Accuracy, Confidence, and Juror Perceptions in Eyewitness Identification

Gary L. Wells, R. C. L. Lindsay, and Tamara J. Ferguson
University of Alberta, Edmonton, Canada

Subjects of both sexes individually witnessed the staged theft of a calculator. The 127 witnesses were then given the opportunity to identify the thief from a six-person picture array; from this sample, 24 accurate-identification witnesses and 18 inaccurate-identification witnesses were cross-examined with either leading or nonleading questions. Jurors were unable to distinguish accurate from inaccurate witnesses across the 42 cross-examination sessions ($d' = .02$). However, jurors in the leading-questions conditions were significantly more likely to believe accurate than inaccurate witnesses ($d' = .35$), whereas the reverse effect held for nonleading questions ($d' = -.39$). Jurors’ attributions of witness confidence were unrelated to witness accuracy, even though these attributions accounted for 50% of the variance in jurors’ decisions to believe witnesses. The poor accuracy-confidence relationship among witnesses is discussed in relation to the research on probability calibration.

Eyewitness identification of criminal suspects is heavily relied on in the criminal justice system. (See Woocher, 1977.) Yet, staged-crime research indicates that eyewitness identifications are often unreliable. Indeed, a recent study indicated that following a staged crime, 34% of the witnesses made no identification from a six-person photo spread, 31% made accurate identifications, and 35% made false identifications (Leippe, Wells, & Ostrom, 1978). Although it is not clear to what extent current research may be presenting an exaggerated picture of the unreliable eyewitness (see Wells, 1978), there is little doubt that the rate of false identifications (i.e., identifying an innocent lineup member) in staged crimes is acceptably high. The current investigation was designed to more fully analyze the implications of false identifications from the perspective of the criminal-justice system.

Although fairly high rates of false identification have been found in the research, the courts may discount the impact that inaccurate witnesses have on the system for at least two reasons. First, according to a differential-attrition hypothesis, inaccurate witnesses may be less likely to appear in court than accurate witnesses. Second, according to a differential-belief hypothesis, inaccurate witnesses may fail more than accurate witnesses to convince the jury of the reliability of their testimony. The primary concern of the current investigation was with the differential-belief hypothesis, although our design also allowed a preliminary assessment of the differential-attrition hypothesis.

The only published research relating to the differential-belief and differential-attrition hypotheses has simply analyzed the relationship between eyewitnesses’ identification accuracy and the eyewitnesses’ confidence in their identifications. Two studies have examined the relationship between identification accuracy and eyewitnesses’ confidence, and both studies reported no significant correlation (Brown, Deffenbacher, & Sturgill, 1977;
Leippe et al., 1978). These results suggest that the differential-attrition and differential-belief hypotheses may be invalid because eyewitnesses' confidence would seem to be the principal determinant of witnesses' willingness to appear in court and jurors' belief of the witness. However, there are reasons to question any general claim from these two studies that eyewitness identification accuracy is unrelated to eyewitness confidence and, hence, to attrition or believability. First, the nonsignificant correlations reported by Leippe et al. and Brown et al. were "incidental" reports of data that had no relationship to the hypotheses of those investigators. Thus, at the very least, there is a clear need to replicate this null effect with independent subjects and operations.

