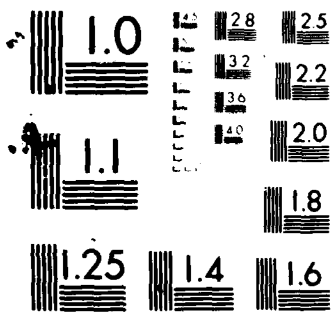
AD-A109 730 DECISION RESEARCH EUGENE OR F/G 5/10
SUBJECTIVE CONFIDENCE IN FORECASTS.(U)
DEC 81 B FISCHHOFF, D MACGREGOR N00014-80-C-0150
UNCLASSIFIED PTR-1092-81-12 NL

1 of 1

END
DATE
FILMED
2 82
DTIC



MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963-A

Technical Report PTR-1092-81-12
Contract No. N00014-80-C-0150
December 1981

SUBJECTIVE CONFIDENCE IN FORECASTS

BARUCH FISCHHOFF
DON MacGREGOR

DECISION RESEARCH
A BRANCH OF PERCEPTRONICS

Prepared for:
OFFICE OF NAVAL RESEARCH
800 North Quincy Street
Arlington, VA 22217

PERCEPTRONICS

6271 VARIEL AVENUE • WOODLAND HILLS • CALIFORNIA 91367 • PHONE (213) 884-7470

unclassified

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER	2. GOVT ACCESSION NO. AD-A109 730	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) Subjective Confidence in Forecasts		5. TYPE OF REPORT & PERIOD COVERED Technical Report
		6. PERFORMING ORG. REPORT NUMBER PTR-1092-81-12
7. AUTHOR(s) Baruch Fischhoff and Don MacGregor		8. CONTRACT OR GRANT NUMBER(s) N00014-80-C-0150
9. PERFORMING ORGANIZATION NAME AND ADDRESS Decision Research, A Branch of Perceptronics, 1201 Oak Street, Eugene, Oregon 97401		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS
11. CONTROLLING OFFICE NAME AND ADDRESS Office of Naval Research, 800 North Quincy Street, Arlington, Virginia 22217		12. REPORT DATE December 1981
		13. NUMBER OF PAGES 47
14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office)		15. SECURITY CLASS. (of this report) unclassified
		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

approved for public release; distribution unlimited

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Calibration
Overconfidence

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

To be useful, a forecast must not only predict the future, but also give some indication of how much confidence to place in that prediction. The appropriateness of people's confidence in their general knowledge has been studied extensively. After briefly reviewing that literature, the present article attempts to make it more directly relevant to forecasters by repeating previous sutdies in the context of confidence in forecasts. The most robust of previous results was strongly replicated: Participants were greatly —

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

unclassified

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered) overconfident in their predictions when dealing with a fairly hard-to-predict set of events. A procedure that had previously proved effective in reducing overconfidence, forcing respondents to provide reasons why their answers might be wrong, was of minor value here. A new result was the discovery of a simple indicator of the quality of people's confidence assessments, whether they ever expressed certitude. Possible implications of these results for producing and using forecasts are discussed.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

SUMMARY

Forecasts have little value to decision makers unless it is known *how much confidence* to place in them. Those expressions of confidence have, in turn, little value unless forecasters are able to assess the limits of their own knowledge accurately.

Previous research has shown very robust patterns in the judgments of individuals who have not received special training in confidence assessment: Knowledge generally increases as confidence increases. However, it increases too swiftly, with a doubling of confidence being associated with perhaps a 50% increase in knowledge. With all but the easiest of tasks, people tend to be *overconfident regarding how much they know*.

These results have typically been derived from studies of judgments of general knowledge. The present study found that they also pertained to confidence in forecasts. Indeed, the confidence-knowledge curves observed here were strikingly similar to those observed previously. The only deviation was the discovery that a substantial minority of judges never expressed complete confidence in any of their forecasts. These individuals also proved to be better assessors of the extent of their own knowledge.

Apparently *confidence in forecasts* is determined by processes similar to those that determine confidence in general knowledge. Decision makers can use forecasters' assessments in a relative sense, in order to predict when they are more and less likely to be correct. However, they should be hesitant to take confidence assessments literally. Someone is more likely to be right when he or she is "certain" than when he or she is "fairly confident;" but there is no guarantee that the certain forecast will come true.

SUBJECTIVE CONFIDENCE IN FORECASTS

Since the destruction of the Second Temple,
prophecy has become the lot of fools.

\- Hebrew expression

What constitutes a wise forecaster? It is not just someone who is usually correct; that definition would give undue deference to those who make forecasts about predictable events. It is not just someone who is seldom proven wrong; that definition would reward the makers of vague and unverifiable forecasts. It is not just someone who provides a confident message with clear implications for action; that definition would promote arrogance over thoughtfulness.

If one is to take action on the basis of a forecast, perhaps the most desirable property is that it be appropriately qualified. That is, one wants to know how much faith to put in it. One measure of the appropriateness of expressions of faith in forecasts is their degree of calibration.[1] For the sake of calibration, all statements of fact are considered to carry with them an implicit or explicit expression of confidence in their truth. When that expression is given quantitative form, the archetypal statement of fact has the form "the probability that statement A is true is X." Statement A may refer to a discrete event (My bank account is overdrawn.); or a continuous one (The balance in my bank account is between -$100 and $150.). It could refer to the past (George Washington died because of poor medical treatment.); present (The capitol of Saudi Arabia is Mecca.); or future (Quebec will be a part of Canada on January 1, 2000.). Only statements about the future represent forecasts, but the evaluation of all such expressions of confidence is similar. Except for situations in which an individual is 100% confident and wrong, it is hard to validate single expressions. However, one can take a set of statements and see if X% of those assigned an X% chance of being correct prove to be correct, once the truth of the statement can be ascertained. The truth of forecasts can be checked by seeing whether the predicted events occur.

The Bayesian, or subjectivist, view of probability underlying calibration studies assumes that probabilities represent an individual's state of knowledge. Hence, it makes sense to aggregate probabilities over a diverse set of statements and see how well, in general, an individual assesses the extent of his or her knowledge.

Crude retrospective assessments of calibration may be derived from looking at the confidence expressions accompanying the performance of real tasks. Thus, one might find evidence of overconfidence in professions that make confident judgments with no demonstrated validity (e.g., predictions of stock price movements [Dreman, 1979; Slovic, 1972], psychiatric diagnoses of dangerousness [Cocozza & Steadman, 1978]). Unfortunately, such evidence is not only imprecise, but also ambiguous whenever "experts" are consulted (and paid) as a function as the confidence they inspire, suggesting that they may be tempted to misrepresent how much they know (Armstrong, 1978).

Among real-world studies, the greatest efforts to ensure candor and explicitness have been with weather forecasters, who are rewarded for good calibration. Their performance is superb (e.g., Murphy & Winkler, 1974, 1977). Whether this success is due to training in calibration or a by-product of their general professional education is unclear. A review of other studies with experts who have not had calibration training suggests that such training, and not just education, is the effective element. Experiments, using problems drawn from their respective areas of expertise but isolated from real-world pressures, have found overconfidence with psychology graduate students (Lichtenstein, & Fischhoff, 1977), bankers (Stael von Holstein, 1972), clinical psychologists (Oskamp, 1962), executives (Moore, 1977), civil engineers (Hynes, & Vanmarcke, 1976), and untrained professional weather forecasters (Root, 1962; Stael von Holstein, 1971).

Overconfidence is also the predominant result of experiments using non-experts responding to general-knowledge questions (Lichtenstein, Fischhoff, & Phillips, in press). Table 1 provides a summary of studies that have attempted to eradicate overconfidence by a variety of manipulations including changing the response mode, offering detailed instructions, raising the stakes hinging on good calibration, and varying the heterogeneity of the item being judged. Each paper is represented by a number which is underlined if the manipulation seemed to improve calibration. From this large set of studies, only three procedures seem to be effective. One is extensive training with personalized feedback. The second is forcing respondents to list reasons why the statement or answer they believe in might be wront (Koriat, Lichtenstein, & Fischhoff, 1980; Study No. 18 in Table 1). The third, and least interesting, is to provide easier tasks. One reflection of people's insensitivity to how much they know is the fact that their mean confidence changes relatively slowly in response to changes in the difficulty of the tasks they face. Thus, when tasks become easier, people's confidence does not rise commensurately, leaving them under-confident for the easiest of tasks. In this light, the preponderance of overconfidence in the literature reflects, in part, the (perhaps natural) tendency not to present people with very easy questions.

The subjectivist interpretation of probability makes no distinction between confidence in statements about the future and confidence in any other kind of statement. Hence, from a formal perspective, one would expect that the results summarized in Table 1 could be generalized to the calibration of forecast probabilities. That is, one could expect to find overconfidence that is impervious to most of the various manipulations described there. Formal equivalence is not, however, the same as psychological equivalence. One might speculate, for example, that all other things being equal, people are less confident in their knowledge about the future because no one knows about the future. Or one might speculate that they are more confident because no one can prove them wrong at the moment of prediction.

