, a subsequent contrast indicated that accurate witnesses who saw a present lineup were significantly more confident than witnesses in the other three cells, $F(1, 66) = 8.820$, $p = .004$, but that witnesses in these latter three groups did not differ from each other.

Most relevant for the present study is that there was a significant interaction between feedback condition and lineup-type on confidence, $F(3, 62) = 4.323$, $p = .008$. Witnesses in the Short condition who saw a present lineup were much more confident than witnesses in the Short condition who saw an absent lineup, $F(1, 62) = 17.973$, $p < .001$. In addition, Short condition witnesses who saw a present lineup were more confident than all other witnesses who also saw a present lineup, $F(1, 62) = 7.142$, $p = .01$. Conversely, Short condition witnesses who saw an absent lineup were much less confident than all other witnesses who also saw an absent lineup, $F(1, 62) = 5.436$, $p = .023$.

Consistent with the overall analyses, most of the difference in confidence between witnesses who saw a present lineup and those who saw an absent lineup in the Short condition was due to the fact that there were many more accurate witnesses in the former group compared to the latter (78% vs. 33%) and that accurate witnesses were more confident than inaccurate ones overall. It was not the case, however, that inaccurate witnesses were less confident than accurate ones in just the absent group (Ms = 3.0 vs. 3.5), but the number of accurate rejectors was so low ($n = 3$) that the means are not very

reliable. The possibility remains, however, that something about the absent lineup besides accuracy or rate of choosing caused people to rate their confidence differently in reaction to the Short decision feedback compared to people who saw a present lineup. In other words, it may be the case that, regardless of whether or not witnesses in the perpetrator-absent who received Short decision time feedback identified someone from the pictures, and regardless of whether or not they were accurate in whatever decision they made, they rated their confidence lower than witnesses who saw a present lineup. It may be that those who saw the absent lineup reasoned that their decision was made too hastily, given that they either rejected the lineup or perhaps reluctantly identified the wrong person. On the other hand, those who saw the present lineup may have been encouraged to hear that their decision was made so quickly since most of them accurately chose the perpetrator from the lineup.

Actual decision time and accuracy. Consistent with the three of the four Kassin (1985) studies in which a simultaneous lineup presentation was used, accurate decisions did not take more or less time overall compared to inaccurate ones (Ms = 32.8 and 33.7, respectively, $r = -.015$, $n = 68$, $p > .05$). Unlike Kassin's work, however, the present data can be broken down by considering the effect of perpetrator-present versus perpetrator-absent lineups on time and accuracy. The actual decision times are presented in Table 11 as a function of lineup-type and accuracy, and further described as either false rejections (i.e., not identifying anyone when the perpetrator is present), hits (i.e., identifying the perpetrator), false identifications (i.e., identifying the *substitute* from a perpetrator-absent lineup), and correct rejections (i.e., not identifying anyone when in fact the perpetrator is not present).

A two-way ANOVA with lineup-type and accuracy as the factors was conducted. There was not a main effect for accuracy, $F(1, 50) < 1$, but there was a main effect for lineup-type. Witnesses who saw a perpetrator-present lineup took much less time to make a

decision compared to those who saw an absent lineup, F (1, 50) = 4.92, p = .031. This effect was qualified, however, by a significant interaction, F (1, 50) = 5.993, p =.018. A simple effects analysis indicated that correct rejectors took more time than all other witnesses to make a decision, but that hits, misses, and inaccurate rejections did not differ from each other. Although the cell sizes are unequal and not large, this result suggests that accurate decisions to not identify anyone from the absent lineup took more time than did inaccurate decisions to not identify anyone from the present lineup. Another, and perhaps more intuitive result was that accurate rejections took longer than accurate identifications.

V. GENERAL DISCUSSION

These three studies were designed to investigate whether or not eyewitness lineup behavior provides cues to the accuracy of witnesses' lineup decisions. The purpose of Study 1 was to investigate the basis on which people judge their own confidence as they watch themselves on videotape making an eyewitness identification. Kassin (1985) demonstrated that witnesses exposed to a tape of their identification task showed a signficant relation between their self-reported confidence and their actual identification accuracy, whereas those in a control group did not. In Study 1, some witnesses saw a tape similar to what witnesses in the Kassin studies saw; in addition, however, some witnesses saw only a silhouette of themselves on tape, some were merely instructed to recreate in their minds what they might look like if they were to see themselves on tape, and still others were given no post-lineup manipulation at all. The results showed that witnesses in both the Full and Silhouette RSA groups demonstrated a significant relation between their confidence and accuracy, but that witnesses in the Imagined RSA and control groups did not.

The fact that witnesses' confidence ratings in the Silhouette RSA condition benefited from viewing the tape of their lineup task does not explain what cues people may be using, but it does eliminate a wide range of plausible facial information as being essential for such judgments. In addition, the fact that witnesses' confidence ratings in the Imagined RSA condition were no better calibrated than were those in the control group suggests that merely simulating the task privately in their minds was not sufficient to improve their confidence calibration. In between these two explanations (facial expressions on the end, mental rehearsal on the other) remain the relatively public verbal and decision-time information available on both the Full and Silhouette RSA tapes.

In Study 2, a subset of those witnesses who made an identification from the lineups in

Study 1 were viewed on videotape and rated by a new group of participant-jurors. Some jurors saw only witnesses' lineup behavior, some saw only witnesses undergoing cross-examination, and others saw both the witnesses' lineup and cross-examination. Three questions were of major interest. First, in the lineup-only condition, how would jurors' exposure to witnesses in normal video versus silhouette-only format affect the correlation between jurors' ascriptions of witness confidence and actual witness accuracy? The results showed that jurors who saw the normal videotapes did not rate witnesses who had made accurate identifications in Study 1 as more confident than inaccurate ones. Jurors who saw witnesses only in silhouette format, however, showed a significant correlation between their ascriptions of confidence and actual witness accuracy.

