Can People Detect Eyewitness-Identification Accuracy Within and Across Situations?

R. C. L. Lindsay, Gary L. Wells, and Carolyn M. Rumpel
University of Alberta, Edmonton, Alberta, Canada

Thefts were staged 108 times for as many witnesses who were subsequently given a photo lineup for identifying the thief. The thefts were staged under conditions designed to yield low (33%), moderate (50%), or high (74%) proportions of correct identifications of the thief. Corroborating past research, the relationship between witnesses' identification accuracy and witnesses' confidence was negligible within conditions. There was no evidence that the confidence-accuracy relationship changed across conditions or that witness confidence changed across theft conditions. A representative sample of 48 witnesses (8 accurate-identification and 8 false-identification witnesses from each of the three theft conditions) was cross-examined. Subjects (n = 96) viewing the cross-examinations showed no ability to detect accurate- from false-identification witnesses within conditions as measured by subjects' belief of witnesses. Although subjects changed their rate of belief of witnesses as a function of the theft conditions (62%, 66%, and 77%, respectively), the rate at which subjects discounted witnesses' testimony was insufficient across conditions. Subjects were shown to be especially over-believing of witnesses when the rate of witness accuracy in that condition was low.

Staged-crime research has shown that false identifications of innocent lineup members is a fairly frequent occurrence (see Wells, 1978). As Wells, Lindsay, and Ferguson (1979) pointed out, however, it is not the rate of false identifications per se that creates problems for the criminal justice system. Instead, it is the rate at which jurors believe false-identification witnesses versus accurate-identification witnesses. The importance of false identifications diminishes as jurors' abilities to detect false identifications increase. Wells et al. (1979) developed a paradigm for estimating the forensic performance abilities of jurors by cross-examining the witnesses to staged crimes and having subject jurors make decisions regarding whether the witnesses had accurately identified the criminal. In the Wells et al. study, accurate-identification witnesses (n = 18) were believed by 80.0% of the subject jurors, whereas inaccurate-identification witnesses (i.e., witnesses who mistakenly identified an innocent suspect, n = 22) were believed by 79.5% of the subject jurors. The virtual identity of these belief rates suggests that jurors may not be able to compensate for eyewitness errors that occur early in the criminal justice process.

The Wells et al. (1979) results are not overly surprising considering that subjects relied heavily on witness confidence when making belief decisions, but witness confidence was relatively nondiagnostic of witness accuracy. There are, however, issues left unresolved in the Wells et al. study. First, Wells et al. (1979) used a "constant crime" design. That is, all witnesses were exposed to the same criminal event for the same amount of time, and so on. However, witnesses who observe a criminal event under poor conditions (e.g., little processing time, poor view of the thief) may not only be less likely to make an accurate identification but also may be appreciably less confident in their identification. That is, though witnesses...
may have little internal knowledge of accuracy, as indicated by the negligible relation between witness accuracy and confidence for a given event, witnesses may be able to infer their likely inaccuracy from the situation (i.e., confidence may change across events). We will refer to this as the situational accuracy–witness confidence relationship to distinguish it from the individual accuracy–witness confidence relationship, the latter being the relationship between accurate and inaccurate witnesses’ confidence within a given crime as in Wells et al. Although the individual accuracy–witness confidence relationship is extremely weak (Brown, Deffenbacher, & Sturgill, 1977; Leippe, Wells, & Ostrom, 1978; Wells et al., 1979), the situational accuracy–witness confidence relationship may be more robust. Subjects who are apparently quite sensitive to witness confidence as a cue to witness accuracy (Wells et al., 1979) may thereby show appropriate changes in their belief of witnesses across situational-accuracy conditions.

The second issue unresolved in the Wells et al. (1979) study concerns the question of whether jurors tend to be overbelieving of witnesses. Among witnesses who made an identification in the Wells et al. study, 74% were accurate and 26% were inaccurate (i.e., made false identifications). Subjects believed witnesses 79.8% of the time and did not believe witnesses 20.2% of the time. Thus, the rate at which subjects believed witnesses and the rate of witness accuracy were fairly close. However, the constant-crime design of Wells et al. again presents an incomplete picture, since this correspondence between subjects’ belief rates and witness accuracy rates may have been accidental. The critical issue is whether the rate of subject belief will decline in situations in which the rate of witness accuracy declines.

