Psychological Science, 1998, **9**, 215-218. (www.psychologicalscience.org/journals/ps/)

Accuracy and Confidence in Person Identification: The Relationship is Strong when Witnessing Conditions Vary Widely

D. Stephen Lindsay
University of Wales, Bangor, United Kingdom; University of Victoria, BC, Canada

J. Don Read
University of Lethbridge, Alberta, Canada

Kusum Sharma
University of Victoria, British Columbia, Canada

We hypothesized that both accuracy and confidence in suspect identifications depend, in part, on participants' ability to identify the target, and that both accuracy and confidence therefore tend to be higher under conditions that lead to good memory for the target. Furthermore, we hypothesized that a substantial correlation between accuracy and confidence will be observed if there is considerable variability across participants in ability to identify the target due to variations in conditions. Consistent with these hypotheses, manipulations that affected accuracy also affected confidence in the same direction, and when data were collapsed across conditions the accuracy-confidence correlation was substantial (mean r = .59).

Psychologists have made hay out of the finding that the correlation between accuracy and confidence in person identification is surprisingly weak (e.g., Bothwell, Deffenbacher, & Brigham, 1987, reported an average correlation of r = .25 across 35 field studies). Interest in this finding is partly due to its counterintuitive nature: Counterintuitive findings fascinate students and demonstrate the superiority of science over common sense. This particular counterintuitive finding is especially powerful because it has enabled psychologists to inform judges and juries that confident eyewitnesses are little more likely to be accurate than non-confident ones.

While not doubting the advantages of science, we suspect that in studies of the accuracy-confidence (AC) relationship a misapplication of scientific precepts has led researchers astray. Although some recent reviews have argued that the AC relationship can be strong under some conditions (e.g., Sporer, Penrod, Read, & Cutler, 1995), none has noted what may be a fatal flaw in the evidence for the general claim that confident identifications are little more likely to be accurate than non-confident ones. Our argument rests on two postulates:

1. Both accuracy and confidence are based, in part, on ability to identify the target (e.g., people would tend to be both more accurate and more confident in identifications of their mothers than in identifications of a briefly glimpsed stranger).

2. Across real-world cases, there is great variability in factors that may affect witnesses' ability to identify the perpetrator (e.g., attention when witnessing, duration of witnessing, delay before test, distinctiveness of the culprit, test difficulty, etc.). Thus it is likely that real-world witnesses vary greatly one from another in ability to identify the perpetrator.

Combined, these postulates suggest that the AC relationship across real-world cases is substantial. From this perspective it is ironic that in previous studies of the AC relationship researchers have used their methodological expertise to minimize variability in ability to identify the target by holding procedures constant across participants. Of course, even in highly controlled studies there is some variability across participants in memory for the target due to individual differences and uncontrolled extraneous variables, but controlled procedures minimize such variability, thereby eliminating the primary basis for the AC relationship across witnesses.

Consistent with this idea, Read, Lindsay, and Nicholls (1997) obtained strong AC relationships when

This research was supported by operating grants to each of the first two authors from the Natural Science and Engineering Research Council of Canada, and by a grant to the second author from the Alberta Law Foundation. We thank Elizabeth Brimacombe for the use of videotapes from which the videos used in this experiment were created, and Mike Hunter and John Vokey for statistical advice.

Correspondence concerning this work can be addressed to D. Stephen Lindsay, Department of Psychology, University of Victoria, P.O. Box 3050 STN CSC, Victoria, B.C., Canada; slindsay@uvic.ca

participants were tested under conditions likely to increase variability in ability to identify the target (e.g., very long delays). The current experiment directly tested the hypothesis that the AC relationship is relatively strong when there is substantial variability across participants in memory for the target by using four conditions: one in which most participants would err, one in which most would respond correctly, and two intermediate conditions. We predicted that (a) the AC relationship within each condition would be modest due to restricted range in memory for the target; (b) both accuracy and confidence would be affected by condition; and (c) the AC relationship across conditions would be substantial.

Method

Participants

Participants were 256 University of Victoria undergraduate students who received extra-credit in an introductory psychology course.