TABLE 1

ATTEMPTS TO EXPLAIN OR REDUCE OVERCONFIDENCE

Strategies		Studied by
Faulty tasks		
Unfair tasks		
	Raise stakes	1,28
	Clarify instructions/stimuli	3,10,12,13,20
	Discourage second guessing	12,20
	Use better response modes	12,13,19,21,22,30,32,33?,<u>34</u>,37?
	Ask fewer questions	15
Misunderstood tasks		
	Demonstrate alternative goal	13
	Demonstrate semantic disagreement	3,13,18,28?
	Demonstrate impossibility of task	12
	Demonstrate overlooked distinction	14?
Faulty judges		
Perfectible individuals		
	Warn of problem	12
	Describe problem	3
	Provide personalized feedback	<u>20</u>
	Train extensively	<u>1</u>,<u>2</u>,<u>4</u>,<u>16</u>,<u>20</u>,<u>24</u>,<u>25</u>,<u>29</u>,<u>32</u>
Incorrigible individuals		
	Replace them	-
	Recalibrate their responses	2,5,23
	Plan on error	-
Mismatch between judges and task		
Restructuring		
	Make knowledge explicit	17
	Search for discrepant information	<u>17</u>
	Decompose problem	-
	Consider alternative situations	-
	Offer alternative formulations	33?
Education		
	Rely on substantive experts	<u>11</u>,15,19,23,27,31,35,36
	Use easier questions	<u>8</u>,<u>9</u>,<u>22</u>,<u>26</u>,<u>29</u>,<u>30</u>
	Educate from childhood	6,7

Note: Each number represents a separate article. Manipulations that have proven at least partially successful are underlined. Those that have yet to be subjected to empirical test or for which the evidence is unclear are marked by a question mark. Details in Fischhoff (in press).

KEY TO TABLE 1

1. Adams & Adams (1958)
2. Adams & Adams (1961)
3. Alpert & Raiffa (in press)
4. Armelius (1979)
5. Becker & Greenberg (1978)
6. Beyth-Marom & Dekel (in press)
7. Cavanaugh & Borkowski (1980)
8. Clarke (1960)
9. Cocozza & Steadman (1978)
10. Dawes (in press)
11. Dowie (1976)
12. Fischhoff & Slovic (1980)
13. Fischhoff, Slovic & Lichtenstein (1977)
14. Howell & Burnett (1978)
15. Hynes & Vanmarcke (1976)
16. King, Zechmeister & Shaughnessy (in press)
17. Koriat, Lichtenstein & Fischhoff (1980)
18. Larson & Reenan (1979)
19. Lichtenstein & Fischhoff (1977)
20. Lichtenstein & Fischhoff (1980)
21. Lichtenstein, Fischhoff & Phillips (in press)
22. Ludke, Stauss & Gustafson (1977)
23. Moore (1977)
24. Murphy & Winkler (1974)
25. Murphy & Winkler (1977)
26. Nickerson & McGoldrick (1965)
27. Oskamp (1962)
28. Phillips & Wright (1977)
29. Pickhardt & Wallace (1974)
30. Pitz (1974)
31. Root (1962)
32. Schaefer & Borcherding (1973)
33. Seaver, von Winterfeldt & Edwards (1978)
34. Selvidge (1980)
35. Staël von Holstein (1971)
36. Staël von Holstein (1972)
37. Tversky & Kahneman (1974)

A study by Fischhoff (1976) found no difference in judgments of the likelihood of hypothetical events set in the future, present, or past. However, the hypotheticality of those events may have weakened some pertinent psychological processes. The studies involving predictions cited above (e.g., Murphy, & Winkler, 1977; Root, 1962) also follow the general patterns observed in non-prediction studies (i.e., overconfidence except with easy tasks or extensive, personalized training). One intriguing possible exception to these patterns is seen in Figure 1, showing a study by Fischhoff and Beyth (1975) in which participants assessed the probability of various possible outcomes of President Nixon's trips to China and the USSR (e.g., he will meet with Chairperson Mao). At the extremes here, one sees the usual overconfidence. About 10% of the events that respondents were 100% certain would happen, failed to happen; about 10% of those that had 0% chance of happening did happen. Nonetheless, over most of the range, subjects were quite well calibrated. An unpublished study by Wright and Wisudha (1979) showed less overconfidence with forecasts than with assessment of general-knowledge questions; fortunately, the forecast questions were also less difficult, suggesting that ease might have been responsible for the difference in calibration.

Reviewing this evidence, anyone interested in eliciting and interpreting expressions of confidence in forecasts or in training forecasters to make such assessments is probably best off assuming that probability assessments for forecasts are no different than those for other problems. The present study attempts to increase or decrease the confidence in forecasts using tasks that are as similar as possible to those used in studies of calibration with general-knowledge questions.

Study 1.

The most widely-used task in calibration studies is the half-range two-alternative question. Given an item with two alternative answers, one of

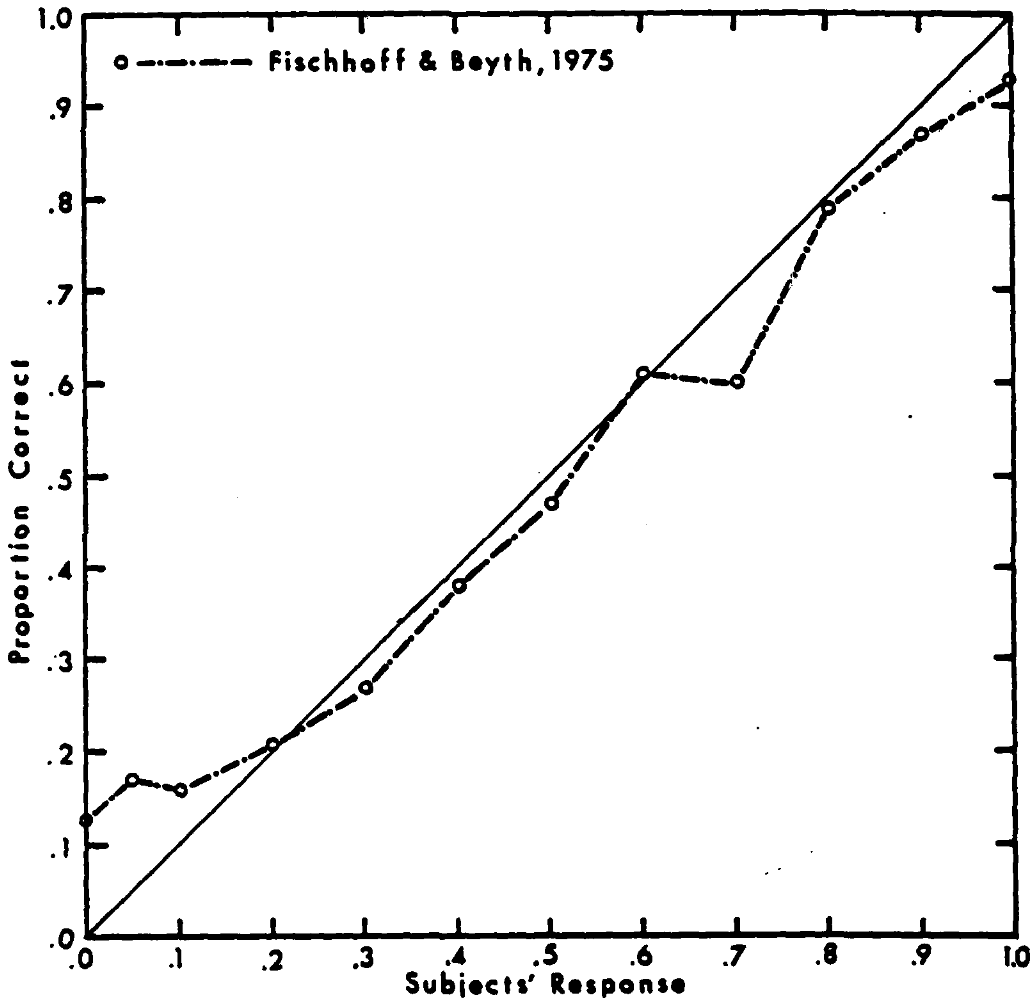


FIGURE 1.
CALIBRATION OF CONFIDENCE JUDGMENTS FOR FORECASTS REGARDING POSSIBLE OUTCOMES OF PRESIDENT NIXON'S TRIPS TO CHINA AND THE USSR.
SOURCE: FISCHHOFF & BEYTH (1975)

which is guaranteed to be true (e.g., absinthe is (a) a precious stone; (b) a liqueur), the respondent must first select the answer that seems more likely to be correct and then assess the probability of that choice being the correct one. Because the more likely answer was to have been chosen, that probability should come from the upper half of the probability range: [.5, 1.0]. Figure 2 shows some typical results observed in studies using such tasks. With all but the easiest tasks, one finds overconfidence, represented by calibration curves resting predominantly under the identity line that would reflect perfect calibration. Being under the identity line means that the percentage of correct answers associated with a particular expressed probability of being correct is smaller than that probability. In such figures, responses are grouped into the intervals [.50, .59], [.60, .69], [.70, .79], [.80, .89], [.90, .99], and [1.00].