Study 1 and the first question of concern in Study 2, then, made some progress in understanding how the RSA effect on witnesses' confidence calibration was achieved by Kassin (1985). Assuming that the Silhouette RSA videotape contained only verbal information, and that decision time was not reliably related to accuracy, the fact that (1) witnesses in both the Full and Silhouette RSA conditions showed signficant confidence calibration, and (2) observers ascribed more confidence to accurate compared to inaccurate witnesses when they saw the Silhouette tape, suggests that something about the witnesses' verbal utterances provides a cue to identification accuracy. This interpretation is further supported by the fact that observers did not ascribe more confidence to accurate compared to inaccurate witnesses when they saw the Full RSA tape; although the Full RSA tape also contains the allegedly useful verbal information, perhaps it was overshadowed by the more salient, yet uninformative, facial information also available in the complete audiovisual presentation. This latter point has been discussed earlier as being conistent with Ekman's (e.g., 1985; 1988) research on audio-only, video-only, and audiovisual presentations of communicators and observers' ability to detect deceit. Ekman's central thesis is that there are communicative channels that contain useful information on which observers can base

their judgments of communicator deciet, but that these channels may not be attended to if other more salient, yet irrelevant, information is also made available. Of course deceit-detection is not the same as accuracy-assessment, but it is appealing to think that similar principles may apply in both circumstances.

The second question of interest in Study 2 concerned the potential benefit of an improved confidence/accuracy relation at the witness level on the accuracy of jurors' belief decisions made on the basis of the witnesses' cross-examination testimony. Because witnesses in the Full and Silhouette RSA conditions is Study 1 showed a signficant confidence/accuracy, but those in the Imagined RSA and control groups did not, it was expected that jurors' belief decisions for witnesses in the former two groups would be more accurate than those for witnesses in the latter two groups. This hypothesis was not suppported.

The third question of interest in Study 2 concerned the probative value of allowing jurors to view videotapes of witnesses' lineup identification tasks in addition to viewing the witnesses' cross-examination testimony. It was hypothesized that the accuracy of jurors' belief decisions who saw both components of the witnesses' behavior would be better than those of jurors who saw only the cross-exam testimony. This hypothesis was also not supported. Jurors who saw both tapes were only slightly, and not significantly, more accurate than were those who saw only the cross-exam.

The results of Study 3 provide good support for the idea that there is a causal relation between witnesses' impressions of their relative decision time and their subsequent self-reported confidence, rather than the mere correlational evidence that was previously available. Although it may still be the case that confident people take less time to make a decision compared to unconfident ones, the results showed that of the people who were told they had made their decision quickly, those who did not know they were accurate rated their confidence higher than people who did not know they were inaccurate. A more

interesting argument based on the results from very small samples is that witnesses who saw a perpetrator-absent lineup, regardless of their decision accuracy, rated themselves as lower in confidence compared to those who saw a perpetrator-present lineup. In addition, the actual time data in Study 3 showed that people took more time to accurately reject an absent lineup than they did to inaccurately reject the lineup when the perpetrator was present. Decision times for inaccurate versus accurate identifiers did not differ.

Legal-policy implications

Overall, the results of the present studies provide little support for the idea that videotapes of lineup identifications are forensically useful to improve judgments of identification accuracy. Of all the questions considered in the present research, there were two hypotheses most relevant to the courts. First, it was predicted that witnesses' whose confidence and identification accuracy were signficantly correlated as a result of the RSA manipulation in Study 1 would be rated more accurately by jurors who saw only their cross-examination testimony is Study 2. This was not supported. The present research does not, therefore, form the basis for a general proposal regarding the benefit of witness-level manipulations that improve witnesses' confidence calibration.

The second hypothesis of relevance to the courts predicted that jurors exposed to videotapes of witnesses' lineup-identification tasks in addition to the standard cross-exam testimony would be better able to discriminate accurate from inaccurate identifiers compared to those jurors who saw only the same witnesses' cross-exam. Jurors in the former group did show a slight advantage, and the addition of the lineup tape certainly did not impair their discrimination ability, but again there is no basis for a recommendation to include such tapes in actual court situations.

Suggestions for further research

How does RSA work? How might the verbal-channel interpretation of the RSA effect be further investigated by future research? The first step would be to provide

observers with an audio-only tape of the witnesses' lineup task to make sure that the silhouette videotapes did in fact contain nothing more than verbal information. The next step would be to investigate another assumption regarding the content of the Silhouette tapes that has already been discussed; namely, that the actual verbal *content* of what the witnesses said probably did little to distinguish accurate from inaccurate identifiers. To support the interpretation that it is the verbal *intonation* associated with accurate and inaccurate choices that allows both witnesses and observers to discriminate accuracy, the content of the audio-tape could be masked so that only the intonation characteristics of the verbal channel could be heard. Several investigators have utilized content-filtering techniques to determine the information available in verbal intonation only. Soskin and Kaufman (1961) and Starkweather (1956), for example, filtered all but very low frequency speech-energy from verbal recordings and found that listeners could discern the speakers' emotional expression. Scherer (1982) has described advances in these filtering methods in a handbook chapter devoted to methodology in nonverbal research.

The present studies are predicated on the assumption that lineup behavior provides cues to identification accuracy that are common to most people, in the same way that research on detection of deceit has described behavioral changes shared by nearly all people when telling a lie. It may be the case, however, that some people and/or some situations provide more information than others, or that some people are better than others at decoding cues. Further research could investigate variables that moderate the RSA effect.

In keeping with a renewed interest among psychological researchers on individual differences and on mulitcausal models of behavior, researchers could further investigate personality variables that may influence any advantage of viewing their own behavior on videotape . As discussed earlier, Kassin (1985) tested some variables (e.g., public vs. private self-consciousness) and Study 1 of the present research investigated the possibility that Snyder's (e.g., 1979) self-monitoring construct moderates how much people learn

from watching their own behavior. There are many other plausible candidates to be tested. Subseqent studies could also look at the role of personality variables (like self-monitoring) on *observers*' abilities to distinguish accuracy-behavior cues from other cues unrelated to accuracy.