Note the relationship between the two issues presented here. The first issue concerns whether subjects will show some reduction in belief of witnesses as witness accuracy rate declines, perhaps via a decline in witness confidence. To the extent that subjects do not change their belief of witnesses across the witness accuracy conditions, the subjects will be making insufficient adjustments to stimulus conditions. However, a change in subjects’ belief rates across witness accuracy conditions does not preclude the possibility that people are overbelieving of witnesses. If, for example, the rates of witness accuracy turn out to be 30%, 50%, and 70% across conditions and the rates at which subjects believe witnesses turn out to be 60%, 70%, and 80%, subjects could be said to be both making adjustments for the conditions and be overbelieving of witnesses.

The overbelief issue becomes more complicated to the extent that one wishes to apply the analysis to jurors’ decisional dilemmas. Unlike our subjects, jurors typically have several pieces of evidence that require the combining of a series of probabilities, of which eyewitness identification is only one piece. Thus, we must avoid suggesting that the belief of an inaccurate witness necessarily results in an error in a given case. Nevertheless, eyewitness identifications are believed to be highly impactful pieces of evidence (see Loftus, 1974, for empirical evidence; see Woocher, 1977, for a legal perspective) and a close alignment between the probability of jurors believing a witness and the probability of witness accuracy is desirable in the long run. Just as with combining of any other series of probabilities, the adequacy of one’s final judgment depends on the adequacy of the individual probabilities.

The first experiment involved the generation of witnesses under conditions designed to produce low, moderate, and high proportions of accurate identifications. This experiment allowed a test of the situational accuracy–witness confidence relationship hypothesis and laid the groundwork for our expectation of how subject jurors would perform. In the second experiment, a representative sample of cross-examined witnesses were presented to subjects for belief decisions that allowed a test of whether subjects make appropriate changes in belief rates across the three conditions.

Experiment 1

Method

Subjects were 108 introductory psychology students who participated for partial fulfillment of a course re-
DETECTION OF EYEWITNESS-IDENTIFICATION ACCURACY

requirement. On arrival, each witness was assigned a cubicle where he or she was asked to fill out a form giving various information (e.g., name, identification number) while the experimenter left presumably to prepare other materials. The witness was unaware of the nature of the study, and it was implied that another subject might be arriving. Shortly afterward, another subject (actually a confederate) opened the cubicle door, entered, found a calculator in the cubicle (that had previously been placed there for the benefit of the confederate), made remarks to the witness, placed the calculator in his (the confederate's) briefcase, and quickly exited.

The situational-accuracy manipulation, which was designed to produce low, moderate, and high proportions of witnesses making accurate identifications of the confederate thief, involved the following variations: In the low-situational-accuracy condition, the thief took a total of 12 sec in the cubicle, remarked to the witness "Hey, is this your calculator?" and wore a knit cap pulled down to cover part of his ears and all of his hair. The moderate-situational-accuracy condition involved the same behavior as in the low-situational-accuracy condition except that the cap was worn higher on the thief's head so that the physical features (especially hair color, texture, etc.) were visible. The high-situational-accuracy condition involved a total of 20 sec in which the thief preceded the question asked in the other two conditions with "Has the experimenter told us what to do yet?" and wore no hat.

Note that the difference between any two situational-accuracy conditions is not necessarily a function of any single factor. As with real crimes, the difference between any two criminal events is multifactorial. Our purpose was not to isolate the factors of a crime that lead to different rates of eyewitness accuracy (see Wells's 1978, discussion of estimator variables) but to generate levels of situational accuracy in a plausibly ecological manner to analyze any effects later in the criminal justice system (e.g., differential witness confidence and/or differential juror belief of witnesses).