The one notable exception to this pattern was the study by Koriat, Lichtenstein, and Fischhoff (1980) in which overconfidence was reduced (although not altogether eliminated) by having respondents provide a reason why each of their chosen answers might be incorrect. Figure 3 shows the effect of this contradicting reason manipulation along with the non-effect of two related manipulations. (In the exhibit, each group's performance on the experimental task is contrasted with its performance on a set of control items for which no reasons were given.) The supporting-reason group provided one reason why their chosen answer might be correct; the both-reasons group gave one supporting and one contradicting reason. The absence of an effect with those groups indicated that the contradicting reason group's calibration had not improved simply as a result of the additional labor involved in writing a reason.

Study 1 replicates the three conditions of Koriat, Lichtenstein, and Fischhoff (1980), using items involving future events.

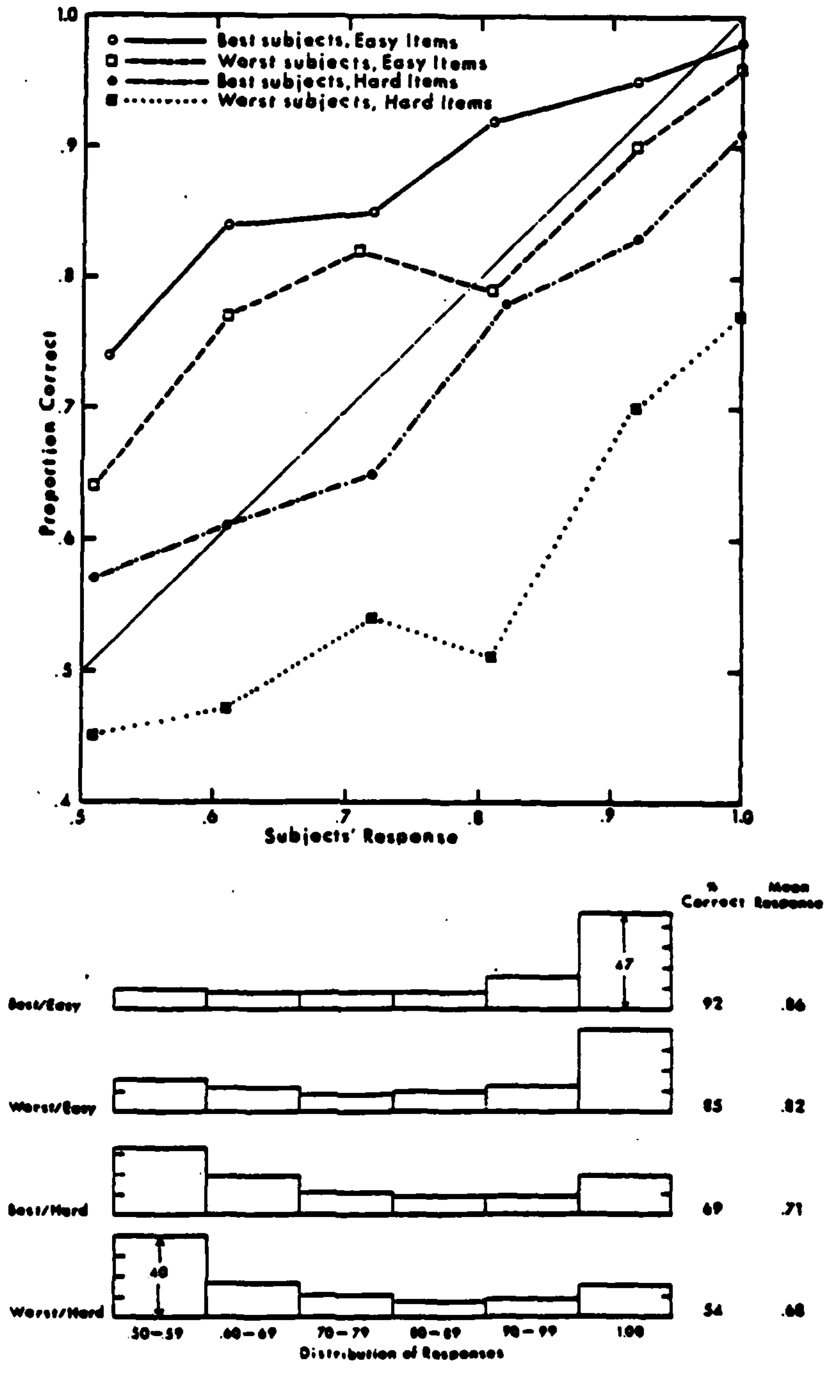


FIGURE 2.
REPRESENTATIVE CALIBRATION CURVES DERIVED FROM STUDIES USING TWO-ALTERNATIVE, HALF-RANGE TASKS.
SOURCE: LICHTENSTEIN & FISCHHOFF (1977).

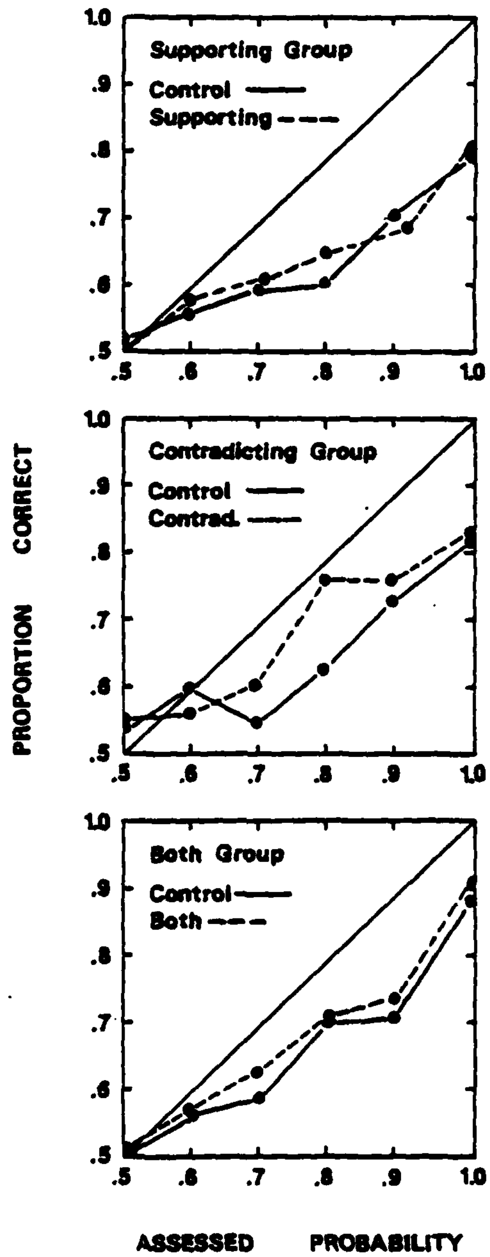


FIGURE 3.
CALIBRATION CURVES FOR INDIVIDUALS PROVIDING SUPPORTING, CONTRADICTING, OR SUPPORTING AND CONTRADICTING REASONS. EACH GROUP'S CALIBRATION IS COMPARED WITH THEIR OWN PERFORMANCE ON A SET OF CONTROL ITEMS.
SOURCE: KORIAT, LICHTENSTEIN & FISCHHOFF (1980)

Method

Design. Each participant responded to 50 two-alternative half-range questions, picking the answer most likely to be correct in each and then assigning it a probability (from .5 to 1.0) of being correct. The first 25 items were done using standard assessment techniques. For each of the last 25 items, after respondents had selected an answer, and prior to providing a probability, they were required to provide a reason supporting their answer, a reason contradicting it, or a reason of each type. Details may be found in Koriat, Lichtenstein, and Fischhoff (1980). A no-reasons group responded to all 50 items without providing reasons.