With respect to situational factors that may moderate the RSA effect, there are interesting alternative interpretations regarding the fact that witnesses' confidence calibration improved after watching either their own Full or Silhouette RSA tape but observers' asciptions of confidence were correlated with actual accuracy only in the Silhouette RSA condition. People are more experienced and familiar with their own behavior than are observers, especially if the observers are strangers, and perhaps that is why witnesses appeared to learn more than the observers by watching their Full RSA tapes. The potential advantage of familiarity with one's own behavior compared to a stranger's could be tested by designing a study in which people rated themselves, a friend, and a stranger on videotape. A familiarity-advantage hypothesis predicts that people would be better at judging the behavior of themselves and their friends compared to a stranger. In fact, people may be better at judging a close friend compared to themselves because they see their friend's behavior more than their own.

In addition to the idea that people merely are more experienced and familiar with their own behavior compared to observers, recent evidence shows that people may have privileged self-knowledge that allows them to access *reasons* for their own behavior that are not apparent to observers (Gavanski & Hoffman, 1987). Observers, then, might have been restricted to a smaller set of less reliable inferences based only on ideas about behavioral cues that are related to accuracy for most people, whereas actors (i.e., witnesses themselves) could rely on a more ideographic conception of themselves.

How might the benefit of exposure to lineups be enhanced? Athough witnesses' confidence calibration was signifcantly improved by the RSA manipulation, the

effect was not extremely large, and exposing jurors to the witnesses' lineup behavior in addition to just their cross-examination showed no effect at all. Under what conditions might exposure to lineup behavior be expected to produce more of an effect in both of these situations? The most obvious variable is the length of the interval between a witness' identification and the point at which she undergoes cross-examination. It is not uncommon in actual criminal investigations for witnesses to testify about their identification several months after actually choosing a suspect from a lineup. During that time, witness confidence is influenced by both internal factors, like commitment to a decision and a desire to appear credible as well as external factors like police pressure and attorney briefings on how best to behave on the stand. Under these circumstances, both witnesses and jurors would be most likely to benefit from viewing a tape of the witnesses' initial lineup identification; RSA witnesses may behave in accordance with the confidence they felt after their identification and jurors benefit by seeing the witness before any post-identification confidence boosting has taken place.

From a theoretical perspective, following up on the suggestion for further regarding degradation of witnesses' verbal statements should improve our understanding of whether or not there are cues other than confidence that may be useful indicators to eyewitness identification accuracy. Is there any basis for the notion that accurate memories for faces are distinguishable from inaccurate ones, as has been demonstrated with people's lengthy verbal statements describing real and imagined experiences? From an applied perspective, is there any probative value in exposing jurors to videotapes of eyewitnesses' previous lineup behavior in actual trial settings? It is an intuitively appealing possibility. Furthermore, even if there is no improvement in jurors' accuracy assessment per se, merely videotaping suspect lineup administrations may serve to help control and standardize the procedure across jurisdictions and eliminate errors that may lead to false accusations and even unwarranted convictions of innocent people.

References

Adams, J. K., & Adams, P. A. (1961). Realism of confidence judgments. Psychological Review, 63, 33-45.

Bauchner, J. E., Kaplan, E. P., & Miller, G. R. (1980). Detecting deception: The relationship of available information to judgmental accuracy in initial encounters. Human Communication Research, 6, 251-264.

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

Blake, M. (1973). Prediciton of recognition when recall fails: Exploring the feeling-of-knowing phenomenon. Journal of Verbal Learning and Verbal Behavior, 12, 311-319.

Borchard, E. M. (1932). Convicting the innocent: Errors of criminal justice. New Haven: Yale University Press.

Bothwell, R. K., Deffenbacher, K., & Brigham, J. C. (1986). Correlation of eyewitness accuracy and confidence: The optimality hypothesis revisited. Journal of Applied Psychology, 72, 691-695.

Brandon, R., & Davies, C. (1973). Wrongful imprisionment: Mistaken convictions and their consequences. London: George Allen & Unwin Ltd.

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. & Cairns, D. L. (in press). The effect of mugshot inspections on eyewitness identification accuracy. Journal of Applied Social Psychology.

Brigham, J. C., & Wolfskeil, M. P. (1983). Opinions of attorneys and law enforcement personnel on the accuracy of eyewitness identifications. Law and Human Behavior, 7, 337-349.

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

Brown & McNeil (1966). The tip-of-the-tongue pehnomenon. Journal of Verbal Learning and Verbal Behavior, 5, 325-337.

Buckout, R. (1974). Eyewitness testimony. Scientific American, 231, 23-31.

Buckout, R. (1976). Psychology and eyewitness identification. Law and Psychology Review, 2, 75-92.

Cutler, B. L., & Penrod, S. D. (1988). Improving the reliability of eyewitness identifications: Lineup construction and presentation. Journal of Applied Psychology, 73, 281-290.

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

Deffenbacher, K. A., Brown, E. L., & Sturgill, W. (1978). Some predictors of eyewitness memory accuracy. In M. M. Gruneberg, P. E. Morris, & R. N. Sykes (Eds.) Practical Aspects of Memory. London: Academic Press.

DePaulo, B. (1985). Telling ingratiating lies: Effects of target sex and target attractiveness on verbal and nonverbal decpetive success. Journal of Personality and Social Psychology, 36, 1302-1309.

Devlin, Honorable Lord P. (1976). Report to the Secretary of sate for the Home Department of the departmental committee on evidence of identification in criminal cases. London: Her Majesty's Stationary Office.

Duval, S., & Wicklund, R. A. (1972). A theory of objective self-awareness. New York: Academic Press.

Ekman, P. (1965). Differential communication of affect by head and body cues. Journal of Personality and Social Psychology, 2, 726-735.

Ekman, P. (1985). Telling lies: Clues to deceit in the marketplace, politics, and marriage. New York: W. W. Norton & Co.

Ekman, P. (1988, June). Behavioral indices of deception. Paper presented at the NATO Advanced Study Institute on Credibility Assessment, Maratea, Italy.

Ekman, P., & Friesen, W. V. (1967). Head and body cues in the judgment of emotion: A reformulation. Perceptual and Motor Skills, 24, 711-724.

