Bulletin of the Psychonomic Society
1985, 23 (4), 311-313

Confidence at recognizing psychological implications

LEILANI GREENING and STEPHEN J. DOLLINGER
Southern Illinois University, Carbondale, Illinois

This study assessed the correspondence between judgment and confidence in those judgments on a task requiring subjects "to read between the lines" and recognize the implications of target persons' word associations. College students attempted to detect which of three "suspects" had imagined committing a theft, on the basis of the "suspects'" responses to a word-association test. In addition to the standard preparatory information given to no-clues subjects, half of the subjects received a list of clues that were diagnostic or contradiagnostic of thief status. Successful judgment was not related to self-evaluated performance or to confidence. However, with aggregations across subjects, a significant correlation was found between group hit rate and mean confidence across the 10 items of the judgment task. This relationship held true for subjects given the standard information and for subjects provided with additional information about valid clues. Finally, calibration curves revealed that subjects tended to be overconfident of their judgments whenever they estimated their success probabilities above 70%. Clue information had a weak effect on hit rate but no effect on confidence.

People are frequently overly confident in their judgments about their knowledge and memory, as well as their judgments about other people (Fischhoff, Slovic & Lichtenstein, 1977; Koriat, Lichtenstein, & Fischhoff, 1980; Lichtenstein & Fischhoff, 1977; Oskamp, 1965). Similarly, people often think they are "better than average" on a number of dimensions (Larwood & Whittaker, 1977; Myers & Ridl, 1979; Svenson, 1981), including their psychological-mindedness (Dollinger, Tylenda, Reader, & Marnett, 1984). Drawing on these findings, the purpose of this study was to investigate people's confidence in their judgments on a task involving recognition of the psychological implications conveyed by target persons' word associations (Dollinger, Reader, Marnett, & Tylenda, 1983).

Our research with the task is grounded in the theoretical categories described by Scheibe (1978, 1979) in his discussion of the "psychologist's advantages" in strategic situations: the possibilities of predicting the other by means of sagacity, authority, or acumen. In brief, sagacity involves prediction based on knowledge of subtle cues or stimulus-response connections and alert attention to such cues. Acumen depends on the psychological judge's ability to penetrate the other's defensive "masks" to accurately perceive his or her intentions, plans, emotions, etc; hence, it involves a kind of empathic understanding of the other's perspective. Authority affords prediction via controlling the reinforcement contingencies in a situation; hence, it is less relevant to the task presently used. A series of studies suggest that the acumen mode is irrelevant to success at the task, but sagacity is quite relevant (Dollinger & Riger, 1984; Dollinger, in press). In particular, this research has shown that word associations but not affect expressions or reaction times provide the critical clue for judgment in this task, and successful judges more commonly report using *just* word associations. They also have a better appreciation of which clues are valid, and they are more attentive to those cues. Given this role of sagacity in the task, a reasonable question to ask is whether subjects who are successful are more confident in their judgments. That is, does confidence correlate with success (which is based on knowing what to look for)?

The study included an experimental manipulation involving the presence or absence of clue information in the task preparation. Although this was intended to replicate a study showing its effect on successful judgment (Dollinger, in press, Study 4), we considered it potentially informative in terms of the confidence-accuracy correspondence. Conceivably, such information could not only improve judgment but also simultaneously enhance subjects' confidence in their judgments, thereby improving the confidence-accuracy correlation.