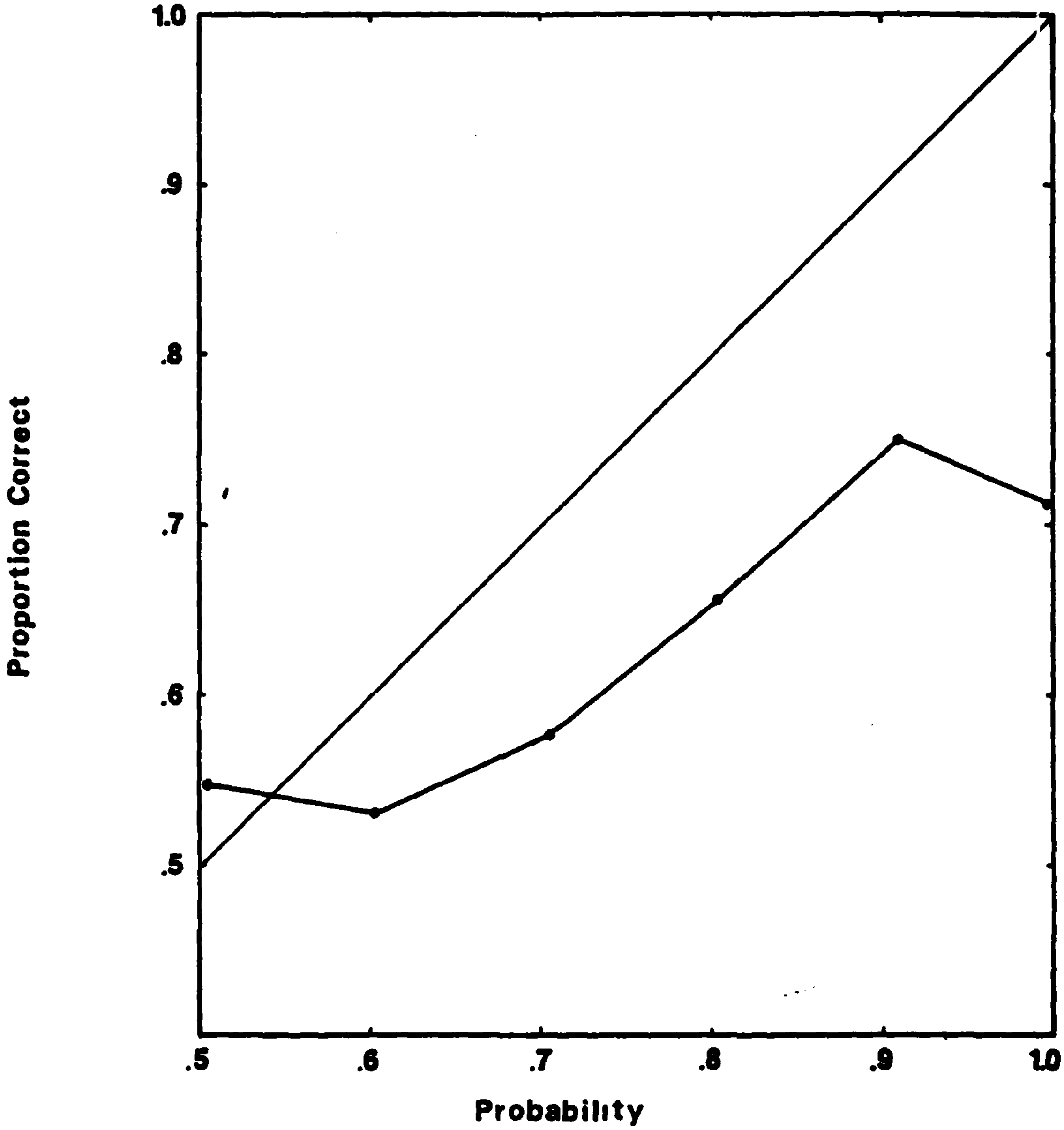
Stimuli. Fifty items were created concerning events that would be consummated within 30 days of the time of the experiments. Some dealt with upcoming local elections (e.g., the mayor of Eugene will be (a) Gus Keller; (b) Catherine Lauris.); others dealt with sporting events (e.g., who will win the following baseball game: (a) Detroit Tigers; (b) California Angels [home team]); others dealt with a variety of topics. These items were separated into two sets so that items dealing with topics that seemed at all related would not appear consecutively or, to the extent possible, in the same set. Each set was used in the control condition for half of one group and in the experimental condition for the other half.

Subjects. One hundred and twelve individuals were recruited through an advertisement in the University of Oregon student paper. They were paid $7 for completing this task as one part of a 1½-hour session. Subject groups recruited in this manner typically are about half male and half female, with an average age of 23. Most are involved with the university community; about 2/3 are students. They treat the tasks in a diligent manner, perhaps akin to a proctored exam. We had hoped to have a larger number of subjects; however, good weather and the proximity of final exams seem to have kept numbers down. In all, there were 32 people in the supporting reasons group, 28 in the contradicting reasons group, 26 in the both-reasons group, and 26 in the group that never gave any reasons.

Results

Main effect. The items we constructed proved to be fairly difficult for subjects, with the proportion of correct forecasts over all 3.447 responses in the control conditions equal to only .618. Associated with these items was a mean confidence of .722. The usual measure of over-or-underconfidence is the difference between these two statistics. Here it equals +.102, indicating that subjects' percentage of correct predictions (in the control condition) should have been higher by 10.2% if their level of confidence was to be justified. The calibration curve corresponding to these responses appears in Figure 4. Respondents' overconfidence is reflected by the fact that most of the curve falls below the identity line. The generally positive slope of the curve indicates that subjects tended to be more knowledgeable when they were more confident. Its flatness, relative to the identity line, indicates that their knowledge did not rise as quickly as did their confidence. This curve pertaining to forecasts looks strikingly like that observed with general knowledge questions of the same difficulty level (e.g., the bottom curves in Figure 2).

Reasons manipulation. Figure 5 contrasts each group of subjects' calibration on the experimental condition with their own performance on the control condition. Thus, it is comparable to Figure 3 from Koriat, Lichtenstein, and Fischhoff (1980). As a rough guide to the stability of these curves, in each, there are approximately 100 (± 30) responses involved in determining the proportions correct associated with probabilities of .6, .7, .8, and .9. If these were all independent responses, that would mean a standard error of estimate of approximately .05; however, subjects typically contributed several responses to each point. Approximately 175 (±30) responses were associated with .5 and about 60 (±30) with 1.0



ALL CONTROL-STUDY 1

FIGURE 4.
CALIBRATION OF ALL RESPONSES TO CONTROL ITEMS IN STUDY 1. CURVE INCLUDES 3,447 RESPONSES PRODUCED BY 112 INDIVIDUALS.

The clearest conclusion to be drawn from Figure 5 is that there are few, if any, systematic differences between the control and experimental conditions for any group. The supporting reasons group, which showed no change at all in the Koriat et al. study, seems to have improved somewhat; however, even these differences seem small relative to statistical variability. The performance of these groups in the control and experimental conditions is summarized several ways in Table 2. Here, we find that the experimental manipulation had little effect on the confidence of supporting or contradicting reasons subjects (slightly increasing it for the former, slightly decreasing it for the latter), but it reduced the mean confidence of both-reasons subjects from .724 to .663. This last change would have cut the overconfidence that those subjects showed in the control condition by 2/3 were there not a concomitant drop in their proportion of correct responses (from .626 to .599). All in all, each of the three groups was somewhat less overconfident in their respective experimental conditions. This modest improvement is also reflected in the group calibration scores shown in Table 2. This score, derived from the partition of the Brier proper scoring rule (see Lichtenstein, Fischhoff, & Phillips, in press), reflects the squared distance between the calibration curve and identity line, weighted by the number of responses at each point. It decreases as calibration improves, becoming zero with perfect calibration.

In Figure 5, each group's performance on the reasons task was compared to their own performance on the control (no reasons) task. Although such within-subject comparisons allow greater sensitivity of analysis, they greatly reduce the number of responses involved in each comparison. If one pools all responses to control questions (as in Figure 4), there appears to be slight improvement in each experimental condition, particularly with the both and contradicting reasons group.

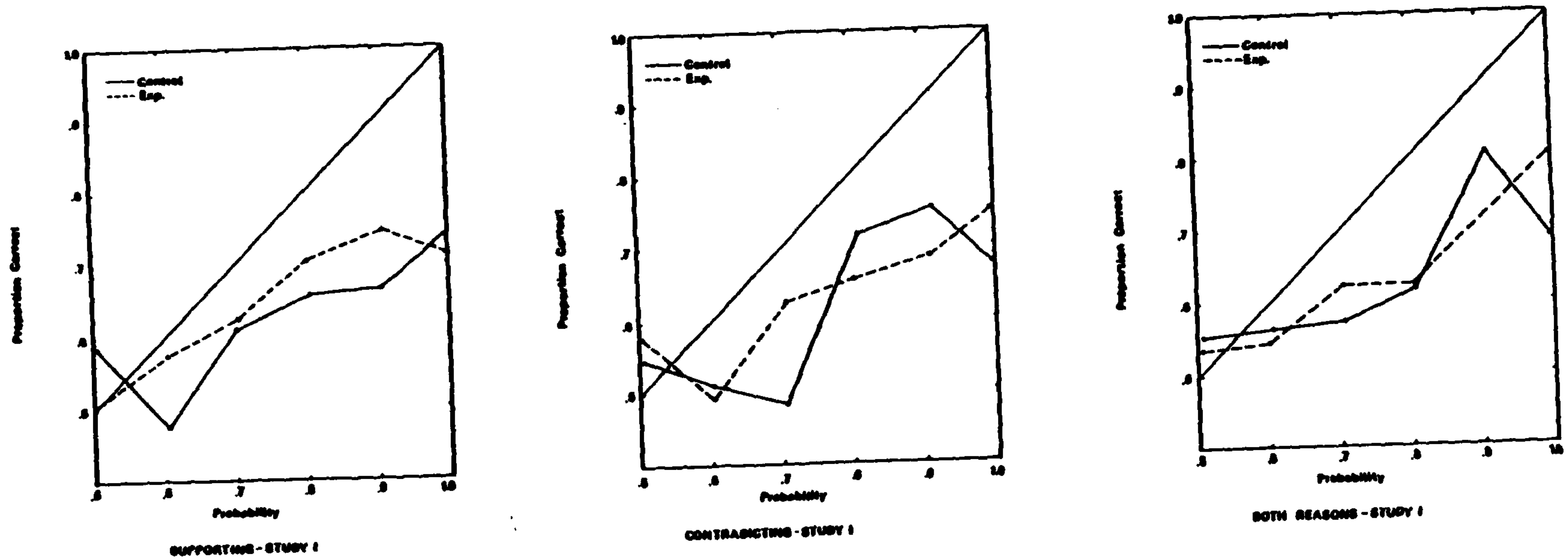


FIGURE 5.
CALIBRATION CURVES FOR INDIVIDUALS PROVIDING SUPPORTING, CONTRADICTING, OR BOTH KINDS OF REASONS IN STUDY 1. CORRESPONDING SUMMARY STATISTICS APPEAR IN EXHIBIT 7.