We felt, however, that there are more important issues than mere replication. Recent evidence, although controversial, suggests that self-reports sometimes show little correspondence to overt behavior. (See Evans & Wason, 1976; Nisbett & Wilson, 1977; Smith & Miller, 1978.) Thus, although self-ratings of confidence on a scale may not be indicative of accuracy, witnesses in the criminal-justice system express their confidence through their overt responses to adversary cross-examination. These overt indicants of confidence may well be related to accuracy. This is one reason why our analysis focuses heavily on the jurors' responses to witnesses' testimony. Also, recent evidence indicates that people may alter their interpretations of an event that they just witnessed if they are led to believe that they may be tested further (Wortman, Costanzo, & Witt, 1973) or that they may have to defend their interpretations (Wells, Petty, Harkins, Kagehiro, & Harvey, 1977). Unlike the procedures used in the criminal-justice system, however, previous eyewitness studies have not led the witnesses to believe that their private, scaled ratings of confidence might be challenged. Given these two considerations, we felt that the proper forensic test of the confidence-accuracy relationship should involve confidence as measured through responses to cross-examination. Furthermore, we expected that the confidence-accuracy relationship would be most likely to surface when the cross-examination involved high pressure and challenging questions. Thus, we used two types of cross-examination: (a) leading questions, which tend to pressure witnesses by at times asserting false premises and which attempt to elicit contradictions in testimony, and (b) nonleading questions, which have an open-ended structure and assert no false premises and, therefore, do not exert high pressure.

Finally, although the confidence-accuracy issue is one of the primary foci of this research, we should note that our central concern is with the differential-belief hypothesis. Whether or not differential belief is mediated by the confidence variable is a strong, but secondary, concern. A 2 X 2 factorial design was used to test the differential-belief hypothesis and related concerns. Jurors observed a witness under cross-examination who had made either a correct or false identification of a suspect following a staged crime. Each witness was cross-examined using either a leading or nonleading cross-examination style. Our principal dependent measure was whether or not the jurors believed that the witness had made a correct identification of the criminal. Witnesses' self-rated confidence in their identifications, jurors' ratings of witnesses' confidence, and jurors' confidence in their believe/not believe decisions constituted the other main measures.

Method

Overview

Three individuals conducted the study; one served as the experimenter, one served as the confederate thief, and the second author served as the cross-examiner. The procedure involved two phases. In the first phase, subject witnesses individually witnessed a staged theft and were subsequently asked to identify the thief from a set of pictures. The "crime" was designed so that the accuracy rate for identifications was well above chance level. From these witnesses, the experi-
menter randomly selected a witness who had made either an accurate or a false identification. In the second phase, the chosen witness was cross-examined using either leading or nonleading questions. Subject jurors observed the cross-examination and subsequently indicated whether or not they believed that the witness had made a correct identification. All subjects were completely debriefed following the cross-examination by being informed of the method and rationale for the study. In addition, subjects were referred to an eyewitness article placed on library reserve.

Subjects

Three hundred twenty-eight introductory psychology students served as subjects for credit in their course. One hundred twenty-seven subjects were randomly selected to serve as witnesses and the remaining 201 subjects served as jurors. Twenty-four accurate-identification witnesses and 18 false-identification witnesses were cross-examined in 42 sessions that had from three to seven jurors each.

Procedure

On arrival, the experimenter explained that subjects would be assigned to cubicles. However, after the first three subjects (witnesses) for each session were assigned to cubicles, the remaining subjects (jurors) were directed to a larger room (the courtroom), where they worked on a filler task until Phase 2. Our intention was to have nine subjects arrive for each session, assign three as witnesses, and use the remaining six as jurors. This allowed us to randomly select either an accurate or inaccurate witness from among the three. However, show-up rates varied so that we could not always have six jurors and three witnesses. The minimum number of jurors in any session was three.

Phase 1: The crime. Upon entering their individual cubicles, the soon-to-be witnesses were asked to fill out their experiment participation cards and were told that instructions regarding their experimental task would follow. After the subjects had been alone in their cubicles for a few minutes, the thief entered, posing as a co-participant. Upon entering, sitting down, and placing her purse on the floor, the thief “discovered” a calculator that had apparently been left by a previous subject. After examining the calculator, the thief stated that she would like to have such a calculator, placed the calculator in her purse, and quickly exited. The thief was in each cubicle for approximately 75 sec; she looked at and spoke to each witness four times during this interval.