METHOD

Judgment Task

Patterned after classic studies on word association (Henke & Eddy, 1909; Jung, 1910; Yerkes & Berry, 1909) and on an experimental psychology laboratory exercise used for many decades, this task requires subjects to judge, on the basis of word associations, which of three videotaped "suspects" has imagined the commission of a theft. The word-association test (WAT) is a 20-word single-response free-association procedure that includes 6 stimulus words considered relevant to the "thief" script. Prior to the videotaping, suspects were assigned randomly to receive one of five story scripts. They were instructed to read and vividly imagine enacting the events described in their script. Within each set of three suspects (triad), one imagined waiting for a companion while seated in a pizza parlor, finding $20 in a wallet dropped nearby by another patron, denying knowledge of it when asked by the patron,

We wish to thank Gordon Pitz for his comments on the manuscript.
Address reprint requests to Stephen J. Dollinger, Department of Psychology, Southern Illinois University, Carbondale, IL 62901.

and then being plagued by the thought of being a thief. In half of the triads, the foil suspects read scripts in which they (1) had just learned that a final exam was optional or (2) were feeling so unpressured and relaxed that they were falling asleep while studying in a library. For the remaining triads, foil suspects imagined (1) witnessing someone steal a book in a bookstore or (2) preparing to respond to a professor's (false) accusation of plagiarism.

After reading their scripts several times, the suspects were escorted individually to another room for a videotaped "psychological examination," consisting of the WAT. The stimulus list included the words *green*, *class*, *steal*, *sing*, *dead*, *student*, *pizza*, *happy*, *thief*, *window*, *friendly*, *money*, *cook*, *grade*, *cold*, *wallet*, *dance*, *village*, *lake*, and *guilty*.

On the videotape, the suspects were depicted seated near the interviewer to establish the direction of visual gaze when eye contact was made. A close-up of the suspect's face and shoulders was shown after the third stimulus word. Order of testing the suspects was randomized within triads. Further details on the task are presented elsewhere (Dollinger et al., 1983; Dollinger & Riger, 1984; Dollinger, in press).

Procedure

In a preliminary meeting, the subjects completed several self-report personality measures (which will not be discussed in detail here because they are irrelevant to the present purposes), they received one of two preparation sheets with instructions concerning the judgment task, and they made an appointment for the subsequent session to take the judgment task. Both preparation sheets described the nature of the task, the WAT list, the "thief" script, and the suggestion to imagine one's own first association to the words prior to and again after reading the thief script. This comprised the standard preparation used in related studies (Dollinger, in press). In addition, one form of the preparation identified 12 valid word clues that significantly discriminated thief and foil suspects on the videotape. For example, the clue most indicative of thief status, PIZZA:PARLOR, was given by 6 of the 10 thieves but by none of the 20 foil suspects. The preparation sheets were presented in sealed form; a statement asking the subject to read the preparation sheet several times before the next session appeared on the face of the packet.

The subjects were scheduled for the judgment task in groups of about 5 (range = 1-7, mean = 4.3). The subjects viewed each triad in sequence and circled the number of the suspect (Suspect 1, 2, or 3) whom they judged to be the thief. The subjects also rated their confidence in each judgment on a scale of 33% to 100%. As explained to the subjects, at 33% the scale reflected the probability of being correct by chance, for example, without even viewing the videotape. A middle range rating of 67% was stated as indicating that one believed he/she would correctly identify two thieves for every one missed, given similar levels of confidence. At the other extreme, 100% confident indicated the belief that there was no possibility of error in judgment. The videotape stopped after each triad and resumed when all subjects had completed their judgments and confidence ratings.

After completing the task, the subjects evaluated their performance by estimating how many thieves they had detected, provided other ratings about the task and about their performance, and then were debriefed.

Subjects

The subjects were 90 college students (34 females, 56 males) who participated in the study in exchange for credit toward their introductory psychology course grade. The participants ranged in age from 18 to 47 years (mean = 20 years). Three subjects who failed to follow instructions were dropped from analyses. Because the confidence measure was included after three groups of subjects had completed the procedure, ns on the confidence variables were 41 and 31 for the clues and no-clues conditions, respectively. Because sex differences were not found on any of the measures, this variable is not considered further.

RESULTS AND DISCUSSION

Preliminary Results

Experimental condition (clues/no clues) had essentially no effect on confidence ratings (all ps > .10) and a weak effect on hit rate ($p < .06$). Consistent with a previous study (Dollinger, in press), the subjects provided with clues attained slightly more hits than those not provided with clues.