TABLE 2

SUMMARY STATISTICS FOR STUDY 1

	Control						Experiment				
Group	N	n	Prop. Cor.	Mean Conf.	Over-conf.	Calib.	n	Prop. Cor.	Mean Conf.	Over-conf.	Calib.
No Reasons	26	1299	.625	.724	.099	.0227	—	—	—	—	—
Supporting	32	800	.608	.726	.118	.0264	798	.639	.735	.096	.0166
Contradicting	28	698	.607	.711	.104	.0239	685	.616	.702	.086	.0211
Both	26	650	.626	.724	.098	.0225	643	.599	.663	.064	.0123
All	112	3447	.618	.722	.104	.0271	2126	.619	.703	.083	.0160

The preceding analyses assume that the experimental manipulations were uniformly effective. Koriat et al. discovered a moderate percentage of items for which reasons were either missing or inappropriate. This was particularly true with the contradicting reasons group, who often gave supporting reasons. Each of the present groups omitted reasons for approximately 10% of all items. When supporting reasons subjects gave reasons, they were almost always appropriate to the task (99% of the time). On the other hand, 11% of the contradicting reasons subjects' reasons were inappropriate, constituting either supporting reasons or vague statements such as "Maybe I'm wrong." For both-reasons subjects, 5% of their supporting reasons were inappropriate, compared with 9% of their contradicting reasons. As in Koriat et al., providing contradicting reasons appears to be a difficult or unnatural task. The total number of these missing and inappropriate responses was not large enough that their elimination changes the calibration curves of Figure 5 appreciably.

Distribution of responses. As mentioned earlier, on the control questions, subjects made roughly equal use of all the responses.6, .7, .8, and .9; they used .5 somewhat more, 1.0 somewhat less. Distributions for the experimental conditions were quite similar except for a slight increase in .5's and decrease in 1.0's. This tendency was particularly marked in the both-reasons group, whose proportion of .5's increased from .234 to .364 and whose proportions of 1.0's dropped from .112 to .048, thus accounting for its reduced overall confidence.

Table 3 shows another aspect of response usage, the percentage of subjects who expressed confidence of 1.0 in at least one of their 30 forecasts. In previous studies with general knowledge questions, typically all or almost all subjects have used 1.0. The fact that only 76% of all subjects did so on the control task suggests some tendency not to express extreme certainty in forecasts. This tendency was highlighted in the experimental tasks, where even fewer subjects used 1.0, particularly for the contradicting and

TABLE 3

USAGE AND NON-USAGE OF 1.0

(PERCENTAGE OF USERS)

Study	1		2		3		All	
Group	Control	Exp.	Control	Exp.	Control	Exp.	Control	Exp.
No Reasons	80.8	---	---	---	---	---	80.8	---
Supporting	75.0	65.6	73.1	63.5	81.3	74.7	77.7	69.7
Contradicting	71.4	50.0	67.4	44.2	76.3	60.5	71.6	51.4
Both	76.9	50.0	75.0	50.0	79.5	61.4	77.1	54.2
All	76.8	55.8	72.0	52.4	79.8	68.2	76.2	60.2

both-reasons groups. The responses of all subjects who did and did not use 1.0 were pooled separately over all experimental groups. For the control conditions, non-users were appreciably better calibrated all along the calibration curve (not shown). Subjects who never expressed extreme confidence were not only less confident, but also more in tune with the extent of their knowledge. With the reasons conditions, the same change was observed, but its size was smaller. Users and non-users of 1.0 had highly similar percentages of correct responses, hence differences in calibration cannot be attributed to differences in difficulty level. As can be seen from the remainder of Table 3, similar patterns were observed in the following studies. Calibration curves for these studies will be reported and discussed later.

Discussion

Lost in this morass of mild and inconclusive effects is the striking main effect shown in Figure 4. Calibration for confidence in forecasts looks just like calibration for confidence in general knowledge, when difficulty level is controlled. These forecasters' accuracy increased with their confidence; however, it did not increase as fast. As confidence rose from .5 to 1.0, the corresponding proportion of correct predictions only increased from .5 to .75. Respondents also tended to be overconfident in the extent of their knowledge, getting 62% of their predictions right, but having a mean probability of .72. The one difference that does emerge is a modest reduction in usage of 1.0. The superior calibration of subjects who never used 1.0 was a promising predictor of individual differences in calibration. Despite this overall similarity, confidence in forecasts did not, however, show the same responsiveness to the reasons manipulations observed in Koriat et al. There was some suggestion of improved calibration with the supporting and contradicting groups. However, the relatively small sample rendered these results somewhat ambiguous. Before reaching any firm conclusion, it seemed appropriate to increase the sample size.

Because the events had already occurred by the time these analyses were completed, it was not possible to add subjects to the existing groups. Instead, a second study was run, replicating the first, but with a new set of events.

Study 2.

Method

The design of Study 2 followed that of Study 1 except for the elimination of the no-reasons group (which completed 50 forecasts without giving any reasons) and an increase in the number of forecasts from 50 to 60. As the study was completed in late October, 1980, a number of the forecasts considered the elections of the following month. A total of 143 subjects, recruited as in Study 1, participated. This number, too, was somewhat less than we had hoped for, but did allow for groups roughly 2/3 larger than in Study 1.

Main effect. The difficulty of the present items in the control tasks proved to be remarkably similar to that of Study 1 (62.9% correct vs. 61.8%), as was subjects' mean confidence (.732 vs. .722). Subjects' overconfidence was correspondingly almost identical (.103 vs. .102).

Reasons manipulation. Table 4 summarizes results for the control and experimental conditions of each of the three groups. Briefly, the only apparent effect on overconfidence was the improvement of the both-reasons group. The other two groups were essentially unchanged. The calibration curves for these groups were so similar to those from Study 1 (Figure 5), that they will not be shown. There were, again, qhite a few missing and inappropriate reasons, particularly for contradictory reasons. Elimination of these responses does not, however, appreciably change the patterns shown in Table 4.

TABLE 4

SUMMARY OF STATISTICS - STUDY 2

	Control						Experiment				
Group	N	n	Prop. Cor.	Mean Conf.	Over-conf.	Calib.	n	Prop. Cor.	Mean Conf.	Over-conf.	Calib.
Supporting	52	1552	.635	.735	.100	.0136	1549	.636	.736	.100	.0180
Contradicting	43	1286	.639	.728	.089	.0134	1282	.605	.705	.101	.0152
Both	48	1439	.614	.733	.119	.0226	1424	.637	.706	.070	.0186
All	143	4277	.629	.732	.103	.0159	4255	.627	.717	.090	.0159

Distributions of responses. Presumably reflecting the increased sample size, the distributions of the three groups' probability assessments on the control tasks were quite similar. In the reasons conditions, usage of .5 tended to increase for all groups, whereas usage of 1.0 decreased somewhat. As in Study 1, a substantial group of subjects never used 1.0 (see Table 3). Some 28% followed this pattern in the control condition and 46.9% in the reasons conditions. This increase was much greater for the contradicting and both-reasons groups, over half of whose subjects never used 1.0 in the experimental conditions. The calibration curves for all subjects who did not use 1.0 showed them to be less overconfident and generally better calibrated than the remaining subjects, both for reasons and controls. As in Study 1, the task was equally difficult for users and non-users.

Discussion

The major results of Study 1 have been replicated: Calibration curves for overconfidence in forecasts resemble those for confidence in general knowledge questions. The reasons manipulations had at best weak effects on overall calibration. The contradicting and both-reasons manipulations did, however, again reduce usage of 1.0. In general, subjects who never used 1.0 were better calibrated than their counterparts.

The overall similarity of the present confidence judgments to those observed elsewhere is encouraging for anyone who would like to exploit that literature for the elicitation and interpretation of forecasts. For example, we would expect the training techniques that have proven effective or ineffective with general knowledge items to have similar effects on the calibration of forecasters. The difference observed here between users and non-users of 1.0 may offer an additional tool for determining how much faith to place in others' confidence assessments. The weakness of the reasons manipulations is, however, disappointing, because it suggests that a simple mechanism that has proven effective in improving calibration is not as robust as one would hope. Before writing off this procedure and discussing some possible implications of this research for forecasting, we will offer one further replication designed to strengthen the reasons manipulation.

Study 3.