Approximately 30 sec later, the experimenter entered the cubicle. The experimenter did not mention the theft for several seconds, thereby giving the witness time to volunteer the information that the theft had occurred. If the witness failed to volunteer the information, the experimenter asked if the witness had seen a calculator in the room. At this point, all witnesses reported the theft. The experimenter then informed all subjects that the theft had been staged for their benefit, gave each witness a questionnaire requesting a description of the thief, and asked the witness to identify the thief from a set of six pictures. The picture array from which witnesses identified the thief was tested for fairness according to the criteria specified in Wells et al. (Note 1). Using Wells et al.’s empirical criteria, the six-person picture array was functionally composed of 5,38 persons. The experimenter was not in the room during the identification test. The witnesses were told that the thief may or may not be in the set of pictures. Subjects responded by checking one of the six pictures or a seventh category labeled “none of them.” Finally, witnesses who identified someone from the set of pictures were asked to indicate their confidence that they had made the correct choice by responding on a 9-point scale with endpoints labeled “not very certain” and “absolutely certain” (9).

Phase 2: The cross-examination. While the witnesses were completing their identification task, the experimenter informed the jurors of their true role in the experiment. Jurors were told that the persons who were sent to the cubicles had witnessed a staged theft of a calculator, that the witnesses had been asked to identify the thief from a set of six pictures, and that they (the jurors) would observe the cross-examination of one of the witnesses who had made an identification. Jurors were told that their task would be to decide whether or not the particular witness, chosen at random, was or was not mistaken in his or her identification. The cross-examiner was blind as to whether or not the witness was accurate in his or her identification, and subjects were informed of that fact.

The experimenter then consulted a predetermined, randomly ordered chart that called for either an accurate or an inaccurate witness. The chart was constructed to allow for an equal number of accurate and inaccurate witnesses to be cross-examined. However, due to the high percentage of accurate witnesses, deviations were allowed whenever there was not an inaccurate witness to be cross-examined in that session. After the experimenter selected a witness, the person was escorted to the courtroom. Witnesses who made no identification were not cross-examined.

At this point, the cross-examiner entered the courtroom and introduced himself as the attorney for the defense of Mary White, a fictitious name assigned to the identified suspect. The cross-examiner proceeded with the cross-examination using either leading or nonleading questions. Both types of cross-examination included 25 questions covering the same material. Leading questions were characterized by the occasional assertion of a false premise, questions that yielded only short answers, and questions that tried to elicit contradictions. For example, the question, “The person you saw had
Table 1  
Mean Confidence for all Accurate and Inaccurate Witnesses and Confidence-Accuracy Correlation Before Cross-Examination

<table>
<thead>
<tr>
<th>Item</th>
<th>Accurate-identification witnesses</th>
<th>Inaccurate-identification witnesses</th>
<th>No-identification witnesses</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number</td>
<td>74</td>
<td>26</td>
<td>27</td>
</tr>
<tr>
<td>%</td>
<td>58</td>
<td>20</td>
<td>21</td>
</tr>
<tr>
<td>M confidence in identification</td>
<td>6.96</td>
<td>5.62</td>
<td>–</td>
</tr>
</tbody>
</table>

a jacket on, didn't she?" was later followed by "Was the jacket she had on tan or brown?" when, in fact, she was not wearing a jacket. Nonleading questions, however, included no false premises, were largely open-ended in nature, and did not actively seek contradictions; for example, "Describe what the person was wearing." For the final two questions, the witness was shown the original mugshots and was asked, "For the benefit of the jury, would you please re-identify the photograph of the person you think you saw? How certain are you that the photograph you just identified is a photo of the person who presumably committed the theft?" The leading and nonleading cross-examinations did not differ in the wording of these two questions. Neither the cross-examiner nor the jury could see the photos. At this point, the defense counsel ended the cross-examination and the witness and jurors were given questionnaires.

Post-Cross-Examination Measures

Jurors' questions. Following the cross-examination, jurors were asked in written form, "Do you believe that the witness correctly identified the person who stole the calculator?" (yes or no). In addition, jurors were asked to indicate (a) how certain they were of that decision on a 9-point scale on which 1 indicated "not at all certain" and 9 indicated "absolutely certain"; (b) to what extent they believed they could generally tell from a witness' testimony whether or not the witness made an accurate identification; (c) what percentage of witnesses they would expect to make an accurate identification, given conditions similar to those in which the current witness participated; and (d) how confident they thought the witness was in his or her ability to identify the thief. For exploratory purposes, several other questions were asked of the jurors. However, responses to these questions are largely irrelevant to the current issues.