Approximately 30 sec after the theft occurred, the experimenter reentered the witness' cubicle and inquired about the calculator. No witness failed to report that the calculator had been taken by someone else. The experimenter explained that the subject had witnessed a "crime" staged for his or her benefit. The witness was then shown a set of six pictures and was asked to examine the pictures carefully to try to pick out the person that he or she thought he or she had seen taking the calculator. The witness could also choose "none of them." The witness was given a questionnaire on which he or she indicated his or her choice and rated his or her confidence that he or she had made the correct decision. The confidence-in-decision rating was made on a 7-point scale with endpoints labeled not at all confident (1) and extremely confident (7). The witness was left alone in the cubicle to make these decisions.

The experimenter returned after the witnesses had made their decisions. Any witness who made an identification was debriefed and dismissed from the study at this point. (Witnesses who failed to make an identification were debriefed and dismissed at this point since a real witness who failed to make an identification would not typically be cross-examined. However, see Wells & Lindsay, 1980.) If the witness agreed to be cross-examined, he or she was asked to sign a release form permitting us to videotape the cross-examination for further use. After the release had been signed, the experimenter led the witness to another laboratory in which the cross-examination was conducted. The cross-examiner was blind to the witness' experimental condition and identification choice.

Results and Discussion

The total number of witnesses was determined by monitoring the identification rate results to assure a minimum of 8 accurate- and 8 inaccurate-identification witnesses in each situational-accuracy condition. The high-situational-accuracy condition yielded 32, 12, and 5 witnesses who made accurate, inaccurate, and no identifications, respectively. The moderate-situational-accuracy condition yielded 14, 14, and 4 witnesses who made accurate, inaccurate, and no identifications. The low-situational-accuracy condition yielded 8, 16, and 3 accurate, inaccurate, and no identifications.

In summarizing the situational-accuracy data, we consider the identification data in terms of its forensic properties. That is, our concern, as outlined earlier, is with the rate of accurate versus inaccurate identification witnesses that would appear in court. The forensic identification rate, therefore, is a ratio of the number of accurate identification witnesses who are willing to be cross-examined to the total number of witnesses who make an identification and are willing to be cross-examined. As indicated above, 12 witnesses made no identification and therefore would not normally be part of a cross-examination procedure in an actual case. In addition, as described in Experiment 2, there were 4 witnesses who made

---

1 The functional size of this photo lineup was 2.89 using the procedures for estimating lineup size outlined in Wells, Leippe, and Ostrom (1979).

2 There are, of course, practices among some criminal-justice investigators that involve having the witness "guess" and subsequently bolster the witness' confidence. We did not follow this questionable practice for several reasons, foremost among them the fact that our general hypothesis involved subjects' abilities to detect accurate from inaccurate witnesses, a hypothesis that may be clouded by any witness-bolstering procedures.
identifications but were not willing to be cross-examined. Deleting these witnesses from their respective conditions yields forensic identification rates of 74% (n = 42), 50%, (n = 26), and 33% (n = 24) for the high, moderate, and low situational-accuracy conditions, respectively. These accuracy rates are significantly different from one another, χ²(2) = 10.94, p < .01, leading us to conclude that we successfully manipulated situational accuracy.

Interestingly, there was no evidence that witnesses refused to be cross-examined or failed to make an identification as situational accuracy decreased. There were one, two, and one witnesses refusing to be cross-examined in the low, moderate, and high situational-accuracy conditions, respectively. In addition, there were three, four, and five witnesses in the low-, moderate-, and high-situational-accuracy conditions, respectively, who refused to make an identification. This suggests that the confidence of the witnesses that they could make an accurate identification was not appreciably altered by the situational-accuracy manipulation. However, the number of witnesses refusing to be cross-examined and making no identification is too small to make any definitive conclusions.