Method

Although this study was essentially a replication of the previous two, a number of changes were introduced in order to strengthen the reasons manipulation: (a) the number of items per page was reduced from 4 to 3 in order to present a less cramped format; (b) subjects were asked to produce not one, but two reasons of the type required by each condition; (c) The instructions were changed to emphasize that the task involved making predictions about future events, and that descriptions of things heard or read, beliefs and associations could all be used as reasons for the predictions made; (d) subjects were asked to make a special effort to be as complete in describing their reasons as possible; (e) subjects were assured that sufficient time had been allotted in the experiment for them to devote thought to the task. All stimuli dealt with events whose outcome would be known during the first week of June, 1981. Technical aspects of subject recruitment caused responses to be elicited on two separate dates, May 15 and May 29, two weeks before the event period and immediately before. On May 15, half of the participants were in each of the supporting and contradicting reasons groups. Comparisons between the corresponding supporting groups at the two times will reveal whether proximity to events has any effect on calibration beyond its effect on difficulty. One hundred and seventy-three individuals participated, with roughly equal numbers on the two dates.

Results

Timing. The proportion of correct responses to control questions was higher by .03 for the supporting group from May 29 than for the May 15 supporting group, perhaps due to the former's closer proximity to the events in question. The May 29 group's confidence (and overconfidence) was correspondingly higher, leaving their calibration curves quite similar. Corresponding changes were seen in the two groups' responses to the experimental condition items, except that the May 29 group was a bit less overconfident (.074 vs. .101). As there is no apparent reason for this anomaly, the two groups' data from the two dates will be pooled in the following analyses.

Main effect. Table 5 shows the same patterns in responses to the control questions as were seen in Studies 1 and 2. Each group is somewhat overconfident. The poor calibration statistic for the contradicting reasons group, despite its relatively low overconfidence, reflects a very flat calibration curve, with only a .12 difference between the proportions correct associated with responses of .5 and 1.0.

Reasons manipulation. As indicated by Table 5, the reasons manipulations slightly reduced overconfidence and slightly improved calibration for all three groups. As shown in Table 3, they also reduced the usage of 1.0. All these effects were somewhat larger for the contradicting and boht-reasons groups. As before, non-users of 1.0 were considerably better calibrated than users.

DISCUSSION

Three clear patterns have emerged from these three studies, each with some possible implications for forecasting practitioners:

TABLE 5

SUMMARY OF STATISTICS - STUDY 3

Group	N	Control n	Control Prop. Cor.	Control Mean Conf.	Control Over-conf.	Control Calib.	Experiment n	Experiment Prop. Cor.	Experiment Mean Conf.	Experiment Over-conf.	Experiment Calib.
Supporting	91	2625	.650	.746	.096	.0198	2610	.657	.745	.088	.0151
Contradicting	38	1098	.655	.724	.069	.0275	1086	.656	.706	.051	.0231
Both	44	1264	.654	.737	.083	.0186	1258	.652	.723	.071	.0113
All	164	4987	.652	.739	.087	.0201	4954	.655	.731	.076	.0172

1. Calibration for confidence assessments regarding forecasts is largely indistinguishable from that pertaining to general knowledge questions. The overconfidence scores and calibration curves observed with the control items here were very similar to those observed with general knowledge items of similar difficulty. On the basis of these results, one should have considerably increased confidence in extrapolating the results of earlier calibration research to confidence in forecasts. Thus, one might expect calibration for forecasts to be relatively unaffected by changes in response mode, incentive payments for correct answers, or familiarity with subject matter (unless accompanied by a change in difficulty), to generalize a few results from Table 1.

2. The only apparent difference between these responses and those observed previously was the appearance of a subsample of subjects who never used 1.0. Over all three studies, such subjects constituted 23.8% of the control groups and 39.8% of the experimental groups. As shown in Table 6, non-users of 1.0 were consistently much better calibrated than users, in all three studies, for both control and experimental items. Figure 6 pools responses of users and non-users across the three studies. Each curve includes 5,000-15,000 responses made by 100 to 300 subjects. Although non-users are somewhat better calibrated for most probability values, the major difference between the groups is at 1.0. Non-users simply do not produce the point that represents the greatest overconfidence, that is, the greatest discrepancy between how often one should be correct and how often one is. On the basis of these results, one might tentatively extend greater credence to the confidence assessments of forecasters who never express complete certitude.

3. The reasons manipulations had consistent but weak effects. In each study, responses in the experimental condition were better calibrated and less overconfident than those in the corresponding control conditions. Over

TABLE 6

SUMMARY STATISTICS FOR USERS AND NON-USERS OF 1.0

GROUP MEANS

Study	1		2		3		All	
Group	Control	Exp.	Control	Exp.	Control	Exp.	Control	Exp.
Prop. Correct								
Users	.617	.625	.635	.625	.656	.653	.638	.639
Non-users	.320	.613	.614	.629	.636	.660	.623	.636
Mean Prob.								
Users	.736	.736	.751	.757	.750	.748	.747	.748
Non-users	.075	.660	.686	.673	.685	.685	.683	.674
Overconfidence								
Users	.119	.111	.115	.132	.094	.095	.108	.109
Non-users	.055	.047	.072	.047	.049	.025	.060	.038
Calibration								
Users	.0249	.0212	.0197	.0257	.0218	.0182	.0215	.0208
Non-users	.0065	.0081	.0065	.0065	.0121	.0079	.0078	.0068

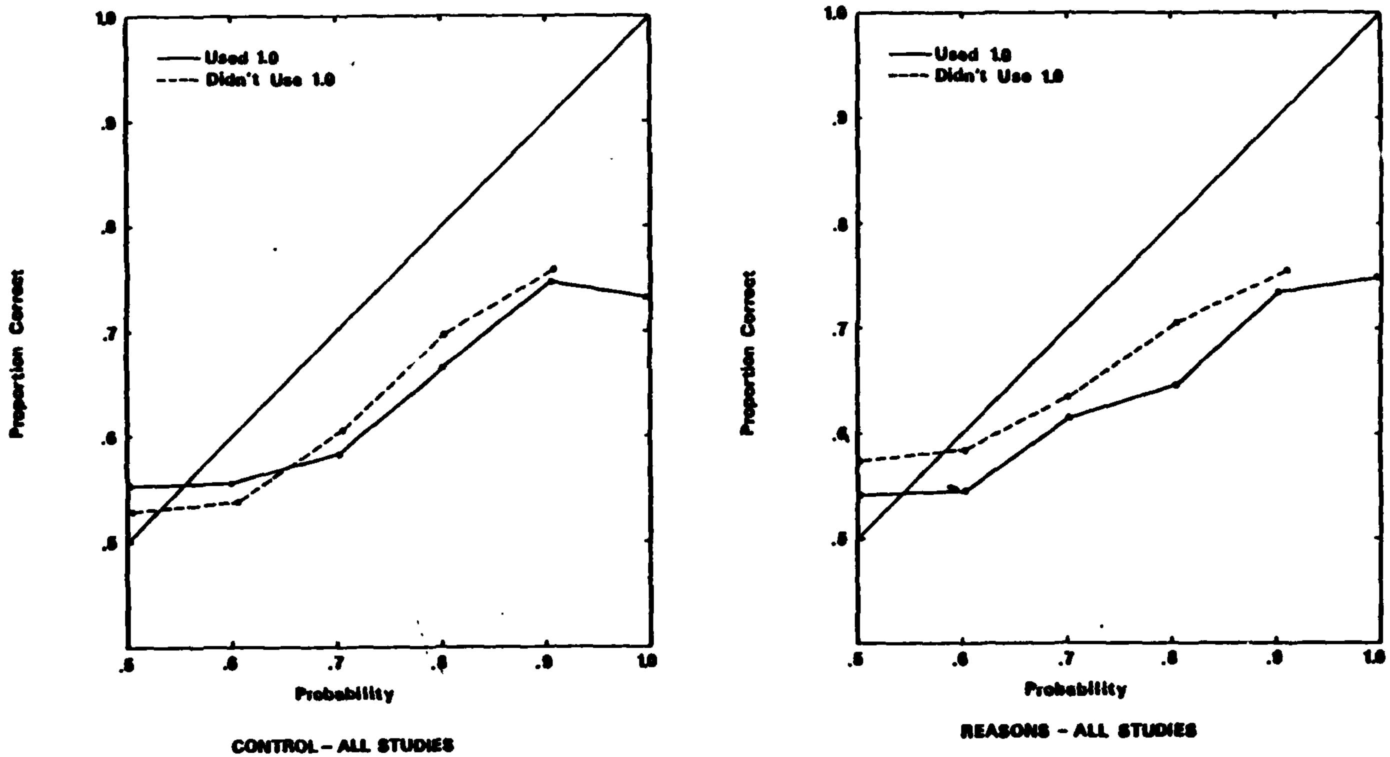


FIGURE 6.
CALIBRATION CURVES FOR USERS AND NON-USERS OF 1.0, POOLED ACROSS STUDIES 1-3. CORRESPONDING SUMMARY STATISTICS ARE GIVEN IN EXHIBIT 11. CURVES INVOLVE APPROXIMATELY 5,000 to 16,000 RESPONSES PRODUCED BY 100 TO 300 INDIVIDUALS.

all three studies, overconfidence decreased by .008 for supporting reasons subjects (from .101 to .093), by .007 for contradicting reasons subjects (from .086 to .079), and by .032 for both-reasons subjects (from .101 to .069). In an applied situation, one might wonder if such modest improvements were worth the additional time and effort the provision of reasons requires. Of course, one might also feel that the provision of explicit reasons has desirable features independent of its impact on calibration. These could include (a) providing a record of the reasons motivating one's forecasts in order to avoid the prejudicial effects of hindsight bias when the time comes to evaluate them, once the event has or has not happened (Fischhoff, 1975); (b) allowing for external review of one's reasoning, perhaps leading to the correction of misconceptions or improved communications (Hogarth & Makridakis, 1981); or (c) helping raise one's alertness to new evidence that should prompt revisions of a forecast (Armstrong, 1978).