Ekman, P., & Friesen, W. V. (1969). Nonverbal leakage and clues to deception. Psychiatry, 32, 88-106.

Ekman, P., & Friesen, W. V. (1974). Detecting deception from the body or face. Journal of Personality and Social Psychology, 29, 288-298.

Ericsson, K. A., & Simon, H. A. (1980). Verbal reports as data. Psychological Review, 87, 215-251.

Eysenck, M. W. (1979). The feeling of knowing a words' meaning. British Journal of Psychology, 70, 243-251.

Fenigstein, A., Scheier, M. F., & Buss, A. H. (1975). Public and private self-consciousness: Assessment and theory. Journal of Consulting and Clinical Psychology, 43, 522-527.

Fischhoff, B., Slovic, P., & Lichtenstein, S. (1977). Knowing with certainty: The appropriateness of extreme confidence. Journal of Experimental Psychology: Human Perception and Performance, 3, 552-564.

Fleet, M. L., Brigham, J. C., & Bothwell, R. K. (1987). The confidence-accuracy relationship: The effects of confidence assessment and choosing. Journal of Applied Social Psychology, 17, 171-187.

Frankfurter, F. (1927). The case of Sacco and Vanzetti. New York: Little Brown.

Gangestad, S., & Snyder, M. (1985a). "To carve nature at its joints": On the existnece on discrete classes in personality. Psychological Review, 92, 317-349.

Gangestad, S., & Snyder, M. (1985b). On the nature of self-monitoring: An examination of latent causal structure. In P. Shaver (Ed.), Self, situations, and social behavior. Beverly Hills: Sage Publications.

Gavanski, I., & Hoffman, C. (1987). Awareness of influences on one's own judgments: The orle of covariation detection and attention to the judgment process. Journal of Personality and Social Psychology, 52, 453-463.

Glenberg, A. M., & Epstein, W. (1985). Calibration of comprehension. Journal of Experimental Psychology: Learning, Memory, & Cognition, 11, 702-718.

Glenberg, A. M., & Epstein, W. (1987). Inexpert calibration of comprehension. Memory and Cognition, 15, 84-93.

Goldstein, A. G., Johnson, K. S., & Chance, J. E. (1981). Does fluency of face description imply superior face recognition? Bulletin of the Psychonomic Society, 13, 15-18.

Gruneberg, M. M., Monks, J., & Sykes, R. N. (1977). Some methodological problems with felling of knowing studies. Acta Psychologia, 41, 365-371.

Hart, J. T. (1965). Memory and the feeling-of-knowing experience. Journal of Educational Psychology, 56, 208-216.

Hart, J. T. (1967). Memory and the memory-monitoring process. Journal of Verbal Learning and Verbal Behavior, 6, 685-691.

Henmon, V. A. C. (1911). The relation of the time a judgment to its accuracy. Psychological Review, 18, 186-201.

Hocking, J. E., Bauchner, J. E., Kaminski, E. P., & Miller, G. R. (1979). Detecting deceptive communication from verbal, visual, and paralinguistic cues. Human Communication Research, 6, 33-46.

Hosch, H. M. (1980). Commentary: A comparison of three studies of the influence of expert testimony on jurors. Law and Human Behavior, 4, 297-302.

Hosch, H. M., & Platz, J. J. (1984). Self-monitoring and eyewitness accuracy. Personality and Social Psychology Bulletin, 10, 289-292.

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

Hosch, H. M., Beck, E. L., & McIntyre, P. (1980). Influence of expert testimony regarding eyewitness accuracy on jury decisions. Law and Human Behavior, 4, 287-296.

Johnson, M. K., & Raye, C. L. (1981). Reality monitoring. Psychological Review, 88, 67-85.

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

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

Krafka, C. & Penrod, S. (1985). Reinstatement of context in a field experiment on eyewitness identifications. Journal of Personality and Social Psychology, 49, 58-69.

Lichtenstein, S., & Fischhoff, B. (1977). Do those who know more also know more about how much they know? Organizational Behavior and Human Performance, 20, 159-183.

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

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, 66, 7[illegible]-80.

Lichtenstein, S., & Fischhoff, B. (1977). Do those who know more also know more about how much they know? Organizational Behavior and Human Performance, 20, 159-183.

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

Lindsay, R. C. L. & Wells, G. L. (1985). Improving eyewitness identifications from lineups: Simultaneous versus sequential lineup presentation. Journal of Applied Psychology, 80, 556-564.

Lindsay, R. C. L. (1986). Confidence and accuracy of eyewitness identifications from lineups. Law and Human Behavior, 10, 229-240.

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

Littlepage, G. E., & Pineault, M. A. (1979). Detection of deceptive facutal statements from the body and face. Personality and Social Psychology Bulletin, 5, 325-328.

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). Eyewitnesses: Essential but unreliable. Psychology Today, Feb., 22-26.

Maier, N. R. F. & Thurber, J. A. (1968). Accuracy of judgment of deception when an interview is watched, heard, and read. Personnel Psychology, 21, 23-30.

Malpass, R., & Devine, P. (1981). Eyewitness identication: Lineup instructions and absence of the offender. Journal of Applied Psychology, 66, 482-489.

Miller, G. R., & Burgoon, J. K. (1982). Factors affecting assessments of witness credibility. In N. L. Kerr & R. M. Bray (Eds.) The psychology of the courtroom. New York: Academic Press.

Munsterberg, H. (1908). On the witness stand: Essays on psychology and crime. New York: Clark Boardman.

Murray, D, M., & Wells, [illegible] L. (1982). Does knowledge that a crime was staged affect accuracy? Journal of Applied Social Psychology, 12, 42-53.

Neil v. Biggers, (1972) 409 U.S. 188.

Nelson, T. O. (1984). A comparison of current measures of the accuracy of feeling-of-knowing predictions. Psychological Bulletin, 95, 109-133.

Nelson, T. O., & Narens, , L. (1980). A new technique for investigating the feeling of knowing. Acta Psychologica, 46, 69-80.

