Journal of Applied Psychology
1989, Vol. 74, No. 2, 356-359

0021-9010/89/$00.75

Eyewitness Accuracy and Confidence: Within-Versus Between-Subjects Correlations

Vicki L. Smith
Northwestern University

Saul M. Kassin
Williams College

Phoebe C. Ellsworth
University of Michigan

Previous researchers using between-subjects comparisons have found eyewitness confidence and accuracy to be only negligibly correlated. In this study, we examined the predictive power of confidence in within-subject terms. Ninety-six subjects answered, and made confidence ratings for, a series of questions about a crime they witnessed. The average between-subjects and within-subject accuracy–confidence correlations were comparably low: $r = .14$ ($p < .001$) and $r = .17$ ($p < .001$), respectively. Confidence is neither a useful predictor of the accuracy of a particular witness nor of the accuracy of particular statements made by the same witness. Another possible predictor of accuracy, response latency, correlated only negligibly with accuracy ($r = -.09$ within subjects), but more strongly with confidence ($r = -.27$ within subjects). This pattern was obtained for both between-subjects and within-subject comparisons. The theoretical and practical implications of these results are discussed.

Common sense suggests that the confidence an eyewitness expresses is a good indicator of the accuracy of the testimony. This intuition appears to be widely held, with 56% of jurors (Brigham & Bothwell, 1983), 76% of undergraduates (Deffenbacher & Loftus, 1982), 73% of police officers, 75% of prosecuting attorneys, and 40% of defense attorneys (Brigham & Wolfskiel, 1983) believing that eyewitness confidence and accuracy are positively correlated. Even the United States Supreme Court, in the case of *Neil v. Biggers* (1972), specified eyewitness confidence as an important predictor of identification accuracy and later reaffirmed this standard in *Manson v. Brathwaite* (1976).

Despite intuition and the Court's assertions, a considerable amount of psychological research indicates that eyewitness confidence is not a reliable predictor of accuracy. Wells & Murray (1984) report that in 18 of the 31 studies they reviewed, accuracy and confidence were not significantly correlated. Even in those studies for which a significant relation was obtained, confidence rarely accounted for more than 5% of the variance in accuracy. The empirical evidence, therefore, indicates that common sense is wrong: A confident witness is not more likely to be accurate than a hesitant witness. This suggests that it is inappropriate to discount the testimony of uncertain witnesses or to trust the accounts of confident witnesses. Confidence does not reflect relative accuracy.

Eyewitness researchers have focused exclusively on the accuracy–confidence relation *between* subjects, comparing the identification accuracy of confident witnesses to the accuracy of hesitant witnesses. But what about the correspondence of accuracy and confidence *within* a given witness? Are confident statements made by a particular witness more likely to be accurate than hesitant statements made by the *same* witness? This question has important implications for the evaluation of eyewitness testimony both in and out of court. When police officers, prosecutors, and jurors listen to a witness recounting a sequence of events, they use this narrative to decide what really happened. It is likely that decision makers place greater weight on a witness's confident statements than on the less confident ones. But does expressed confidence reliably predict accuracy *within* subjects?

Although the eyewitness literature has not considered the within-subject correlation of accuracy with confidence, research in the verbal learning domain suggests that the correlation should be positive and large. For example, in some studies subjects were asked a series of general knowledge questions. For those questions that were not answered correctly, subjects rated the probability that they would recognize the correct answer on a multiple-choice test. The results typically indicate that these feeling-of-knowing ratings are useful predictors of subsequent recognition (e.g., Blake, 1973; Hart, 1965, 1966, 1967; Nelson, Gerler, & Narens, 1984; Schacter, 1983).[1]

This research was conducted while Vicki Smith was supported by a National Science Foundation Graduate Fellowship. Saul Kassin was on sabbatical from Williams College and supported by a postdoctoral fellowship provided by the National Institute of Health to the Psychology-and-Law Program at Stanford University. Phoebe Ellsworth was a Fellow at the Center for Advanced Study in the Behavioral Sciences, Stanford, California.

Phoebe Ellsworth is grateful to the John D. and Catherine T. MacArthur Foundation, to the James McKeen Cattell Fund, and to the Center staff and Fellows for time, space, unsurpassable services, and intellectual stimulation.

We would like to thank Elizabeth Loftus for providing the slides used in this study, and some of the memory questions. We would also like to thank Patrice Callahan and Doris Patton for their help in testing subjects.

Correspondence concerning this article should be addressed to Vicki L. Smith, Department of Psychology, Northwestern University, Evanston, Illinois 60208.