Main Results

Table 1 presents correlations between confidence and accuracy on individual triads. Accuracy here was coded 1 for hits, 0 for misses; hence, entries are point-biserial correlations. As shown in Table 1, confidence correlated significantly with hits on three triads (Triads 1, 4, and 5) for the no-clues condition and on one triad (Triad 5) for the clues condition (all ps < .01). In addition, a negative correlation emerged for the clues condition on Triad 4, albeit of marginal significance ($p < .06$). This correlation was also significantly different from the comparable value for the no-clues condition ($z = 3.51$, $p < .001$). Thus, at the level of individual responses, accurate and inaccurate judges were equally confident in most of their judgments. For subjects given either preparation, a positive correlation emerged for a moderately easy triad (Triad 5). Among no-clues subjects, a positive correlation was found for a very easy triad (Triad 1). And the two conditions differed in the direction of correlation on a moderately difficult triad (Triad 4). It may be that the added information gave these subjects a better appreciation of the multiple cues present, particularly for a difficult triad.

Each subject's hit rate and mean confidence were computed across the 10 triads, and these variables were found to be unrelated in both the clues condition ($r = .10$) and the no-clues condition ($r = -.14$). With estimated number of hits as an alternative measure of confidence, again no significant relationship was found (rs = .08 for the clues condition and .11 for the no-clues condition). (The latter correlations are based on respective ns of 46 and 44.) Thus, total test performance across the 10 judgments did not correlate with confidence or estimated performance.

A different impression was suggested by considering whether subjects as a group are less confident on the more difficult than on the easier triads. For this level of analy-

Table 1
Mean Confidence, Triad Difficulty, and Confidence × Accuracy Correlations for the No-Clues (NC) and Clues (C) Conditions

	Confidence (%)		Difficulty		Correlation	
Triad	NC	C	NC	C	NC	C
1	75	79	.86	.96	.53*	.25
2	62	65	.80	.89	−.06	.04
3	69	66	.68	.74	−.01	.22
4	49	52	.39	.39	.51*	−.30
5	64	73	.84	.76	.50*	.49*
6	51	55	.50	.61	.15	.22
7	55	61	.32	.26	−.28	−.16
8	56	62	.66	.70	.08	.24
9	56	61	.16	.28	−.14	−.10
10	65	61	.43	.61	.08	.18

Note—Triad difficulty is the proportion of subjects passing the item (i.e., "catching the thief"). Correlations are between rated confidence in a judgment and accuracy (1 = correct, 0 = incorrect). *$p < .01$.

sis, the unit of observation was the triad; the two variables were the groups' mean confidence on each triad and the proportion of judges within that group who made a correct decision (i.e., triad difficulty). Here, we found significant rank-order correlations of .82 for the clues condition ($p < .01$) and .66 for the no-clues condition ($p < .05$). These correlations imply that subjects as a group have some appreciation of the ease and difficulty of the various judgments. The different results for the across-subjects and across-triads levels of analysis implies the possibility that confidence has little validity in an absolute sense for a given judgment (e.g., "this decision is difficult"). However, confidence may have more validity in a relative sense (e.g., "this decision is easier than the last one"). Additionally, the "feeling of confidence" in one's confidence judgments may derive from the latter. The results may also be explained by the increased reliability of both accuracy and confidence when these variables are based on composites across 30-40 subjects within a group.

The correspondence observed in rank-order correlations is also evident when data are tabled to reflect calibration curves—hit rates for each level of confidence. Confidence ratings were combined to form eight categories from very low confidence (33%-39%) to certainty (100%). As depicted in Table 2, responses were pooled across subjects (within each group) and across triads. Inspection of Table 2 reveals three trends. First, with increasing confidence, hit rates generally rose. Second, the subjects in the no-clues condition were fairly well calibrated below 70% confidence, but not above that point. In other words, ratings above 70% tend to reflect overconfidence by these subjects. With two important exceptions, this held true for the subjects in the clues group. First, these subjects were more successful than they realized when they assigned ratings below 50%—that is, here they were *under*confident. And second, they were well calibrated when they were certain: For the six occasions in which clues subjects were certain of their judgments, they were always correct. However, given the instructions, the subjects in both groups may have been reluctant to state absolute certainty, and therefore the small number of cases makes the latter impression very tentative.