A superior test of whether witness confidence decreased with situational accuracy is contained in the analysis of variance (ANOVA) on witnesses’ self-rated confidence. A 2 (individual witness accuracy) × 3 (situational accuracy) between-subjects ANOVA (adjusted via Edwards's, 1972, procedure for unequal cell size) yielded a significant main effect for individual witness accuracy, F(1, 86) = 5.15, p < .05, no effect for situational accuracy, F(2, 86) = 1.08, ns, and no interaction (F < 1). The mean confidence scores for the low-, moderate-, and high-accuracy conditions, respectively, were 4.3, 4.3, and 5.0 for accurate witnesses and 3.4, 3.8, and 3.9 for inaccurate witnesses.

The results do not support our hypothesis regarding a situational accuracy–witness confidence relationship. The results do corroborate Wells et al.’s (1979) data showing a weak but statistically significant difference between accurate and inaccurate witnesses in confidence. The strength or weakness of the individual accuracy–witness confidence relationship is contained in the proportions of variance on the individual accuracy dimension accounted for by witness confidence. The percentage of variance in witness accuracy accounted for by witness confidence in the low-, moderate-, and high-situational-accuracy conditions was 9.0, 1.9, and 9.6, respectively. Because these values are comparable to Wells et al.’s obtained value of 8.4%, we suspect that subjects will not be able to use confidence reliably to detect individual witness accuracy at a level exceeding chance.

Nevertheless, the witness confidence data do not necessarily eliminate our suspicion that subjects who observe cross-examinations may reduce their belief rates as situational accuracy declines. Specifically, it remains a possibility that subjects will take into account (via the cross-examination) the nature of the witnessing situation even though witnesses, as measured by self-rated confidence, do not appear to take the witnessing situation into account.

Experiment 2

Method

Subjects were 96 introductory psychology students who participated in partial fulfillment of a course requirement. On arrival, each subject was directed to a viewing room that included a television monitor and a headphone set. Subjects were told that a crime had been staged some time ago for the benefit of a witness. The witness had agreed to be cross-examined with regard to his/her identification and the subjects would see a videotape of that cross-examination. Each subject was told that his or her role would be that of a juror attempting to decide whether the witness made an accurate identification of the criminal or had identified an innocent person. Subjects were told that this would be done on the basis of observing a cross-examination of the witness.

Although each subject eventually observed four witnesses, subjects were unaware of the fact that there would be more than one witness until all data were collected on the first witness. Also, subjects did not know the total number of witnesses to be observed until after their judgments of the final witness were made. Keeping subjects uncertain regarding whether another witness would be observed helped assure that the subjects’ judgments of belief were made independently for each witness. This procedure also allowed for a test of the dependence or independence of subjects’ belief decisions across trials by comparing first decisions to later decisions.
Cross-examinations of witnesses. As noted earlier, each witness in Experiment 1 who made an identification was asked to sign a release form indicating his or her willingness to be cross-examined and have the cross-examination videotaped. Four of the witnesses refused to be cross-examined: one inaccurate witness from each of the three situational-accuracy conditions and one accurate witness from the moderate situational-accuracy condition. Each willing witness was then escorted to a courtroom to be cross-examined by a second experimenter. Observers, described to the witness as jurors, were present in the courtroom, creating the atmosphere of a trial.

The cross-examination included a swearing in of the witness followed by 15 questions. The questions included open-ended enquiries (e.g., “Describe what the person was wearing?”), questions that required the respondent to use a specific dimension (e.g., “How long was the person in the room?”), questions that probed the extent to which the conditions of identification were facilitative or inhibitive of accuracy (e.g., “Was that really long enough to get a good look at the thief?”), and questions that challenged the witnesses’ credibility (e.g., “Isn’t it possible, just possible, that you saw my client somewhere else and remembered his face even though he was not in the room?”).3

The cross-examiner was described as an attorney for the defense of the accused and delivered the examination consistent with that role (i.e., attempting to discredit the witness). The cross-examiner was kept blind to the true accuracy of the witnesses until the conclusion of the experiment.4 Videotaping was done with a high-quality camera that was stationary throughout the experiment. The camera field encompassed the sitting witnesses from the ankles to the head.