It is worth noting in this context that the most dramatic effect demonstrated by Koriat, Lichtenstein and Fischhoff (1980) was found with a much more involved procedure than that depicted in Figure 3 and repeated in the present studies. In a separate experiment, they required subjects to complete a 2 x 2 matrix giving reasons for and against each of the two possible answers. Ten, rather than thirty, items were used in that study. The more ambitious and focused manipulation reduced confidence by .023, while increasing the percentage of correct answers by .040, thereby reducing overconfidence by .063. Perhaps one must conclude that provision of one or two reasons for each of a fairly large number of items cannot hurt, but it cannot be counted to help very much.

The most consistent effect of the reasons manipulations, in particular the provision of contradicting or both reasons, was to increase the proportion of subjects who never used 1.0. As mentioned, these non-users were better calibrated than users in both the control and experimental conditions. Indeed, one might speculate that the primary effect of the reasons manipulations is to indirectly convince some people never to be entirely certain.

ACKNOWLEDGEMENTS

This research was supported by the Office of Naval Research under contract N00014-80-C-0150 to Perceptronics, Inc. We would like to thank Gerry Hanson and Mark Layman for their help in various parts of this enterprise.

FOOTNOTES

1. An alternative use of "calibration" found in the forecasting literature is "to extimate the relationships (and constant terms) in a forecasting model" (Armstrong, 1978, p. 477). In addition, several other terms are at times used to describe the calibration of probability assessments (see Lichtenstein, Fischhoff & Phillips, in press).

References

Adams, J. K., & Adams, P. A. Realism of Confidence Judgments. Psychological Review, 68 (1961), 33-45.

Adams, P. A., & Adams, J. K. Training in Confidence Judgments. American Journal of Psychology, 71 (1958), 747-751.

Alpert, W., & Raiffa, H. A Progress Report on the Training of Probability Assessors. In D. Kahneman, P. Slovic, and A. Tversky (Eds.), Judgment under uncertainty: Heuristics and biases. New York: Cambridge University Press, in press.

Armelius, K. Task Predictability and Performance as Determinants of Confidence in Multiple-cue Judgments. Scandinavian Journal of Psychology, 20 (1979),19-25.

Armstrong, J. S. Long-Range Forecasting: From Crystal Ball to Computer. New York: John Wiley & Sons, Inc., 1978.

Becker, B. W., & Greenberg, M. G. Probability Estimates by Respondents: Does weighting improve accuracy? Journal of Marketing Research, 15 (1978), 482-486.

Beyth-Marom, R., & Dekel, S. Thinking Under Uncertainty: A Textbook for Junior High School Students. In press (in Hebrew)

Cavanaugh, J. C., & Borkowski, J. G. Searching for Meta-Memory Connections. Developmental Psychology, 16 (1980), 441-453.

Clarke, F. R. Confidence Ratings, Second-Choice Responses, and Confusion Matrices in Intelligibility Tests. Journal of the Acoustical Society of America, 32 (1960), 35-46.

Cocozza, J. J., & Steadman, H. J. Prediction in Psychiatry: An Example of Misplaced Confidence in Experts. Social Problems 25 (1978), 265-276.

Dawes, R. M. *Confidence in Intellectual Judgments vs. Confidence in Perceptual Judgments*. In Essays in Honor of Clyde Coombs, in press.

Dowie, J. On the Efficiency and Equity of Betting Markets. Economica, 43 (1976), 139-150.

Dreman, D. Contrarian Investment Strategy. New York: Random House, 1979.

Fischhoff, B. Hindsight ≠ Foresight: The Effect of Outcome Knowledge on Judgment Under Uncertainty. Journal of Experimental Psychology: Human Perceptiona and Performance, 1 (1975), 288-299.

Fischhoff, B. The Effect of Temporal Setting on Likelihood Estimates. Organizational Behavior and Human Performance, 15 (1976), 180-194.

Fischhoff, B. Debiasing. In D. Kahneman, P. Slovic, & A. Tversky (Eds.), Judgment under Uncertainty: Heuristics and Biases. New York: Cambridge University Press, in press.

Fischhoff, B., & Beyth, R. "I Knew It Would Happen":—Remembered Probabilities of Once-Future Things. Organizational Behavior and Human Performance, 13 (1975), 1-16.

Fischhoff, B.,& Slovic, P. A Little Learning...: Confidence in Multi-Cue Judgment. In R. Nickerson (Eds.), Attention and Performance, VIII. Hillsdale, N. J.: Lawrence Erlbaum, 1980.

Fischhoff, B., Slovic, P., & Lichtenstein, S. Knowing With Certainty: The appropriateness of Extreme Confidence. Journal of Experimental Psychology: Human Perception and Performance, 3 (1977), 552-564.

Hogarth, R. M., & Makridakis, S. Forecasting and Planning: An Appraisal. Management Science, 27 (1981), 115-138.

Howell, W. C. & Burnett, S. A. Uncertainty Measurement: A Cognitive Taxonomy. Organizational Behavior and Human Performance, 22 (1978), 45-68.

Hynes, M., & Vanmarcke, E. Reliability of embankment Performance Predictions. _Proceedings of the ASCE Engineering Mechanics Division Specialty Conference_. Waterloo, Ontario, Canada: University of Waterloo Press, 1976.

King, J. F., Zechenmeister, E. B., & Shaughnessy, J. J. Judgment of Knowing: The Influence of Retrieval Practice. _American Journal of Psychology_, in press.

Koriat, A., Lichtenstein, S., & Fischhoff, B. Reasons for Confidence. _Journal of Experimental Psychology: Human Learning and Memory_, _6_ (1980), 107-118.

Larson, J. R., & Reenan, A. M. The Equivalence Interval as a Measure of Uncertainty. _Organizational Behavior and Human Performance_, _23_ (1979), 49-55.

Lichtenstein, S., & Fischhoff, B. Do Those Who Know More Also Know More About How Much They Know? The Calibration of Probability Judgments. _Organizational Behavior and Human Performance_, _20_ (1977), 159-183.

Lichtenstein, S., & Fischhoff, B. Training for Calibration. _Organizational Behavior and Human Performance_, _26_ (1980), 149-171.

Lichtenstein, S., Fischhoff, B., & Phillips, L. D. Calibration of Probabilities: The State of the Art to 1980. In D. Kahneman, P. Slovic, & A. Tversky (Eds.), _Judgment Under Uncertainty: Heuristics and Biases_. New York: Cambridge University Press, in press.

Ludke, R. L., Stauss, F. F., & Gustafson, D. H. Comparison of Five Methods for Estimating Subjective Probability Distributions. _Organizational Behavior and Human Performance_, _19_ (1977), 162-179.

Moore, P. G. The Manager's Struggle with Uncertainty. Journal of The Royal Statistical Society, Series A, 140 (1977), 129-165.

Morris, P. A. Decision Analysis Expert Use. Management Science, 20 (1974), 1233-1241.

Murphy, A. H., & Winkler, R. L. Subjective Probability Forecasting Experiments in Meteorology: Some Preliminary Results. Bulletin of the American Meteorological Society, 55 (1974), 1206-1216.

Murphy, A. H., & Winkler, R. Can Weather Forecasters Formulate Reliable Probability Forecasts of Precipitation and Temperature? National Weather Digest, 2 (1977), 2-9.

Nickerson, R. S., & McGoldrick, C. C., Jr. Confidence Ratings and Level of Performance on a Judgmental Task. Perceptual and Motor Skills, 20 (1965), 311-316.

Oskamp, S. The Relationship of Clinical Experience and Training Methods to Several Criteria of Clinical Prediction. Psychological Monographs, 76 (1962), (28, Whole No. 547).

Phillips, L. D., & Wright, G. N. Cultural Differences in Viewing Uncertainty and Assessing Probabilities. In H. Jungermann, and G. deZeeuw (Eds.), Decision Making and Change in Human Affairs. Amsterdam: D. Reidel, 1977.

Pickhardt, R. C., & Wallace, J. B. A Study of the Performance of Subjective Probability Assessors. Decision Sciences, 5 (1974), 347-363.

Pitz, G. F. Subjective Probability Distributions for Imperfectly Known Quantities. In L. W. Gregg (Ed.), Knowledge and Cognition. New York: Wiley, 1974.