Materials

Three videos were made of each of two targets (Julie and Wayne): The 10-s tape showed brief close-up and full-body views of the target, and the 1- and 3-minute tapes showed the target from a variety of perspectives, in a variety of settings, etc. Each line-up consisted of head-and-shoulder frontal color photographs of eight individuals who resembled one another, and the words "Not Present." For each target, four line-ups were prepared: Two target-absent (TA) and two target-present (TP) (the two versions of each line-up type differed in terms of the positions of the photographs). The two kinds of materials were made at different times and the targets were wearing different clothing in the videos versus line-ups (as would be typical of real-world cases). Confidence was rated on a scale from 1 ("purely guessing") to 11 ("100% confident") with the odd-numbered response options labeled "slightly confident," "fairly confident," "quite confident," and "very confident."

Procedure

Informed consent was obtained from participants, and their rights were duly protected. A predetermined schedule evenly counterbalanced assignment to condition, assignment of targets (Wayne vs. Julie) to the TA versus TP line-ups, order in which the TA and TP tests were performed, and which version of each line-up (which differed in positions of the photographs) was used. Participants were tested individually in one of four conditions:

Worst condition. Participants viewed the two 10-s videos with instructions to identify the locations in which the videos were filmed. After reporting where they thought each video was filmed, participants performed an unrelated (Stroop) experiment for approximately 15 mins., after which they were shown a line-up and asked if the person in the first video was or was not present and, if so, to point to that person's photograph (if not, they were to select the "not present" option). Participants then rated their confidence in this judgment, then performed the identification task and confidence rating for the person in the second video.

Medium condition. Same as the "worst" condition except that participants were told to study the person's appearance in each video because they would later be asked to identify them.

Good condition. Same as the "medium" condition except that the 1-m videos were used and the tests were administered immediately after the two videos.

Best condition. Same as the "good" condition except that the 3-m videos were used.

Results

The data can be broken down for analysis in a variety of ways. For example, separate analyses could be conducted on TA versus TP line-ups, Test 1 versus 2, or choosers (those who identified a photograph) versus non-choosers (those who responded "Not Present"). We have analyzed the data in each of these ways; the absolute values of the results vary, but the pattern is similar. Here we focus (as planned) on analyses of choosers collapsed across TA and TP line-ups. These analyses provide the strongest parallel to real-world cases, in which it is usually not known if the perpetrator is in the line-up and interest focuses on positive identifications. The complete data set is available from the first author.

Table 1 summarizes the data for each target. As expected, accuracy increased across conditions (Julie: $\chi^2(3, N = 121) = 37.68$, $\underline{p} < .001$; Wayne: $\chi^2(3, N = 143) = 64.67$, $\underline{p} < .001$). One-way analyses of variance indicated that confidence also increased reliably across condition (Julie: $\underline{F}(3, 117) = 15.70$, $\underline{MSe} = 4.66$, $\underline{p} < .001$; Wayne: $\underline{F}(3, 139) = 41.17$, $\underline{MSe} = 4.28$, $\underline{p} < .001$). Condition affected confidence in both accurate (Julie: $\underline{F}(3, 57) = 3.45$, $\underline{MSe} = 4.10$, $\underline{p} < .03$; Wayne: $\underline{F}(3, 68) = 3.77$, $\underline{MSe} = 2.91$, $\underline{p} < .02$) and inaccurate identifications (Julie: $\underline{F}(3, 56) = 2.67$, $\underline{MSe} = 4.59$, $\underline{p} < .06$; Wayne: $\underline{F}(3, 67) = 8.31$, $\underline{MSe} = 4.15$, $\underline{p} < .001$).

For the Wayne target the AC correlation was significant within each condition, whereas for the Julie target the AC correlation was significant only in the Medium condition. The weighted mean of the four within-condition AC correlations for Wayne was $r = .44$, whereas that for Julie was $r = .30$ (for the difference, $z = 1.92$, $p < .03$). The match between appearance in the video and line-up seemed to be poorer for Julie than Wayne, which may have lowered confidence and weakened the AC relationship for Julie (cf. Read, Vokey, & Hammersley, 1990). The AC correlation did not reliably differ across conditions ($F < 1$ for both Julie and Wayne).

Of central interest is the AC correlation when data are collapsed across conditions. The AC correlation for Wayne collapsed across conditions was $r = .68$, which is significantly greater than the mean of the within-condition AC correlations for Wayne ($z = 2.75$, $p < .003$). Similarly, the AC correlation for Julie collapsed across conditions was $r = .51$, which is significantly greater than the mean of the within-condition AC correlations for Julie ($z = 4.22$, $p < .001$). Collapsing across the two targets, the weighted mean of the within-condition AC correlations was .36 (confidence accounted for 13% of the variation in accuracy), whereas the AC correlation across conditions was .59 (confidence accounted for 35% percent of the variation in accuracy). Figure 1 depicts frequency histograms of choosers' confidence ratings for accurate and inaccurate identifications for each condition. Figure 2 depicts those data collapsed across conditions.