Sampling of witnesses. There were 8 accurate-identification and 8 inaccurate-identification witnesses from each of the three situational-accuracy conditions (for a total of 48 witnesses) from Experiment 1 who were presented to subjects in this experiment. To assure representativeness of these 48 witnesses, the average confidence of witnesses within each cell of the 3 (situational accuracy) × 2 (individual accuracy) design was as close as possible to that obtained in Experiment 1. The sampled means for witness confidence for the low-, moderate-, and high-situational-accuracy conditions, respectively, were 4.3, 4.3, and 5.0 for accurate witnesses and 3.5, 3.8, and 4.0 for inaccurate witnesses.

Dependent measures. After observing each cross-examination, the subjects made a decision on whether to believe that the witness had made an accurate identification by pressing one of two buttons. Following each decision, subjects rated their confidence that they (the subjects) had made a correct decision regarding whether to believe the witness by pressing one of seven buttons labeled from “not at all sure” (1) to “extremely sure” (7).

Design. Each subject observed four witnesses who were all selected from the same cell of the 3 (situational accuracy) × 2 (individual accuracy) design. Because there were eight witnesses in each cell of the 3 × 2 design, we treated the two sets of four witnesses as a between-subjects replication factor, yielding 12 cells with four witnesses each. In addition, dividing witnesses into those with confidence scores of 5 or more versus those with confidence scores of 4 or less yielded two witnesses in each of the 12 cells who could be characterized as high-confidence witnesses and two witnesses in each cell characterized as low-confidence witnesses. Thus, the full design was a 3 (situational accuracy) × 2 (individual accuracy) × 2 (witness confidence level) × 2 (replication) factorial with the witness confidence level as a within-subjects variable.

Results and Discussion

There were no main effects or interactions of the replication factor on any of the dependent measures (belief, correctness of decision, or confidence in decision). Thus, it is unlikely that the pattern of results to be reported was produced by the unique characteristics of the witnesses who were used as stimuli. All subsequent analyses have been collapsed over the replication factor.

To test the influence of presenting multiple witnesses to the subjects, order effects on the dependent measures were examined. Only the results for the belief measure are discussed here, although the results were similar for correctness-of-decision and confidence-in-decision measures. Collapsing across the independent variables, the rates of belief for witnesses presented first, second, third, and fourth were 68%, 68%, 69%, and 71%, respectively. These rates do not differ significantly (F < 1). Also, the pattern in the overall results is present in the subjects’ first decisions. For example, the overall rates of belief for all witnesses, all high-confident witnesses, and all low-confident witnesses were 68%, 77%, and 59%, respectively. Belief rates of all first witnesses, high-confident first witnesses, and low-confident first witnesses were 66%, 74%, and 59%, respectively. Clearly the data from the first witnesses closely resembles the overall data. It seems reasonable to assume, therefore, that the repeated-measures nature of the current design did not affect the general results.

Belief data. The belief data were analyzed by assigning a score of 0%, 50%, or 100% for the percentage of witnesses that

3 Cross-examination scripts are available from the second author.
4 The third author served as the cross-examiner.
Table 1
Percentages of Subjects' Decisions to Believe Witnesses as Functions of the Confidence and Situational-Accuracy Conditions of Witnesses

<table>
<thead>
<tr>
<th>Witness confidence</th>
<th>Situational accuracy</th>
<th></th>
<th></th>
<th>M</th>
</tr>
</thead>
<tbody>
<tr>
<td>Low</td>
<td>Low</td>
<td>47</td>
<td>54</td>
<td>76</td>
</tr>
<tr>
<td></td>
<td>Moderate</td>
<td>76</td>
<td>76</td>
<td>78</td>
</tr>
<tr>
<td></td>
<td>High</td>
<td>62</td>
<td>66</td>
<td>77</td>
</tr>
</tbody>
</table>

each juror believed in each cell of the 3 (situational accuracy) x 2 (individual accuracy) x 2 (level of witness confidence) design. The ANOVA showed no main effect or interaction for individual witness accuracy. However, there was a main effect for witness confidence, F(1, 84) = 13.50, p < .001, a main effect for situational accuracy, F(2, 84) = 3.27, p < .05, and a significant Witness Confidence x Situational Accuracy interaction, F(1, 84) = 3.02, p = .05. Table 1 presents the mean percentage scores for the situational-accuracy and witness-confidence conditions.