In sum, this study showed that people's confidence in their inferences about others based on verbal implications ("reading between the word associations") generally has little relation to the accuracy of those individual inferences. However, aggregating across many subjects affords better predictions: That is, the group's mean confidence is a fair representation of the difficulty of the decision. Future studies of the confidence-accuracy relationship in this task might take several directions. First, it would be interesting to determine whether different types of judges (e.g., accurate-underconfident, inaccurate-overconfident) show qualitative or quantitative differences in cue utilization. This could be done by examining which word associations they identify as clues in a paper-and-pencil version of the task. Second, the effects of different response formats might be studied, such as confidence in a given judgment relative to the last several judgments, or confidence that a given suspect is the thief relative to others in the triad.

Table 2
Hit Rate for Each Level of Confidence

	No Clues		Clues	
Confidence	Ratio	Percent	Ratio	Percent
33-39	18/47	38	30/63	48
40-49	27/56	48	29/48	60
50-59	20/48	42	26/50	52
60-69	35/55	64	48/79	61
70-79	22/33	67	36/58	62
80-89	17/27	63	32/44	73
90-99	24/36	67	47/62	76
100	6/ 8	75	6/ 6	100

Note—Ratio signifies the ratio of hits/(hits + misses) for a given level of confidence.

REFERENCES

DOLLINGER, S. J. (in press). Sagacious judgment via word association. *Journal of Personality and Social Psychology.*

DOLLINGER, S. J., READER, M. J., MARNETT, J. P., & TYLENDA, B. (1983). Psychological-mindedness, psychological-construing, and the judgment of deception. *Journal of General Psychology*, **108**, 183-191.

DOLLINGER, S. J., & RIGER, A. L. (1984). On penetrating the "mask": The role of sagacity and acumen in a word-association/clinical-judgment task. *Journal of Personality and Social Psychology*, **46**, 145-152.

DOLLINGER, S. J., TYLENDA, B., READER, M. J., & MARNETT, J. P. (1984). Self-ascribed psychological-mindedness: A self-serving bias in "penetrating the mask." *Journal of Social Psychology*, **124**, 127-129.

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

HENKE, F. G., & EDDY, M. W. (1909). Mental diagnosis by the association reaction method. *Psychological Review*, **16**, 399-409.

JUNG, C. G. (1910). The association method. *American Journal of Psychology*, **21**, 219-269.

KORIAT, A., LICHTENSTEIN, S., & FISCHHOFF, B. (1980). Reasons for confidence. *Journal of Experimental Psychology: Human Learning and Memory*, **6**, 107-118.

LARWOOD, L., & WHITTAKER, W. (1977). Managerial myopia: Self-serving biases in organizational planning. *Journal of Applied Psychology*, **62**, 194-198.

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

MYERS, D. G., & RIDL, J. (1979, August). Can we all be better than average? *Psychology Today*, pp. 89, 95-98.

OSKAMP, S. (1965). Overconfidence in case-study judgments. *Journal of Consulting Psychology*, **29**, 261-265.

SCHEIBE, K. E. (1978). The psychologist's advantage and its nullification: Limits of human predictability. *American Psychologist*, **33**, 869-881.

SCHEIBE, K. E. (1979). *Mirrors, masks, lies, and secrets*. New York: Praeger.

SVENSON, O. (1981). Are we less risky and more skillful than our fellow drivers? *Acta Psychologica*, **47**, 143-148.

YERKES, R. M., & BERRY, C. S. (1909). The association reaction method of mental diagnosis (Tatbestandsdiagnostik). *American Journal of Psychology*, **20**, 22-37.

(Manuscript received for publication February 20, 1985.)