O'Connor, J. (1974). "That's the man": A sobering study of eyewitenss identification and the polygraph. St. John's Law Review, 1, 27-29.

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

Schacter, D. L. (1983). Feeling of knowing in episodic memory. Journal of Experimental Psychology: Learning, Memory, & Cognition, 9, 39-54.

Scherer, K. R. (1982). Methods of research on vocal communication: Paradigms and paramters. In K. R. Scherer & P. Ekman (Eds.), Handbook of methods in nonverbal behavioral research. Cambridge: Cambridge University Press.

Schooler, J. W., Gerhard, D., & Loftus, E. F. (1986). Qualities of the unreal. Journal of Experimental Psychology: Learning, Memory, & Cognition, 12, 171-181.

Slovic, P. (1972). Psychological study of human judgment: Implications for investment decision making. Journal of Finance, 27, 779-799.

Snyder, M. (1979). Self-monitoring processes. In L. Berkowitz (Ed.), Advances in experimental social psychology (Vol. 12, pp. 85-128). New York: Academic Press.

Soskin, W. F., & Kaufman, P. E. (1961). Judgment of emotion in word-free voice samples. Journal of Communication, 11, 73-80.

Starkweather, J. A. (1956). The communication value of content-free speech. American Journal of Psychology, 69P, 121-123.

Stern, W. (1904). Realistic experiments. In U. Niesser (Ed. and Trans.), Memory Observed (pp. 95-108). San Francisco: Freeman & Co. (Reprinted from Beitrage zur Psychologie der Aussage, 2, 1-31)

Stern, L. W. (1939). The psychology of testimony. Journal of Abnormal and Social Psychogical, 34, 3-20.

Tulving, E., & Madigan, S. A. (1970). Memory and verbal learning. Annual Review of Psychology, 21, 437-484.

Undeutsch, U. (1982). Statement reality analysis. In A. Trankell (Ed.), Reconstructing the past: The role of psychologists in criminal trials. Stockholm: Norsted.

Underwood, B. J. (1969). Attributes of memory. Psychological Review, 76, 559-573.

Wells, G. L. (1984). The psychology of lineup identifications. Journal of Applied Psychology, 14, 89-103.

Wells, G. L. (1986). Expert psychological testimony: Empirical and conceptual analyses of effect. Law and Human Behavior, 10, 83-96.

Wells, G. L., & Lindsay, R. C. L. (1985). Methodological notes on the accuracy-confidence relationship in eyewitness identifications. Journal of Applied Psychology, 64, 440-448.

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, 688-696.

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

Wells, G. L., Lindsay, R. C. L., & Tousignant, J. P. (1980). Effects of expert psychological advice on human performance in judging the validity of eyewitness testimony. Law and Human Behavior, 4, 275-286.

Wells, G. L., & Murray, D. M. (1983). What can psychology say about the Neil vs. Biggers criteria for judging eyewitness identification 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. Cambridge: Cambridge University Press.

Wells, G.L. & Turtle, J.W. (1986). Eyewitness identification: The importance of lineup models. Psychological Bulletin, 99, 320-329.

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

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

Wolfskiel, M. P., & Brigham, J. C. (1985). The relationship between accuracy of prior description and facial recognition. Journal of Applied Psychology, 70.

Yarmey, A. D., & Jones, H. P. T. (1983). Is the psychology of eyewitness identification a matter of common sense? In S. M. A. Lloyd-Bostock & B. R. Clifford (Eds.), Evaluating witness evidence. Chichester: Wiley.

Yuille, J., Raskin, D., & Steller, M. (in press). Statement validity analysis.

Zajonc, R. B. (1980). Feeling and thinking: Preferences need no inferences. American Psychologist, 35, 151-175.

Table 1

Breakdown of Lineup Decision Data for Study 1

		Decision Type			
Lineup Type	n	Accurate IDs	Inaccurate IDs	Accurate Rejections	Inaccurate Rejections
Perpetrator present	50	20	6	*	24
Perpetrator absent	39	*	10	29	*
Overall	89	20	16	29	24

*Indicates that such a decision cannot occur.

Table 2

Lineup Type, Gender, Choosing Rates, Accuracy, and Decision Time Data for Study 1 Broken Down by RSA Condition

	RSA Condition				
Item	Full	Silhouette	Imagined	Control	Total or M
n	23	19	22	25	89
Males	7	11	8	9	35
Females	16	8	14	16	54
Perpetrator present lineups	14	12	12	12	50
Perpetrator absent lineups	9	7	10	13	39
Number of IDs	11	8	9	8	36
Accurate	7	4	5	4	20
Inaccurate	4	4	4	4	16
Overall Accuracy (%)	61	47	55	56	55
Average Decision Time (secs)	25.4	22.7	34.1	27.1	27.4

Table 3

Confidence, Probdbility, Difficulty, and Estimated Decision Time Data for Study 1 Broken Down by RSA Condition

Item	RSA Condition				
	Full	Silhouette	Imagined	Control	Mean
n	23	19	22	25	
Confidence (1 - 7)	5.1	5.2	5.0	4.6	5.0
Est. probability (0 - 100%)	73.2	72.3	67.4	64.9	69.2
Judged difficulty (1 - 7)	4.6	4.7	4.5	4.2	4.5
Estimated decision time (secs)	67.4	60.1	83.4	69.2	70.4

Table 4

Witnesses' Self-Reported Confidence and Estimated Probability of Accuracy in Study 1 as a Function of RSA Condition and Actual Identification Accuracy

RSA and Dependent Measure	Identification Accuracy	
	Inaccurate	Accurate
Full RSA		
n	9	14
Confidence	4.6	5.5
Probability	61.0	81.0
Silhouette RSA		
n	10	9
Confidence	4.6	5.8
Probability	65.5	79.9
Imagined RSA		
n	10	12
Confidence	4.9	5.1
Probability	66.0	68.6
Control		
n	11	14
Confidence	4.7	4.5
Probability	70.0	60.9
Mean		
n	40	49
Confidence	4.7	5.2
Probability	65.9	72.0