As evidenced by Table 1, the main effect for witness confidence is due to a tendency for subjects to believe high-confident witnesses more than low-confident witnesses. The main effect for situational accuracy is due to a tendency for subjects to believe high-situational-accuracy witnesses more than moderate- or low-situational-accuracy witnesses. Note, however, that both main effects are modified by the interaction, which indicates that although subjects reduced their belief rates as situational accuracy decreased, the effect is almost totally due to the effect occurring in the low-witness-confidence conditions. We suspect that the reason subjects only modified their belief rates when dealing with low-confidence witnesses was due to the low-confidence witnesses increasing the subjects' feelings of uncertainty. This in turn made subjects search for other possible predictors of accuracy (i.e., the witnessing situation).

Correctness of subjects' decisions. We also analyzed the belief data by converting it into a correctness-of-decision index (as in Wells et al., 1979). This was done by assigning a score of 0, 50, or 100 for the percentage of correct belief decisions made by each subject in the 3 (individual accuracy) x 2 (situational accuracy) x 2 (confidence of witness) design. A correct decision is either believing an accurate witness or not believing an inaccurate witness. Although this analysis is statistically redundant with the belief analysis, it is an instructive aid for a somewhat different view of the data. Specifically, it allows a more direct assessment of how well subjects were performing across the conditions. This performance measure is particularly important to the extent that one wishes to analogize the subjects to jurors. This analysis more directly answers such questions as whether the subjects were more likely to make decision errors in one set of circumstances (conditions) than in another.

The ANOVA on the correctness measure showed that situational accuracy had no main effect on the correctness of subjects' decisions. This means that subjects were no more likely to make distinctions between accurate and inaccurate witnesses in one situational-accuracy condition than in any other situational-accuracy condition (see Figure 1). There was, however, an individual accuracy main effect, F(1, 84) = 57.22, p < .001, which takes the form of subjects being far more likely to be correct when viewing accurate witnesses (M = 70%) than when viewing inaccurate witnesses (M = 33%). This main effect is simply due to the fact that subjects had a tendency to believe witnesses more than not believe them, which in turn made subjects more likely to be correct when the witness should have been believed (i.e., when the witness was accurate) than when the witness should not have been believed (i.e., when the witness was inaccurate).

Finally, the correctness-of-decision analysis revealed a significant Individual Accuracy x Situational Accuracy interaction, F(1, 84) = 3.80, p < .05. The percentages of correct subject decisions for the low-, moderate-, and high-situational-accuracy conditions, respectively, were 63, 66, and 80 for the accurate witnesses and 39, 34, and 25 for the inaccurate witnesses. The pattern of
this interaction reflects the fact that the change in subjects' belief rates across conditions only affected the type of error that was made (i.e., believing an inaccurate witness or disbelieving an accurate witness) rather than the likelihood that an error was made. Specifically, the increase in subjects' belief rates across conditions decreased the probability that subjects would make decision errors only if the witness was accurate and increased the probability that subjects would make decision errors if the witness was inaccurate.

Estimates of subjects' correctness when considering base rates. The data regarding subjects' correctness of decisions across levels of situational accuracy change somewhat when controlling for the base rates of witness accuracy. That is, the factorial design of the current study represents accurate and inaccurate witnesses equally across levels of situational accuracy. As demonstrated in Experiment 1, however, the cells of the 3 (situational accuracy) X 2 (individual accuracy) are not equally represented in nature. For example, if we had sampled witnesses randomly from an event such as our high-situational-accuracy condition, the expected probability that any given witness would be accurate is .74. Similarly, the probability that any given witness from the low-situational-accuracy condition would be accurate is .33. These probabilities are derived from the forensic witness accuracy rates obtained in Experiment 1.