Discussion

As predicted, manipulations that affected accuracy also affected confidence in the same direction. This finding indicates that confidence ratings are not entirely relative to the conditions under which they are made: If they were, participants would use the scale to indicate their "confidence given the conditions," and therefore confidence would not differ across conditions. Interestingly, the conditions that led to higher confidence in correct identifications also led to higher confidence in false identifications. This suggests a role of metamemory in determining the confidence with which identifications are made--that is, participants tested under conditions likely to lead to good performance tended to be confident even when responding incorrectly (cf. Lüüs & Wells, 1994; Read, 1995).

The central finding of our experiment was that, as predicted, when the data were collapsed across conditions the AC correlation was substantial. Why has the role of variability in witnessing conditions been ignored in previous considerations of the AC relationship? Research by R. C. L. Lindsay, Wells, and Rumpel (1981) may have discouraged exploration of this issue. These investigators manipulated conditions to produce three levels of identification accuracy. The AC relationship was weak in each condition, and confidence did not differ across conditions.

Why did R. C. L. Lindsay et al. (1981) not find an effect of condition on confidence? First, their conditions produced a narrower range of accuracy levels (33%, 50%, and 70%) than the four conditions of the current experiment (11%, 44%, 78%, and 86%). Second, their comparisons involved only 42, 26, and 24 participants in the three conditions, respectively (compared to 83, 64, 58, and 59 in our four conditions). Third, R. C. L. Lindsay et al. used a 7-point confidence scale, whereas we used an 11-point confidence scale. Finally, Wells (personal communication, October 1997) proposed that the fact that the R. C. L. Lindsay et al. participants knew they would later be cross examined about their identification judgments may have restricted differences across conditions in choosers' confidence.

The AC relationship we observed when data were collapsed across conditions ($r = .59$) is substantial, but it is not huge. Indeed, stronger AC relationships have been obtained in some studies that did not seek to create great variability in participants' ability to identify the target (Sporer et al., 1995). It may be that such studies serendipitously created substantial variability across participants in ability to identify the target. Consistent with this, Read et al. (1997) found that the AC correlation among choosers who performed an expected identification after a 9-month delay was $r = .72$ (compared to $r = .18$ after a shorter delay); presumably the long delay enabled individual differences to create large variability across participants in ability to identify the target. Conditions designed to produce "optimal" performance (as per the Bothwell et al., 1987, "optimality hypothesis") may also serendipitously increase variability in memory for the target (e.g., long exposure durations may increase individual differences in memory for the target). Similarly, differences in the

[2] Collapsing across targets is technically inappropriate, because some participants were choosers on both the TA and TP line-ups and hence contributed two pairs of data points. However, the fact that the AC correlations obtained when data are collapsed across targets were very close to the mean of the AC correlations for the Julie and Wayne targets indicates that this violation of the independence assumption did not grossly distort the results.

amount of variability across participants in memory for the target may explain why the AC relationship tends to be stronger in field studies than in laboratory studies.

How strong is the AC relationship across real-world eyewitness suspect identifications? On the one hand, we suspect that the current study underestimates the amount of real-world variability across witnesses in ability to identify the perpetrator. Across real-world cases, some witnesses have only a fleeting incidental glimpse of the perpetrator whereas others spend considerable time in intense face-to-face interactions; some witnesses make an identification within an hour of witnessing the culprit and others only after weeks; some perpetrators have strikingly memorable faces and others do not; in some cases the appearance of the perpetrator as witnessed and in the line-up are very similar and in others they are very dissimilar; etc. These considerations suggest that ability to identify the perpetrator varies greatly from one real-world witness to another and this, in turn, suggests that the real-world AC relationship is strong.

On the other hand, other factors (e.g., suggestive interviewing procedures before or after an identification is made) can dramatically weaken the AC relationship (e.g., Lüüs & Wells, 1994; Wells & Bradfield, in press), and our experiment likely underestimates the extent to which such factors weaken the AC relationship in real-world cases (especially when confidence is assessed at trial rather than immediately after an identification is made). Future research using procedures like ours to simulate real-world levels of variability in memory for the target could assess the extent to which such factors compromise the AC relationship in real-world cases.