Witnesses' questions. Following the cross-examination, witnesses were asked seven questions, four of which were rated on 9-point scales. Witnesses were asked to rate how certain they were that they had chosen the correct person (1 = not at all certain), how much the cross-examination affected their level of confidence (1 = greatly reduced confidence), how they felt the cross-examiner had treated them during questioning (1 = friendly), and the extent to which jurors could tell in general how accurate a witness was (1 = definitely cannot tell). Two questions asked what percentage of witnesses could make an accurate identification under similar conditions and what percentage of jurors believed that the witness had made a correct identification; both of these were answered on 11-point scales ranging, in 10% intervals, from 100% (at the beginning) to 0%. Witnesses also responded "yes" or "no" to a question asking them to recall if there were any cross-examination questions that upset them or caught them off guard.

Results

Witnesses' Data

The identification and confidence data obtained from all of the witnesses in Phase 1 are presented in Table 1. Unlike Leippe et al. (1978) and Brown et al. (1977), the overall correlation between accuracy and confidence was small, but significant ($r = .29$, $p < .05$). However, for the 42 witnesses who were cross-examined, neither pre- nor post-cross-examination correlations between accuracy and confidence were significant (both $r_s = .01$).

Regarding the differential-attrition hypothesis, three witnesses refused to be cross-examined. Two witnesses were accurate and one was inaccurate. Given the low attrition rate, no statistical analyses were performed to test the differential-attrition hypothesis.

Witnesses' responses to the seven post-cross-examination items were analyzed using separate $2 \times 2$ between-subjects analyses of vari-

---

3 Cross-examination scripts are available from the first author.
Table 2
Proportion of Jurors Making Correct Decisions and Proportion of Jurors Believing the Witnesses as Functions of Witness Accuracy and Type of Cross-Examination

<table>
<thead>
<tr>
<th>Witnesses' identification accuracy</th>
<th>Proportion of jurors believing the witnesses</th>
<th>Proportions of jurors making correct decision</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Leading questions</td>
<td>Nonleading questions</td>
</tr>
<tr>
<td>Inaccurate</td>
<td>.73</td>
<td>.86</td>
</tr>
<tr>
<td>Accurate</td>
<td>.84</td>
<td>.76</td>
</tr>
</tbody>
</table>

ance (Type of Cross-Examination × Witness Accuracy). Only one of the seven items resulted in a significant effect. Specifically, there was a main effect for accuracy on the witnesses' estimates of what percentage of witnesses they felt could make an accurate identification under the same conditions that they (the witnesses) were in. Accurate witnesses felt that 57.19% of the witnesses could be accurate, whereas inaccurate witnesses felt that 42.56% of the witnesses could be accurate, $F(1, 159) = 6.63$, $p = .014$.

**Jurors' Data**

The five juror questions were also analyzed by separate $2 \times 2$ between-subjects analyses of variance. Because our design necessitated that sessions be nested within witnesses, all analyses of variance on jurors' data were calculated with the sessions effects removed from the error term. Thus, instead of having 197 degrees of freedom in each error term, there are 159 degrees of freedom. A least-squares weighted-means analysis was used on jurors' data.