Table 5

Mean Lineup-Decision Times (in secs) for Witnesses in Study 1

Decision Type	Decision Accuracy	
	Accurate	Inaccurate
Identifiers	(n = 20)	(n = 16)
Average time on target (Acc) or foil (Inacc)	12.6	6.9
Average time on all other pictures	4.3	4.5
Nonidentifiers	(n = 29)	(n = 24)
Average time on substitute (Acc) or target (Inacc)	5.1	4.7
Average time on all other pictures	3.7	2.9

Table 6

Correlations Between Decision Times and Confidence and Accuracy for Witnesses in Study 1

	1	2	3	4
1. Total decision time	x	.236*	-.259**	.452**
2. Actual accuracy		x	.179*	.190*
3. Self-reported confidence			x	-.101
4. Self-estimated decision time				x

n = 79

* one-tail $p < .05$

** one-tail $p < .01$

Table 7

Correlations Between Jurors' Ascriptions of Witness Confidence and Actual Witness Accuracy on the Basis of Lineup Behavior Only as a Function of Video Type in Study 2

RSA Type	n	Correlation
Overall	64	.108
Normal (Full, Imagined, & Control)	48	.013
Full RSA	16	.221
Imagined RSA	16	-.042
Control	16	-.116
Silhouette	16	.444*

* $p < .05$

Table 8

Juror Accuracy and Ratings of Witness Confidence as a Function of Juror Condition in Study 2 and Witness RSA Condition in Study 1

	Juror Condition			
Witness RSA Condition	Lineup Only	Cross-Exam Only	Both	Mean
Full RSA				
Juror accuracy (%)	56.2	68.7	50.0	58.3
Ascribed witness confidence	3.6	4.8	4.5	4.3
Silhouette RSA				
Juror accuracy	62.5	31.3	50.0	47.9
Ascribed witness confidence	5.1	4.9	4.8	4.9
Imagined RSA				
Juror accuracy	75.0	31.3	56.2	54.2
Ascribed witness confidence	3.9	4.7	4.4	4.3
Control				
Juror accuracy	50.0	56.2	68.7	58.3
Ascribed witness confidence	4.6	4.4	4.3	4.4
Mean				
Juror accuracy	60.9	46.9	56.2	54.7
Ascribed witness confidence	4.3	4.7	4.5	4.5

Table 9

Witnesses' Estimated Decision Time (in seconds) in Study 3 as a Function of Decision-Time Feedback and the Presence or Absence of the Perpetrator in the Lineup

Lineup-type	Decision-Time Feedback				
	Long (n = 18)	Short (n = 18)	Average (n = 17)	Control (n = 17)	Mean (n = 70)
Perpetrator-present lineups	43.6	16.8	39.3	31.3	32.7
Perpetrator-absent lineups	101.9	57.4	54.4	61.4	69.1
Mean	72.7	33.2	34.1	26.4	51.0

Table 10

Confidence Ratings (from 1 = not at all to 7 = extremely) for Witnesses in Study 3 as a Function of Feedback Condition and the Presence or Absence of the Perpetrator in the Lineup

	Decision-Time Feedback				
Lineup-type	Long (n = 18)	Short (n = 18)	Average (n = 17)	Control (n = 17)	Mean (n = 70)
Perpetrator-present lineups	4.9_a	6.0_b	4.2_a	4.8_a	4.8
Perpetrator-absent lineups	4.6_a	3.3_c	4.5_a	4.6_a	4.3
Mean	4.7	4.7	4.4	4.6	4.6

NOTE: Values with different subscripts differ at the .05 level.

Table 11

Mean decision times for witnesses in Study 3 as a function of decision accuracy and the presence or absence of the perpetrator in the lineup.

	Decision accuracy	
Lineup-type	Inaccurate	Accurate
Perpetrator-present	False Rejections	Hits
	30.2_a ($n = 7$)	20.9_a ($n = 24$)
Perpetrator-absent	Substitute IDs	Accurate Rejections
	28.4_a ($n = 11$)	57.5_b ($n = 12$)

NOTE: Values with different subscripts differ at the .01 level.

APPENDIX 1

ATTITUDE INVENTORY

General Instructions

The following statements concern your personal reactions to a number of different situations. No two statements are exactly alike, so consider each statement carefully before answering. If a statement is TRUE or MOSTLY TRUE as it applies to you, then circle the "T" next to that statement. If a statement is FALSE or NOT USUALLY TRUE as it applies to you, then circle the "F" next to that statement.

Please remember that your answers to these questions are completely confidential. We are interested in group responses, so you need not even fill in your ID number on the answer sheet.

Please try to answer all of the questions, but if you find a question personally objectionable, you do not have to answer it.

Do not spend too much time on any one question, we are interested in your "first impressions".

There are no right or wrong answers; we are interested only in what you think.

PLEASE TURN NOW TO THE NEXT PAGE . . . .

REMEMBER: If a statement is TRUE or MOSTLY TRUE, then circle the "T". If a statment is FALSE or NOT USUALLY TRUE, then circle the "F".

(1) T F I find it hard to imitate the behavior of other people.

(2) T F At parties and social gatherings, I do not attempt to do or say things that others will like.

(3) T F I can only argue for ideas which I already believe.

(4) T F I can make impromptu speeches even on topics about which I have almost no information.

(5) T F I guess I put on a show to impress or entertain people.

(6) T F I would probably make a good actor.

(7) T F In a group of people I am rarely the center of attention.

(8) T F In different situations and with different people, I often act like very different persons.

(9) T F I am not particularly good at making other people like me.

(10) T F I am not always the person I appear to be.

(11) T F I would not change my opinions (or the way I do things) in order to please someone else or win their favor.

(12) T F I have considered being an entertainer.

(13) T F I have never been good at games like charades or improvisational acting.

(14) T F I have trouble changing my behavior to suit different people and different situations.

(15) T F At a party I let others keep the jokes and stories going.

(16) T F I feel a bit awkward in company and do not show up quite as well as I should.

(17) T F I can look anyone in the eye and tell a lie with a straight face (if for the right end).