Our findings converge with most previous research in indicating that when conditions of witnessing and testing are held constant (e.g., if several people witnessed and were tested on the same perpetrator under the same conditions) confidence is a poor indicator of accuracy. Even when conditions of witnessing and testing vary widely across witnesses, the AC relationship is far from perfect; in our study, some participants were very confident in false identifications and others were very non-confident in accurate identifications, and even when data were collapsed across conditions confidence accounted for only 35% of the variability in accuracy. Finally, our data do not challenge evidence that jurors are overly swayed by the apparent confidence with which identifications are made (e.g., Lieppe, Manion, & Romanczyk, 1992; Penrod & Cutler, 1995; Wells, 1993) (although it would be interesting to repeat studies of jurors with procedures like ours, in which some witnesses have good reasons for being more confident than others).

Our findings do indicate that it is misleading to claim that, in general, confident eyewitness identifications are little more likely to be accurate than non-confident ones. Consistent with commonsense intuition, when witnessing conditions vary widely confident witnesses are substantially more likely to be accurate than non-confident ones.

References

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.

Leippe, M. R., Manion, A. P., & Romanczyk, A. (1992). Eyewitness persuasion: How and how well do fact finders judge the accuracy of adults' and children's memory reports? Journal of Personality and Social Psychology, 63, 181-107.

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

Lüüs, C. A., & Wells, G. L. (1994). The malleability of eyewitness confidence: Co-witness and perseverance effects. Journal of Applied Psychology, 79, 714-724.

Penrod, S., & Cutler, B. (1995). Witness confidence and witness accuracy: Assessing their forensic relationship. Psychology, Public Policy, and Law, 1, 817-845.

Read, J. D. (1995). The availability heuristic in person identification: The sometimes misleading consequences of enhanced contextual information. Applied Cognitive Psychology, 9, 91-122.

Read, J. D., Lindsay, D. S., & Nicholls, T. (1997). The relationship between accuracy and confidence in eyewitness identification studies: Is the conclusion changing? In C. P. Thompson, D. Bruce, J. D. Read, D. Hermann, D. Payne, & M. P. Toglia (Eds.), Eyewitness memory: Theoretical and applied perspectives (pp. 107-130). Mahwah, NJ: Erlbaum.

Read, J. D., Vokey, J. R., & Hammersley, M. (1990). Changing photos of faces: Effects of exposure duration and photo similarity on recognition and the accuracy--confidence relationship. Journal of Experimental Psychology: Learning, Memory, & Cognition, 16, 870-882.

Sporer, S. L., Penrod, S. L., Read, J. D., & Cutler, B. (1995). Choosing, confidence, and accuracy: A meta-analysis of the confidence-accuracy relation in eyewitness identification studies. Psychological Bulletin, 118, 315-327.

Wells, G. L. (1993). What do we know about eyewitness identifications? American Psychologist, 48, 553-571.

Wells, G. L., & Bradfield, A. L. (in press). "Good, you identified the suspect:" Feedback to eyewitnesses distorts their reports of the witnessing experience. J of Applied Psychology.

Table 1

Number of Choosers, Percent Accurate Identifications, Mean Confidence Ratings, and Accuracy-Confidence Correlation Coefficients for Choosers by and across Conditions for each Target

Target Condition	N	Percent Accurate	Confidence: All	Confidence: Accurate	Confidence: Inaccurate	A*C Correlation
Julie						
Worst	37	14%	4.38^{a} (2.06)	5.60^{a} (2.30)	4.19^{a} (1.99)	.24
Medium	30	47%	6.23^{b} (2.37)	7.29^{ab} (1.86)	5.31^{ab} (2.44)	$.42^{*}$
Good	27	70%	6.48^{b} (2.16)	6.84^{a} (2.01)	5.63^{ab} (2.39)	.26
Best	27	85%	8.07^{c} (2.04)	8.30^{b} (2.08)	6.75^{b} (1.26)	.28
All	121	50%	6.13 (2.52)	7.39 (2.14)	4.85 (2.23)	$.51^{***}$
Wayne						
Worst	46	9%	4.67^{a} (2.20)	7.00^{a} (2.94)	4.45^{a} (2.03)	$.33^{*}$
Medium	34	41%	7.88^{b} (2.24)	9.29^{b} (1.86)	6.90^{b} (1.97)	$.53^{**}$
Good	31	84%	8.45^{b} (2.08)	8.89^{b} (1.75)	6.20^{ab} (2.39)	$.48^{**}$
Best	32	88%	9.53^{c} (1.63)	9.82^{c} (1.36)	7.50^{b} (2.08)	$.48^{**}$
All	143	50%	7.34 (2.81)	9.22 (1.80)	5.44 (2.33)	$.68^{***}$

Note: Numbers in parentheses are standard deviations. For the AC correlation coefficients, * = $p < .05$, ** = $p < .01$, and *** = $p < .001$. For the confidence ratings, cell means in a column with different superscripts differ at the .05 level by Fisher's least significant difference test.