The principle measure, juror belief, was coded in two ways. First, the jurors' belief scores were coded as 1 (believe) and 0 (not believe). An analysis of variance using this procedure yielded no significant main effects (both $F$s < 1), but a significant interaction, $F(1, 159) = 4.01$, $p < .05$. The proportions of jurors who believed the witnesses are presented in Table 2 by condition. As evidenced in Table 2, the nature of this interaction is that jurors were more willing to believe accurate than inaccurate witnesses in the leading-questions condition, whereas the opposite trend occurred for the nonleading-questions condition. We also coded the juror-belief data with regard to whether the juror was or was not correct in his or her belief decision (scored as 1 and 0, respectively). The analysis of variance on this data resulted in a significant main effect for witness accuracy, $F(1, 159) = 122.4$, $p < .001$, a significant main effect for type of cross-examination, $F(1, 159) = 3.91$, $p < .05$, and no interaction, $F < 1$. Proportions of correct jurors are also presented in Table 2. Inspection of Table 2 shows that the cross-examination effect is due to the fact that jurors were more likely to make a correct decision in the leading-questions conditions than in the nonleading-questions conditions. The strong effect for the witness-accuracy variable may simply reflect jurors' relatively high rate of belief (79.8% overall), thus making jurors far more likely to make the correct decision when the witness was accurate than when the witness was inaccurate.

To determine whether or not jurors were able to discriminate significantly between accurate and inaccurate witnesses, we used the signal detection parameter $d'$ (Elliott, 1964). That is, we treated the cases in which the witness was accurate as a case of signal-plus-noise and the cases in which the witness was inaccurate as noise alone. Hits were cases in which the witness was accurate and the juror believed the witness; false alarms were cases in which the witness was inaccurate and the juror believed the witness. The strong effect for the witness-accuracy variable may simply reflect jurors' relatively high rate of belief (79.8% overall), thus making jurors far more likely to make the correct decision when the witness was accurate than when the witness was inaccurate.

To determine whether or not jurors were able to discriminate significantly between accurate and inaccurate witnesses, we used the signal detection parameter $d'$ (Elliott, 1964). That is, we treated the cases in which the witness was accurate as a case of signal-plus-noise and the cases in which the witness was inaccurate as noise alone. Hits were cases in which the witness was accurate and the juror believed the witness; false alarms were cases in which the witness was inaccurate and the juror believed the witness.

---

3 This analysis is a statistically liberal test. Our choice of a sensitive analysis stems from our preference for, if anything, erring in favor of jurors' abilities rather than using an insensitive statistical analysis that may prematurely brand jurors as inept in making judicial decisions.
As might be expected from examining Table 2, our overall $d'$ was clearly nonsignificant ($d' = .02, z = .54, p > .05$). However, because of the significant main effect for cross-examination style on the correctness measure, we calculated separate $d$'s within cross-examination types. Within the leading-questions condition, the proportions of hits (.84) and false alarms (.73) produced evidence that jurors could reliably detect accurate from inaccurate witnesses ($d' = .35, z = 3.98, p < .01$). Surprisingly, when $d'$ is calculated for the nonleading questions condition, the resultant value is negative ($d' = -.39, z = 4.53, p < .01$) owing to the fact that false alarms exceeded hits (.86 vs. .76, respectively).

Jurors' confidence that they made a correct decision was unaffected by cross-examination type or witness accuracy. However, jurors in the leading-questions conditions felt significantly less confident that they could generally tell from a witness' testimony whether the witness was accurate than did jurors in the nonleading-questions conditions, $M = 3.83$ and 5.32, respectively; $F(1, 159) = 9.83, p < .01$. Also, jurors in the leading-questions conditions expected a significantly lower percentage of witnesses on the average to be accurate than was expected by jurors in the nonleading-questions condition, $M = 54.6\%$ and 62.9\%, respectively; $F(1, 159) = 8.62, p < .01$. No other main effects or interactions were significant on the jurors' questions.

**Principal Intercorrelations**

Five variables seemed particularly fruitful for further analysis regarding their relationships to one another: jurors' decisions to believe or disbelieve the witness; jurors' confidence in their decision to believe or disbelieve the witness; jurors' correctness in their decision to believe or disbelieve the witness; jurors' attributions of confidence to the witness; and witnesses' self-rated confidence. An intercorrelation matrix was constructed for these five variables and is presented in Table 3. The results of this analysis reveal that jurors' correctness in their decisions to believe or disbelieve the witness was unrelated to any of these variables. Although the correctness variable was unrelated to any other variable, the remaining four variables were all significantly related to one another. Perhaps most impressive is the fact that jurors' ratings of the witnesses' confidence accounted for nearly 50\% of the variance in jurors' decisions of whether or not to believe the witness.