[1] Most studies of feeling-of-knowing consider only a subset of items that the subject is unable to recall. However, recent data collected by Vicki L. Smith indicate that these feeling-of-knowing ratings are also

Similar results were obtained in experiments on text recall (Stephenson, 1984; Stephenson, Brandstatter, & Wagner, 1983; Stephenson, Clark, & Wade, 1986). Subjects in these studies either heard or read a short text. They were then asked to recall details of the text and rate their confidence in their recall. Subjects' confidence in right answers significantly exceeded their confidence in wrong answers. In one study, Stephenson (1984) measured the within-subject correlation of accuracy with confidence directly and found that despite large individual differences, the average within-subject correlation was large, $r = .48$.

These studies suggest that a person's expressions of confidence can, under some circumstances, reflect relative states of knowing. It is not clear, however, that these findings generalize beyond verbal learning to eyewitness testimony. Eyewitnesses are exposed to events that are perceptually richer, more complex, and more stressful than a general knowledge test or reading text. Situational pressures inherent in being a witness to a crime may also affect the calibration of accuracy and confidence. For example, the feeling that "I was there, so I should know" may prompt eyewitnesses to fill in details they do not remember clearly or to overstate their confidence in their memories. This could substantially attenuate the correlation of accuracy and confidence that emerges in other contexts.

Two face recognition studies have considered the calibration of accuracy and confidence in tasks that more closely approximate an eyewitnessing situation by testing recall for visual, rather than verbal, stimuli. Brown, Deffenbacher, and Sturgill (1977) showed subjects two live groups of five "criminals" each. An hour and a half later, subjects viewed a series of mugshots. For each, they determined whether or not the photo represented one of the criminals and rated their confidence in their decision. One week later, subjects made lineup identifications and again rated their confidence. In both the mugshot and lineup phases, the between-subject accuracy–confidence correlation was nonsignificant ($r = .05$, for mugshot; $r = .12$, for lineup). Moreover, across subjects, mean confidence for correct answers was highly correlated with mean confidence for incorrect answers ($r = .51$, $p < .01$, for mugshot; $r = .70$, $p < .01$, for lineup). This latter finding indicates that people who are confident when they are correct also tend to be confident when they are incorrect. Brown et al. did not, however, report within-subject correlations of accuracy and confidence.

Deffenbacher, Leu, and Brown (1981) showed subjects 50 photographs of faces. One week later, subjects tried to identify these faces and rated their confidence in their identifications. The within-subject correlations of accuracy and confidence ranged from −.05 to +.60, with a mean of .31. Although these data suggest a reasonably large positive correlation within subjects on this face recognition task, it is not clear whether the results of this experiment would generalize to other eyewitnessing situations. The accuracy–confidence correlation between subjects in this experiment was .48, which is large compared with most other studies. It is possible that the ideal viewing conditions under which this face recognition task was performed account for the large between-subjects correlation (see Bothwell, Deffenbacher, & Brigham, 1987; Deffenbacher, 1980). If so, these optimal viewing conditions may also have increased the within-subject correlation, making generalization to nonoptimal eyewitnessing situations problematic.

Furthermore, in both of these studies, the face recognition tasks focused on multiple identifications, yet seldom in the real world does a single eyewitness identify multiple suspects. Thus, information about a witness's confidence in one identification relative to other identifications by the same witness is usually unavailable to the trier of fact. More frequently, witnesses describe their recollections of a sequence of events. In this study, we sought to test the within-subject correlation of accuracy and confidence for this kind of testimony. Subjects watched an event, answered a series of questions, and rated their confidence in each of their answers. With these data we were able to make both within-subject and between-subjects comparisons of accuracy and confidence.

A second objective was to explore the utility of response latency as a predictor of accuracy. In between-subjects terms, Kassin (1985) found that accurate responses were made somewhat more quickly than inaccurate responses ($r = -.17$) but that subjects' response times were highly correlated with their confidence ratings: the less time subjects took to respond, the more confident they were ($r = -.40$). Although the causal relation between these variables is as yet uncertain, these data suggest that witnesses may be inferring the accuracy of their own decisions, and hence their confidence, from response latency cues. In our research, these findings were explored on a within- as well as a between-subjects basis.