When we consider the fact that accurate and inaccurate witnesses are not, in certain real-world conditions, equally likely to be cross-examined, it becomes apparent that errors on the part of our subjects who are judging those witnesses will not be equally distributed across those conditions. Specifically, any changes across conditions in the prior likelihood of viewing an accurate witness must be compared with any change across conditions in the likelihood of subjects believing the witnesses.

Because subjects believed accurate and inaccurate witnesses equally within situational-accuracy conditions, we can ignore
this within-condition factor. The estimated likelihood of a subject making a correct belief decision can be considered solely on the basis of the correspondence between the probability of subjects believing/disbelieving witnesses in a given condition and the prior probability that a witness would be accurate/inaccurate in that condition. The estimated probability of a correct belief decision is then the sum of two products: (a) prior probability of witness accuracy times the probability of subject belief plus (b) prior probability of witness inaccuracy times the probability of subject disbelief. Calculating this for each situational-accuracy condition shows that an estimated 46%, 50%, and 63% correct decisions would have been made by our subjects had we randomly sampled witnesses from our low-, moderate-, and high-situational-accuracy conditions, respectively. This estimated correctness of subjects' belief decisions when controlling for witness accuracy base rates is depicted in Figure 1.

This estimated-correctness analysis is not the only possible measure of the expected correctness of our subjects' decisions. We can also directly measure the congruence between the probabilities of subjects' belief of witnesses and witness accuracy by taking the difference between the two probabilities in each situational-accuracy condition. The results show values of .03, .16, and .29 for the high-, moderate-, and low-situational-accuracy conditions, respectively. This suggests that subjects were best calibrated in the higher situational-accuracy conditions.

Subjects' confidence. The confidence data were scored by using the mean confidence for each subject within each of the 12 cells. The ANOVA produced a significant main effect for witness confidence, \( F(1, 84) = 18.53, p < .001 \), and a Witness Confidence \( \times \) Situational Accuracy interaction, \( F(1, 84) = 6.96, p < .01 \). The main effect is due to subjects being more confident of their own decisions when the witness was high in confidence (\( M = 5.4 \)) than when the witness was low in confidence (\( M = 4.9 \)). The interaction is due to this difference being greater in the low-situational-accuracy conditions than in the other situational-accuracy conditions. At this point we do not fully understand this interaction; the main effect for witness confidence on subjects' confidence in their own decisions, however, corroborates our previous supposition that low witness confidence increases subjects' uncertainty, perhaps causing subjects to search for other predictors of witness accuracy, most notably the situation.

General Discussion

Experiment 1 corroborated previous findings showing a relatively small relationship between individual witness accuracy and witness confidence. Percentages of witness accuracy accounted for by witness confidence have varied from 0 (Brown et al., 1977; Leippe et al., 1978) to 6.8 (average across cells of Experiment 1 in this study) to 8.4 (Wells et al., 1979). Assuming that subjects have no better predictor of witness accuracy than witness confidence, and assuming the strength of the correlation between witness confidence and attributions of witness confidence to be around .55 (as in Wells et al., 1979), we estimate that it would take a minimum of 100 witnesses and as many subjects to produce statistically significant (\( p = .05 \)) evidence that people believe accurate more than inaccurate witnesses within situational-accuracy conditions. Even under these conditions the difference between the probability of a person believing an accurate versus an inaccurate witness would be negligible (around .025).

Experiment 1 suggests that witnesses show no particular sensitivity to the fact that they witnessed a crime under conditions in which the probability of their making a false identification increases over 250% across conditions (i.e., from 26% to 67%). Those witnesses who were in conditions producing 67% identifications of an innocent party were no less likely to make an identification, no less willing to be cross-examined, and no less confident than were witnesses in conditions that yielded only 26% false identifications.

Experiment 2 showed that the subjects had some appreciation for the witnessing conditions in that they reduced their belief rates from 77% in the high-situational-ac-
curacy condition to 62% in the low-situational-accuracy condition. On the other hand, subjects' compensation for the witnessing conditions fails to match the slope of change in witness accuracy rates. This discrepancy is quantitatively described as a Linear (witness accuracy across situational accuracy conditions) × Linear (subjects' belief of witnesses across situational accuracy conditions) interaction, $F(1, 84) = 4.12, p < .05$.