**Discussion**

We began our investigation with the question of whether or not a witness who makes a false identification of a defendant will have as much courtroom impact (as defined by the proportion of jurors who believe the witness) as a witness who makes an accurate identification. Our overall $d'$ analysis indicated that jurors' belief decisions were unrelated to whether or not the witness that they observed had made an accurate identification. This independence between witnesses' accuracy and jurors' belief inferences is less distressing, however, when the $d'$ analyses are calculated within cross-examination types. Specifically, there was a significant positive $d'$ in the leading cross-examination condition, indicating that jurors could discriminate be-

<table>
<thead>
<tr>
<th>Table 3</th>
<th>Principal Intercorrelations</th>
</tr>
</thead>
<tbody>
<tr>
<td>Variable</td>
<td>1</td>
</tr>
<tr>
<td>1. Jurors' believe/not believe decisions</td>
<td>—</td>
</tr>
<tr>
<td>2. Jurors' confidence in their decisions</td>
<td>—</td>
</tr>
<tr>
<td>3. Jurors' correctness in their decisions</td>
<td>—</td>
</tr>
<tr>
<td>4. Jurors' attributions of confidence to the witness</td>
<td>—</td>
</tr>
<tr>
<td>5. Witnesses' self-rated confidence</td>
<td>—</td>
</tr>
</tbody>
</table>

*Note. All correlations used the mean of jurors' responses for each session and thus are based on $n = 42$. 
* $p < .05$. 
** $p < .01$. 

between accurate and inaccurate witnesses. Unexpectedly, however, the nonleading cross-examination condition yielded a significant negative $d'$ value. The negative $d'$ value also indicates that jurors could discriminate between accurate and inaccurate witnesses, but jurors' decisions to believe appear to be governed by a significantly incorrect decision rule. In psychophysical terms, a negative $d'$ seems to indicate that the subject responder has reversed his or her interpretation of noise only versus signal plus noise.

Even though jurors' estimates of the witnesses' confidence were strongly related to jurors' decisions to believe or disbelieve the witness, it appears that the confidence of the witness was not the signal that allowed jurors to make accurate discriminations in the leading cross-examination conditions. Nor did the confidence of the witness serve as a faulty signal that could account for jurors' inaccurate discriminations in the nonleading cross-examination conditions. The evidence for this is contained in the fact that witnesses' confidence, whether measured by the witnesses themselves or through jurors' estimates, was unrelated to witnesses' accuracy for the 42 cross-examined witnesses. This was true of both the leading and nonleading cross-examination conditions. In fact, the only witness measure that was significantly related to witnesses' accuracy was witnesses' responses to the question, "What percentage of witnesses would be accurate under these conditions?" The fact that this question produced a significant effect for accuracy and no interaction with cross-examination style suggests that it may have fruitful validity in helping to distinguish between accurate and inaccurate witnesses.

We also found no evidence to support the differential-attrition hypothesis. In fact, the data show a nonsignificant reversal effect wherein two of our accurate-identification witnesses refused to be cross-examined, whereas only one of our false-identification witnesses refused.

Interestingly, subjects' verbal reports of to what extent they believe that they can tell whether a witness is accurate by listening to her or his testimony produced results that were opposite to what might be expected based on the $d'$ analyses. Specifically, jurors in the leading-question conditions felt less competent at detecting whether a witness was accurate than did subjects in the nonleading-question conditions. This apparent discrepancy between subjects' verbal reports and their behavioral performance may parallel Evans and Wason's (1976) contention that people have little or no access to higher order cognitive processes.