Method

Ninety-six Stanford undergraduates participated individually in sessions lasting approximately 1 hr. They first watched a slide presentation of an automobile accident in which a pedestrian was hit by a car (cf. Loftus, Miller, & Burns, 1978). Each slide was visible for two seconds, and the entire series lasted approximately 1½ min. Subjects then completed a 20-min filler task unrelated to the experiment, after which they answered 20 questions about the slides. The questions were presented sequentially on an IBM microcomputer in a two-alternative forced-choice format. Subjects pressed the return key as soon as they were ready to respond and at the same time stated their answer to the experimenter. The computer recorded response latency, and the experimenter recorded the specific answers. Subjects were told: "Your responses to these questions are being timed, so please respond as quickly as you can. It is very important, however, that you not make any errors, so make sure you know the correct answer before pressing the return key. If you find yourself making errors, please slow down."

Following each question, subjects entered their confidence estimates on a 10-point scale (where 1 = *pure guess* and 10 = *absolutely sure*). Twenty questions about the slides were asked in this manner. After another 20-min filler task, subjects answered an additional 20 questions.[2]

strong predictors of subsequent recognition when all questions are considered: those for which recall was successful and those for which recall failed; overall $r = .55$ within subjects.

[2] Our experimental procedure was designed to test two separate issues: the within-subject accuracy–confidence relation and the effects of misleading information on original memory for an event. The latter purpose required two test periods separated by a filler task. One question in the first group of 20 presented misleading information to half of the subjects; one question from the second group of 20 assessed subjects' endorsement of the misleading information. The results of the misleading information analyses were inconclusive and are not reported here. The two questions involving misleading information were eliminated

The memory questions used in this study were presented in forced-choice format. Eyewitnesses to events in the real world are generally not required to respond to forced-choice questions but have the option of saying "I don't know." Although the forced-choice format in this study precluded a specific "I don't know" response, subjects could reflect their uncertainty in their choices by entering a "1" on the confidence scale, indicating that their choice was a *pure guess*.

Results and Discussion

Overall, witnesses were correct on an average of 63% of their answers (range = .45 to .79; $SD = 0.08$), exceeding chance performance at $p < .01$ ($z = 2.5$). Their mean level of confidence was 5.33 (range = 2.61 to 7.58; $SD = 1.18$). Neither accuracy nor confidence showed floor or ceiling effects that might artificially depress the correlations.

We computed the accuracy–confidence correlation across subjects for each of the 38 questions, averaged these correlation coefficients, and found that the average between-subjects correlation was positive, but small, $r = .14$ (median $r = .20$; range = −.32 to +.44). To test whether it was significant, we transformed the correlation coefficient for each question into a z score, averaged the z scores, and tested this average against zero. Because the variability of the correlation coefficients was small (SD of z scores = .20) and the number of questions included in the average relatively large, this average z score was significantly greater than zero, $t(37) = 4.59$, $p < .001$. Still, confidence accounted for only 2% of the variance in accuracy. These results confirm those of past studies: Witnesses who are confident in their testimony are not substantially more accurate than those who are not.

Next, we computed accuracy–confidence correlations for each subject across the 38 questions and averaged these correlation coefficients for all 96 subjects.[3] As with the between-subjects correlation, the average within-subject correlation was positive, but small, $r = .17$ (median $r = .17$; range = −.21 to +.60). To test its significance, each subject's correlation coefficient was transformed into a z score, the z scores were averaged, and the average tested against zero. As before, although the correlation was significantly greater than zero, $t(95) = 10.69$, $p < .001$, confidence accounted for only 3% of the variance in accuracy.

We considered the possibility that accuracy and confidence might be more highly correlated within subjects who were characteristically more or less confident. For example, the less confident subjects might not have rated themselves as confident except on those items for which they were absolutely sure, and they may have been right about those. This turned out not to be the case. For subjects above the median in confidence (median = 5.25), the accuracy–confidence correlation was .16; for those below the median, the correlation was .17.

It is not clear why the eyewitness context differs from the verbal learning domain in the correspondence of accuracy and confidence within subjects. Why are subjects who are asked to recall general knowledge information, or details of a text they have read, able to correctly assess the contents of their memories and reflect that information in their expressions of confidence, whereas eyewitnesses are not? It is possible that our within-subject correlations were attenuated by situational pressures that encourage eyewitnesses to confidently assert all of their responses (regardless of how sure they really are). If this is the case, subjects' confidence ratings should be uniformly high. However, confidence was quite variable within subjects: 96% of our subjects expressed confidence that varied across 7 or more of the 10 possible scale values. Across all 96 subjects, the average within-subject standard deviation on the confidence scale was 2.67. Thus, uniformly high confidence does not account for the small correlations we obtained.