(18) T F I may deceive people by being friendly when I really dislike them.

THANK YOU. PLEASE INDICATE TO THE EXPERIMENTER THAT YOU ARE FINISHED. DO NOT TURN TO THE NEXT PAGE.

APPENDIX 2

LINEUP-DECISION CONFIDENCE ASSESSMENT

Now that you have made a lineup decision (either you chose someone or you did not choose anyone), please answer the questions below.

1. How confident are you that you have made the correct decision? (circle one)

Not at all confident			Moderately confident			Extremely confident
1	2	3	4	5	6	7

2. What do you think is the probability (in percent) that you have made a correct decision?
(provide a number between 0% and 100% that indicates the probability that you are correct)

I think there is a _____ % probability that I am correct.

3. If you chose someone from the lineup, would you be willing to testify under oath that the person you identified is in fact the person in the film? (check one)

____ YES ____ NO

If you did not choose anyone, ignore this question.

4. Just based on your own impressions, how difficult would you say the identification-task was? (circle one)

Not at all difficult			Moderately difficult			Extremely difficult
1	2	3	4	5	6	7

5. How much time do you think you spent looking at the pictures? _____ secs/mins
(write in a number and circle secs or mins)

APPENDIX 3

IMPRESSION INVENTORY

Now that you've watched the short film showing the two men playing tennis, please answer the four questions below. There are no right or wrong answers, just respond on the basis of your impressions.

1. Which of the two players, the blond-haired player (further from the camera) or the dark-haired player (closer to the camera), do you think is the better tennis player? (check one)

___ BLOND ___ DARK

2. Which of the two players, blond or dark, do you think has more experience at playing tennis? (check one)

___ BLOND ___ DARK

3. If you were asked to recommend one of the players as a tennis instructor, who would you choose? (check one)

___ BLOND ___ DARK

4. Do you think the two players are good friends, or do you think they just met that day at the tennis court? (check one)

___ GOOD FRIENDS ___ JUST MET

THANK YOU. PLEASE INDICATE TO THE EXPERIMENTER THAT YOU ARE FINISHED. DO NOT TURN TO THE NEXT PAGE.

APPENDIX 4

CROSS-EXAMINATION

1. Could you describe for me, very briefly, what happened in the film?

2. How would you describe your opportunity to see the man in the film? Would you say you got a good view of him, or what?

3. About how long did you get to see him for?

4. Do you think that was enough time to <u>really</u> get a good look at him?

5. Isn't it true, though, that the man was facing <u>away</u> from you for most of the time?

6. Can you describe what the man was wearing?

7. Do you think you'd be able to recognize his clothes if you were to see them again?

8. Is it true that you have identified the picture of a suspect as the man who you saw in the film?

9. Do you think it's possible that you've only seen the man you chose from the pictures somewhere else, and that you just <u>think</u> he's the man from the film?

10. Do you think you know the woman in the film?

11. Do you think you'd be able to recognize <u>her</u> if you saw her again?

APPENDIX 5

IMPRESSIONS OF WITNESS' LINEUP-IDENTIFICATION BEHAVIOR

Now that you've watched the witness look through the pictures and make an identification, please answer the following questions:

1. Do you think the witness correctly identified the thief? In other words, do you think the person who the witness chose from the pictures is the person who committed the crime in the film? (CHECK ONE)

_____ YES _____ NO

2. How confident do you feel about your answer to Question 1? (CIRCLE ONE)

Not at all confident			Moderately confident			Extremely confident
1	2	3	4	5	6	7

3. If you didn't have to make a "YES" or "NO" decision, what would you say is the probability (in percent) that the witness made a correct decision?

(PROVIDE A NUMBER BETWEEN 0 AND 100. IF YOU ANSWERED "YES" TO QUESTION 1, THEN YOUR NUMBER SHOULD BE GREATER THAN 50%. IF YOU ANSWERED "NO" TO QUESTION 1, THEN YOUR NUMBER SHOULD BE LESS THAN 50%)

"I think there is a _____ % probability that the witness is correct."

4. How confident would you say the witness appeared to be? (CIRCLE ONE)

Not at all confident			Moderately confident			Extremely confident
1	2	3	4	5	6	7

5. How long do you think the witness took to look at all of the pictures? _____ secs / mins (CIRCLE ONE)

THANK YOU. YOU MAY NOW LEAVE YOUR CUBICLE AND GO SEE THE EXPERIMENTER AT THE END OF THE HALL.

APPENDIX 6

IMPRESSIONS OF WITNESS' INTERVIEW BEHAVIOR

Now that you've watched the witness answer the questions regarding the film and the person chosen from the pictures, please answer the following questions:

1. Do you think the witness correctly identified the thief? In other words, do you think the person who the witness chose from the pictures is the person who committed the crime in the film?
(CHECK ONE)

_____ YES _____ NO

2. How confident do you feel about your answer to Question 1?
(CIRCLE ONE)

Not at all confident			Moderately confident			Extremely confident
1	2	3	4	5	6	7

3. If you didn't have to make a "YES" or "NO" decision, what would you say is the probability (in percent) that the witness made a correct decision?
(PROVIDE A NUMBER BETWEEN 0 AND 100. IF YOU ANSWERED "YES" TO QUESTION 1, THEN YOUR NUMBER SHOULD BE GREATER THAN 50%. IF YOU ANSWERED "NO" TO QUESTION 1, THEN YOUR NUMBER SHOULD BE LESS THAN 50%)

"I think there is a _____ % probability that the witness is correct."