As depicted in Figure 1, this trend interaction follows a pattern wherein the tendency for subjects to be overbelieving of witnesses is inversely related to the situationally induced accuracy rate of witnesses. Thus, our earlier question of whether jurors are overbelieving of witnesses can be answered with a qualified “yes.” Specifically, to the extent that our subjects' decision processes can be generalized to jurors, then jurors appear overbelieving of witnesses when the probability of witness accuracy is low and particularly if witness confidence is also high. The next section addresses this question of generalization.

**Generalizing to the Courtroom**

To what extent can parallels be drawn between the outcomes of Experiments 1 and 2 and the criminal justice process? There are, of course, several differences. First, the witnesses in Experiment 1 knew that it was not a real crime at the time they indicated their willingness to be cross-examined. This simulated setting may overestimate people's general willingness to follow through with an accusation and testify in court. Nevertheless, the witnesses in this situation had every reason to want to be correct and to admit low confidence if they felt uncertain about their accuracy because they were aware that the experimenter knew the correct answer. Given the negligible relation between witness confidence and witness accuracy, therefore, we could expect a real crime with real consequences to produce similar ratios of accurate to inaccurate witnesses willing to be cross-examined, though the overall rate of witnesses willing to testify may be lower than that obtained here. In other words, whereas a real criminal situa-

---

Interval distances for trend analysis are based on the witness accuracy rates within conditions. The inequality of distances and cell sizes are corrected using Gaito's (1965) procedure.
Another aspect of our subjects' task that is not in line with the ecology of a juror's task is that our subjects saw four witnesses and made four judgments. Nevertheless, our analysis of the data revealed that this repeated-measures factor had no effect on the pattern of the results, since subjects' decisions on Trial 1 produced the same pattern of results as was found over all four trials.

Another aspect of our situation that is unlike the typical court case is the short interval between the witnessing of the event and the cross-examination. However, short intervals provide little opportunity for interference, leveling, or assimilation, which in turn should have provided good conditions for witness-to-subject communications (Allport & Postman, 1947). Thus, we might be less optimistic about the ability of witnesses to effectively communicate to jurors in an actual case (since the passage of time should enhance distortions) than was found in the current studies.

Finally, there are several differences between the subjects in Experiment 2 and actual jurors that we cannot address at this time in terms of their potential import, namely age, social status, and related differences between university students and those on jury rolls. We have no reason at this point to believe that a sample of jurors with a mean age of 43 would respond to witnesses differently than would 22-year-olds or that people with steady jobs would respond differently than would people who are yet to enter the job market. But possible differences cannot be dismissed without further evidence.

There are other types of generalization issues that need empirical assessment. For example, future research should address whether or not other variables that have marked effects on the rate of false identifications (e.g., race of witness and race of criminal; see Brigham & Barkowitz, 1978; Malpass & Kravitz, 1969) produce sufficient levels of discounting on the part of jurors. Similarly, research should address whether or not the wording of witness interrogations (e.g., Loftus, 1975; Loftus, Altman, & Geballe, 1975; Loftus, & Palmer, 1974; Loftus & Zanni, 1975) ultimately affects jurors' decisions and to what extent jurors are appreciative of the degree to which unconscious transference (Loftus, 1976) affects false identification rates. Psychologists' ability to benefit jurors' decisions regarding eyewitness identification requires both an understanding of the distribution of witnesses' false alarms (Goldstein, Stephenson, & Chance, 1977) and empirical examination of jurors' implicit eyewitness theories.

It is difficult to know how expert psychological advice on eyewitness matters should be handled in the courtroom, or whether it is needed in the courtroom, unless we understand the cognitive processes of people's responses to eyewitnesses. The current paradigm, originally developed in Wells et al. (1979), provides the means for a literature that will assess weaknesses in jurors' implicit theories of eyewitness performance.

References


Received January 16, 1980