Figure 1: Choosers' confidence in correct and incorrect identification judgments by witnessing condition.

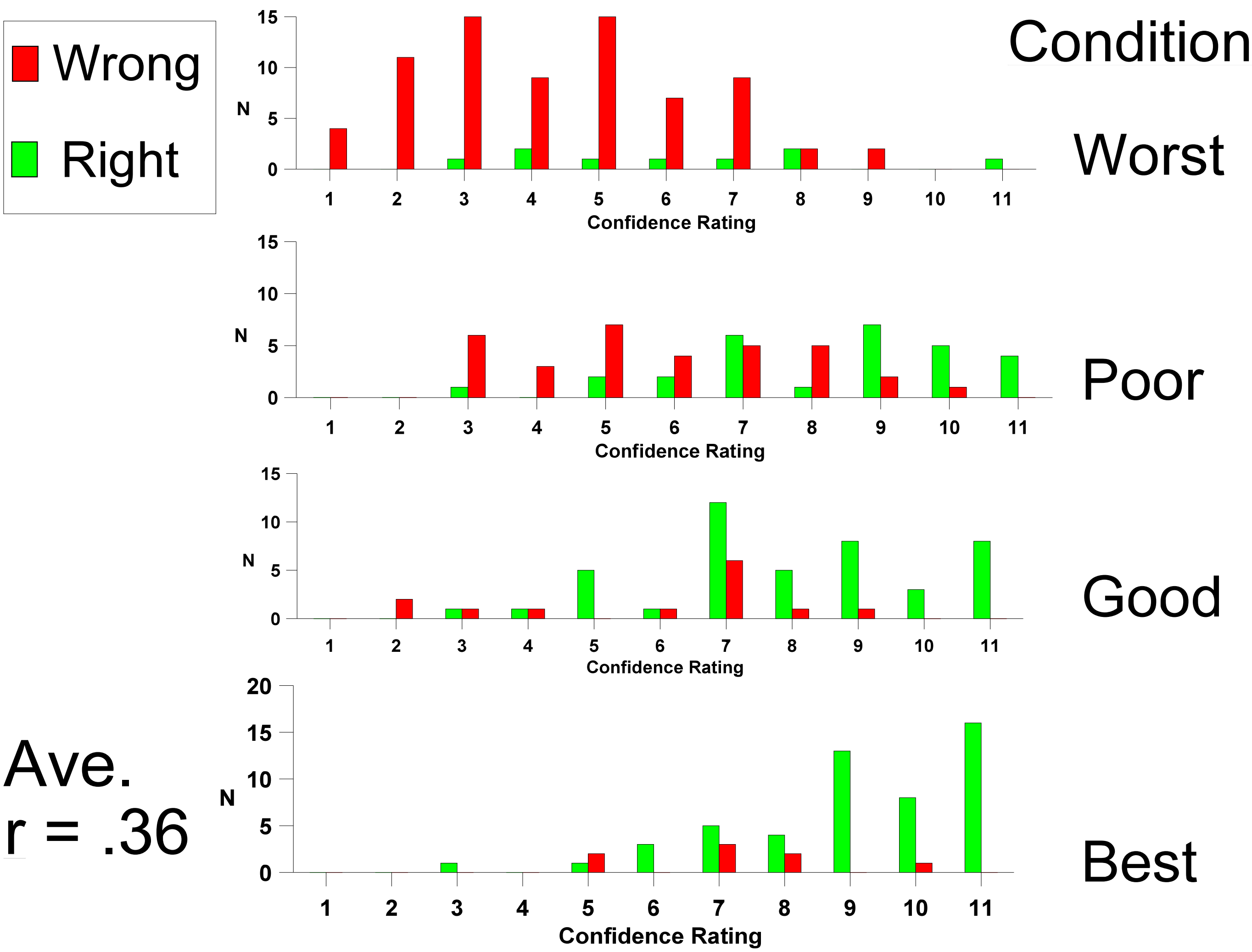

Figure 2: Choosers' confidence in correct and incorrect identification judgments collapsed across witnessing conditions.