4. How confident would you say the witness appeared to be?
(CIRCLE ONE)

Not at all confident			Moderately confident			Extremely confident
1	2	3	4	5	6	7

5. How believable would you say the witness was?
(CIRCLE ONE)

Not at all believable			Moderately believable			Extremely believable
1	2	3	4	5	6	7

APPENDIX 7

IMPRESSIONS OF WITNESS' LINEUP AND INTERVIEW BEHAVIOR

Now that you've watched the witness answer the questions regarding the film and the person chosen from the pictures, please answer the following questions:

1. Do you think the witness correctly identified the thief? In other words, do you think the person who the witness chose from the pictures is the person who committed the crime in the film?
(CHECK ONE)

_____ YES _____ NO

2. How confident do you feel about your answer to Question 1?
(CIRCLE ONE)

Not at all confident			Moderately confident			Extremely confident
1	2	3	4	5	6	7

3. If you didn't have to make a "YES" or "NO" decision, what would you say is the probability (in percent) that the witness made a correct decision?
(PROVIDE A NUMBER BETWEEN 0 AND 100. IF YOU ANSWERED "YES" TO QUESTION 1, THEN YOUR NUMBER SHOULD BE GREATER THAN 50%. IF YOU ANSWERED "NO" TO QUESTION 1, THEN YOUR NUMBER SHOULD BE LESS THAN 50%)

"I think there is a _____ % probability that the witness is correct."

4. How confident would you say the witness appeared to be?
(CIRCLE ONE)

Not at all confident			Moderately confident			Extremely confident
1	2	3	4	5	6	7

5. How long do you think the witness took to look at all of the pictures? ____ secs / mins
(CIRCLE ONE)

6. How believable would you say the witness was?
(CIRCLE ONE)

Not at all believable			Moderately believable			Extremely believable
1	2	3	4	5	6	7

APPENDIX 8

INSTRUCTIONS TO PARTICIPANTS

The experiment you are participating in today has to do with people's reactions to, and impressions of, eyewitness testimony and eyewitness identification. In a previous experiment, other participants viewed a simulated crime (a purse-snatching) on videotape. Afterward, participants were asked to imagine that were witnesses to the crime depicted on the videotape. Then they were asked if they could identify the thief from a number of very similar-looking pictures. All participants then looked through a series of color slides, one picture at a time. They were allowed to spend as much time as they liked looking at each picture. For each picture, witnesses were asked to say the number of the picuture out loud ("Number 1, Number 2, " etc.) and to say "Yes" or "No", in answer to the question "Is that the man you saw in the videotape?". All participants were videotaped as they looked through the pictures.

In this experiment, we are interested in people's impressions of the witnesses' behavior as they looked through the pictures. In just a few minutes you will be seeing a videotape of one the witnesses from the experiment described above. PLEASE WATCH THE VIDEOTAPE CLOSELY. Afterward, the experimenter will bring a response-sheet to your cubicle, and you will be asked to rate the witness on a number of dimensions (accuracy, confidence, etc.). PLEASE READ THE INSTRUCTIONS AT THE TOP OF THE RESPONSE-SHEET, AND FOR EACH QUESTION, VERY CAREFULLY.

When you are finished with the response-sheet, please take your belongings, leave your cubicle quietly, and see the experimenter at the end of the hallway. There, you will receive credit for your participation and a written debriefing describing the experiment in more detail.

Thank you for participating.

APPENDIX 9

INSTRUCTIONS TO PARTICIPANTS

The experiment you are participating in today has to do with people's reactions to, and impressions of, eyewitness testimony and eyewitness identification. In a previous experiment, other participants viewed a simulated crime (a purse-snatching) on videotape. Afterward, participants were asked to imagine that were witnesses to the crime depicted on the videotape. Then they were asked if they could identify the thief from a number of very similar-looking pictures. All participants then looked through a series of color slides, one picture at a time. They were allowed to spend as much time as they liked looking at each picture. Following their identification, all witnesses were asked a number of questions regarding the simulated crime and about the man who they had identified from the pictures. All witnesses were videotaped as they answered these questions.

In this experiment, we are interested in people's impressions of the witnesses' behavior as as they answered the questions. In just a few minutes you will be seeing a videotape of one the witnesses from the experiment described above. PLEASE WATCH THE VIDEOTAPE CLOSELY. Afterward, the experimenter will bring a response-sheet to your cubicle, and you will be asked to rate the witness on a number of dimensions (accuracy, confidence, etc.). PLEASE READ THE INSTRUCTIONS AT THE TOP OF THE RESPONSE-SHEET, AND FOR EACH QUESTION, VERY CAREFULLY.

When you are finished with the response-sheet, please take your belongings, leave your cubicle quietly, and see the experimenter at the end of the hallway. There, you will receive credit for your participation and a written debriefing describing the experiment in more detail.

Thank you for participating.

APPENDIX 10

INSTRUCTIONS TO PARTICIPANTS

The experiment you are participating in today has to do with people's reactions to, and impressions of, eyewitness testimony and eyewitness identification. In a previous experiment, other participants viewed a simulated crime (a purse-snatching) on videotape. Afterward, participants were asked to imagine that were witnesses to the crime depicted on the videotape. Then they were asked if they could identify the thief from a number of very similar-looking pictures. All participants then looked through a series of color slides, one picture at a time. They were allowed to spend as much time as they liked looking at each picture. For each picture, witnesses were asked to say the number of the picuture out loud ("Number 1, Number 2, " etc.) and to say "Yes" or "No", in answer to the question "Is that the man you saw in the videotape?" All participants were videotaped as they looked through the pictures.

Following their identification, all witnesses were asked a number of questions regarding the simulated crime and about the man who they had identified from the pictures. All witnesses were again videotaped as they answered these questions.

In this experiment, we are interested in people's impressions of the witnesses' behavior as they looked through the pictures and as they answered the questions. In just a few minutes you will be seeing a videotape of one the witnesses fr he experiment described above. First you will see the witness looking through the pic ur s. Following a brief pause, you will see the witness answering the questions. PLEASE WATCH THE VIDEOTAPE CLOSELY. Afterward, the experimenter will bring a response-sheet to your cubicle, and you will be asked to rate the witness on a number of dimensions (accuracy, confidence, etc.). PLEASE READ THE INSTRUCTIONS AT THE TOP OF THE RESPONSE-SHEET, AND FOR EACH QUESTION, VERY CAREFULLY.

When you are finished with the response-sheet, please take your belongings, leave your cubicle quietly, and see the experimenter at the end of the hallway. There, you will receive credit for your participation and a written debriefing describing the experiment in more detail.

Thank you for participating.