The data indicate that although jurors' decisions to believe the witness are highly related to their ratings of the witnesses' confidence, the confidence-accuracy relationship is very poor. Recent research in the area of probability calibration suggests that low accuracy-confidence relationships may be related to a lack of relevant "feedback experience." Specifically, recent research by Lichtenstein and Fischhoff (1977, Experiment 2) indicated that there was practically no relationship between accuracy and confidence on a task that was novel (e.g., identifying whether a handwriting specimen was American or European). However, after allowing subjects to study a different sample of handwriting specimens, correctly labeled as American or European, subjects' confidence was significantly related to accuracy of decision.

Do people get the kind of feedback in daily encounters that should allow them to be well calibrated regarding confidence-accuracy relationships in eyewitness identification? Perhaps not. In fact, we argue that daily encounters and the assumptions on which they operate might serve to maintain the poor accuracy-confidence relationship and enhance overconfidence. First we note that with rare exceptions, the recognition of one's own friends is not the same task as that required in criminal identifications. Criminal identifications typically involve a single exposure followed in time by a recognition test. We often see someone in a hallway, on the street, or at a party who we believe we recognize, but, not knowing the person well, we simply smile or make some ambiguous gesture of recognition. Such gestures are almost always reciprocated and, thus, we consider our belief (i.e., that we made a correct recognition) to be
confirmed. However, in signal detection terms, we have failed to distinguish between a "hit" and a "false alarm," with the verdict almost always being that it was a hit. Our perceptual set has led us to think of such events as confirmations. Of course, in some cases we do get feedback. For example, at a social gathering a third party might take initiative of the form, "Have you and John met?" However, the nature of social intercourse usually takes the form of a pause following such a question so that if one person says "yes," the other person will be able to avoid an embarrassing "no" response. Indeed, there seems to be a general belief that if one person says "yes," both individuals believe that they have previously met; the remainder of the conversation is to determine where or when they met. While this may provide the person who says "no" with a disconfirmation, the overall nature of daily experience is that if one believes that he or she recognizes another person, it is practically disconfirmable. Thus, people's daily experiences may lead them to believe that if they make a recognition, they are correct. Because they lack veridical feedback experience, people fail to be sensitive to factors leading to correct and incorrect recognitions which, in turn, leads to eyewitnesses' poor accuracy–confidence calibration.

Perhaps, as in Lichtenstein and Fischhoff (1977, Experiment 2), eyewitnesses can be trained for accuracy–confidence calibration. However, the current evidence indicates that if one believes that he or she recognizes another person, it is practically disconfirmable. Thus, people's daily experiences may lead them to believe that if they make a recognition, they are correct. Because they lack veridical feedback experience, people fail to be sensitive to factors leading to correct and incorrect recognitions which, in turn, leads to eyewitnesses' poor accuracy–confidence calibration.

Generalizing to the Courtroom

Although the implications of the current research for criminal justice are rather obvious, we must for the moment insert a limit on the extent to which we are willing to generalize. Specifically, it must be remembered that the confidence–accuracy data and the jurors' detection data were based on the use of a "constant crime" design. By constant crime, we mean that each witness saw approximately the same event for the same amount of time, and so on. Suppose, however, that the crime was different for each witness. It may be that accurate and inaccurate witnesses do not differ in their confidence within a given crime, but, given a sample of witnesses who observed different crimes, the relationship may exist. It seems only reasonable, for example, that witnesses who observe a criminal for 15 sec will be both less accurate and less confident than witnesses who observe a criminal for 15 min. Thus, although the confidence–accuracy relationship may be weak or nonexistent as measured between witnesses within an event, it may exist as measured between events.

A similar generalization problem exists over whether jurors are "overbelieving" of witnesses or not. The current data indicated that among witnesses who made a choice, 74% were accurate; a value that is reasonably close to the 79.8% figure that represents the jurors' rate of belief. However, it is not at all clear what would happen to the jurors' rate of belief if the crime would have been of the type that produced a low rate of accuracy. It is possible that the jurors' rate of belief is around 80% irrespective of the actual rate of witness accuracy.

Reference Note


References


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


Received September 5, 1978