Turning to response latency, the average between-subjects and within-subject correlations with accuracy were of similar magnitude: $r = -.05$ ($p < .05$) and $r = -.09$ ($p < .001$), respectively. The negative correlations indicate that accurate responses were made somewhat more quickly than inaccurate responses. Although these correlations were significant, response time accounted for less than 1% of the variance in accuracy. The correlations of response time with confidence were also comparable: between-subjects $r = -.28$ ($p < .001$) and within-subject $r = -.27$ ($p < .001$). Thus, subjects were more confident in answers they made quickly than in those made more slowly. These correlations were significant, with response time accounting for approximately 8% of the variance in confidence.

These results replicate those obtained by Kassin (1985). Response time was more highly correlated with confidence than with accuracy. This pattern holds whether the comparison is made between or within subjects. As Kassin noted, this pattern suggests that if subjects base their confidence estimates on response time (inferring accuracy by observing their own behavior), they are misleading themselves.

In sum, whether the confidence–accuracy question is framed in the between-subjects terms considered in previous research ("Are confident eyewitnesses more accurate than hesitant ones?"), or in the within-subject terms considered in this study ("Are a witness's confident statements more accurate than his or her hesitant ones?"), the answer is the same: Confidence is not a good predictor of accuracy. Common sense and the Supreme Court notwithstanding, confidence is not a useful indicator of the accuracy of a particular witness or of the accuracy of particular statements made by the same witness. The present data indicate that relying on confidence to assess the credibility of a witness's statements may be dangerously misleading. Probative evidence may be ignored because it is not confidently asserted, and errors believed because the witness is certain.

from our analyses of the accuracy–confidence relation. We used each subject's answers and confidence ratings for the remaining 38 questions to address the accuracy–confidence issue.

[3] Because half of our subjects received misleading information in one of the two questions excluded from these analyses, we first tested to see whether this manipulation had any effect on the accuracy–confidence relation for the remaining questions. When we compared those subjects who received misleading information with those who did not, we found that the average within-subject correlations did not differ significantly: $r = .19$ and $r = .16$, respectively ($z = 0.13$, *ns*). The data for these two groups were pooled for the remaining analyses.

References

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

Bothwell, R. K., Deffenbacher, K. A., & Brigham, J. C. (1987). Correlation of eyewitness accuracy and confidence: Optimality hypothesis revisited. *Journal of Applied Psychology, 72,* 691–695.

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., & Wolfskiel, M. P. (1983). Opinions of attorneys and law enforcement personnel on the accuracy of eyewitness identification. *Law and Human Behavior, 7,* 337–349.

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

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

Deffenbacher, K. A., Leu, J. R., & Brown, E. L. (1981). Memory for faces: Testing method, encoding strategy, and confidence. *American Journal of Psychology, 94,* 13–26.

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

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

Hart, J. T. (1966). Methodological note on feeling-of-knowing experiments. *Journal of Educational Psychology, 57,* 347–349.

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

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

Loftus, E. F., Miller, D. G., & Burns, H. J. (1978). Semantic integration of verbal information into a visual memory. *Journal of Experimental Psychology: Human Learning and Memory, 4,* 19–31.

Manson v. Brathwaite, 432 U.S. 98, 97 S.Ct. 2243 (1976).

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

Nelson, T. O., Gerler, D., & Narens, L. (1984). Accuracy of feeling-of-knowing judgments for predicting perceptual identification and relearning. *Journal of Experimental Psychology: General, 113,* 282–300.

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

Stephenson, G. M. (1984). Accuracy and confidence in testimony: A critical review and some fresh evidence. In D. J. Muller, D. E. Blackman, & A. J. Chapman (Eds.), *Psychology and law: Topics from an international conference* (pp. 229–248). Chichester, England: Wiley.

Stephenson, G. M., Brandstatter, H., & Wagner, W. (1983). An experimental study of social performance and delay on the testimonial validity of story recall. *European Journal of Social Psychology, 13,* 175–191.

Stephenson, G. M., Clark, N. K., & Wade, G. S. (1986). Meetings make evidence? An experimental study of collaborative and individual recall of a simulated police interrogation. *Journal of Personality and Social Psychology, 50,* 1113–1122.

Wells, G. L., & Murray, D. M. (1984). Eyewitness confidence. In G. L. Wells and E. F. Loftus (Eds.), *Eyewitness testimony: Psychological perspectives* (pp. 155–170). New York: Cambridge University Press.

Received October 29, 1986
Revision received July 12, 1988
Accepted July 16, 1988 ■