Root, H. E. Probability Statements in Weather Forecasting. Journal of Applied Meteorology, 1 (1962), 163-168.

Schaefer, R. E., & Borcherding, K. The Assessment of Subjective Probability Distribution: A Training Experiment. Acta Psychologica, 37 (1973), 117-129.

Seaver, D. A., von Winterfeldt, D., & Edwards, W. Eliciting Subjective Probability Distributions on Continuous Variables. Organizational Behavior and Human Performance, 21 (1978), 379-391.

Selvidge, J. Assessing the Extremes of Probability Distributions by the Fractile Method. Decision Sciences, 11 (1980), 493-502.

Slovic, P. Psychological Study of Human Judgment: Implications for Investment Decision Making. Journal of Finance, 27 (1972), 779-799.

Staël von Holstein, C.-A. S. An Experiment in Probabilistic Weather Forecasting. Journal of Applied Meteorology, 10 (1971), 635-645.

Staël von Holstein, C.-A. S. Probabilistic Forecasting: An Experiment Related to the Stock Market. Organizational Behavior and Human Performance, 8 (1972), 139-158.

Tversky, A., & Kahneman, C.-A. S. Judgment Under Uncertainty: Heuristics and Biases. Science, 185 (1974), 1124-1131.

Wright, G., & Wisudha, A. Differences in Calibration for Past and Future Events. Paper presented at the Seventh Research Conference on Subjective Probability, Utility and Decision Making. Göteborg, Sweden, 1979.

OFFICE OF NAVAL RESEARCH

TECHNICAL REPORTS DISTRIBUTION LIST

CDR Paul R. Chatelier
Office of the Deputy Under Secretary of Defense
OUSDRE (E&LS)
Pentagon, Room 3D129
Washington, D.C. 20301

Engineering Psychology Programs
Code 422
Office of Naval Research
800 North Quincy Street
Arlington, VA 22217 (5 cys)

Manpower, Personnel & Training Programs
Code 270
Office of Naval Research
800 North Quincy Street
Arlington, VA 22217

Operations Research Programs
Code 411-OR
Office of Naval Research
800 North Quincy Street
Arlington, VA 22217

Statistics & Probability Program
Code 411-S&P
Office of Naval Research
800 North Quincy Street
Arlington, VA 22217

Information Systems Program
Code 411-IS
Office of Naval Research
800 North Quincy Street
Arlington, VA 22217

CDR K. Hull
Code 410B
Office of Naval Research
800 North Quincy Street
Arlington, VA 22217

Physiology & Neuro Biology Programs
Code 441
Office of Naval Research
800 North Quincy Street
Arlington, VA 22217

Commanding Officer
ONR Eastern/Central Regional Office
ATTN: Dr. J. Lester
Bldg. 114, Section D
666 Summer Street
Boston, MA 02210

Commanding Officer
ONR Western Regional Office
ATTN: Dr. E. Gloye
1030 East Green Street
Pasadena, CA 91106

Office of Naval Research
Scientific Liaison Group
American Embassy, Room A-407
APO San Francisco, CA 96503

Director
Naval Research Laboratory
Technical Information Division
Code 2627
Washington, D.C. 20375

Dr. Michael Melich
Communications Sciences Division
Code 7500
Naval Research Laboratory
Washington, D.C. 20375

Dr. Robert G. Smith
Office of the Chief of Naval Operations, OP987H
Personnel Logistics Plans
Washington, D.C. 20350

Human Factors Department
Code N215
Naval Training Equipment Center
Orlando, FL 32813

Dr. Alfred F. Smode
Training Analysis & Evaluation Group
Naval Training Equipment Center
Code N-OOT
Orlando, FL 32813

Dr. Albert Colella
Combat Control Systems
Naval Underwater Systems Center
Newport, RI 02940

Dr. Gary Poock
Operations Research Department
Naval Postgraduate School
Monterey, CA 93940

Mr. Warren Lewis
Human Engineering Branch
Code 8231
Naval Ocean Systems Center
San Diego, CA 92152

Dr. A.L. Slafkosky
Scientific Advisor
Commandant of the Marine Corps
Code RD-1
Washington, D.C. 20380

Mr. Arnold Rubinstein
Naval Material Command
NAVMAT 0722 - Rm. 508
800 North Quincy Street
Arlington, VA 22217

Commander
Naval Air Systems Command
Human Factors Programs
NAVAIR 340F
Washington, D.C. 20361

CDR Robert Biersner
Naval Medical R&D Command
Code 44
Naval Medical Center
Bathesda, MD 20014

Dr. Arthur Bachrach
Behavioral Sciences Department
Naval Medical Research Institute
Bethesda, MD 20014

CDR Thomas Berghage
Naval Health Research Center
San Diego, CA 92152

Dr. George Moeller
Human Factors Engineering Branch
Submarine Medical Research Lab
Naval Submarine Base
Groton, CT 06340

Head
Aerospace Psychology Department
Code L5
Naval Aerospace Medical Research Lab
Pensacola, FL 32508

Dr. James McGrath
CINCLANT FLT HQS
Code 04E1
Norfolk, VA 23511

Navy Personnel Research & Development Center
Planning & Appraisal Division
San Diego, CA 92152

Dr. Robert Blanchard
Navy Personnel Research & Development Center
Command & Support Systems
San Diego, CA 92152

LCDR Stephen D. Harris
Human Factors Engineering Division
Naval Air Development Center
Warminster, PA L8974

Dr. Julie Hopson
Human Factors Eningeering Division
Naval Air Development Center
Warminster, PA 18974

Mr. Jeffrey Grossman
Human Factors Branch
Code 3152
Naval Weapons Center
China Lake, CA 93555

Human Factors Engineering Branch
Code 1226
Pacific Missile Test Center
Point Mugu, CA 93042

CDR W. Moroney
Code 55MP
Naval Postgraduate School
Monterey, CA 93940

Dr. Joseph Zeidner
Technical Director
U.S. Army Research Institute
5001 Eisenhower Avenue
Alexandria, VA 22333

Director, Organizations &
Systems Research Laboratory
U.S. Army Research Institute
5001 Eisenhower Avenue
Alexandria, VA 22333

U.S. Air Force Office of Scientific
Research
Life Sciences Directorate, NL
Bolling Air Force Base
Washington, D.C. 20332

Chief, Systems Engineering Branch
Human Engineering Division
USAF AMRL/HES
Wright-Patterson AFB, OH 45433

Dr. Earl Alluisi
Chief Scientist
AFHRL/CCN
Brooks, AFB, TX 78235

Dr. Kenneth Gardner
Applied Psychology Unit
Admiralty Marine Technology
Establishment
Teddington, Middlesex TW11 OLN
ENGLAND

Director, Human Factors Wing
Defense & Civil Institute of
Environmental Medicine
P.O. Box 2000
Downsview, Ontario M3M 3B9
CANADA

Dr. A.D. Baddeley
Director, Applied Psychology Unit
Medical Research Council
15 Chaucer Road
Cambridge, CB2 2EF
ENGLAND

Dr. Robert T. Hennessy
NAS-National Research Council
2101 Constitution Ave., N.W.
Washington, D.C. 20418

Dr. M.G. Samet
Perceptronics, Inc.
6271 Variel Avenue
Woodland Hills, CA 91367

Dr. Robert Williges
Human Factors Laboratory
Virginia Polytechnic Institute
& State University
130 Whittemore Hall
Blacksburg, VA 24061

Dr. Alphonse Chapanis
Department of Psychology
The Johns Hopkins University
Charles & 34th Streets
Baltimore, MD 21218

Dr. Ward Edwards
Director, Social Science Research
Institute
University of Southern California
Los Angeles, CA 90007

Dr. Charles Gettys
Department of Psychology
University of Oklahoma
455 West Lindsey
Norman, OK 73069

Dr. Kenneth Hammond
Institute of Behavioral Science
University of Colorado
Room 201
Boulder, CO 80309

Dr. James H. Howard, Jr.
Department of Psychology
Catholic University
Washington, D.C. 20064

Dr. William Howell
Department of Psychology
Rice University
Houston, TX 77001

Dr. Christopher Wickens
University of Illinois
Department of Psychology
Urbana, IL 61801

Defense Technical Information Center
Cameron Station, Bldg. 5
Alexandria, VA 22314 (12 cys)

Dr. Judith Daly
System Sciences Office
Defense Advanced Research Projects Agency
1400 Wilson Blvd.
Arlington, VA 22209

Dr. Robert R. Mackie
Human Factors Research, Inc.
5775 Dawson Avenue
Goleta, CA 93017

Dr. Gary McClelland
Institute of Behavioral Sciences
University of Colorado
Boulder, CO 80309

Dr. Jesse Orlansky
Institute for Defense Analyses
400 Army-Navy Drive
Arlington, VA 22202

Dr. T. B. Sheridan
Department of Mechanical Engineering
Massachusetts Institute of Technology
Cambridge, MA 02139

Dr. Paul Slovic
Decision Research
1201 Oak Street
Eugene, OR 97401

Dr. Harry Snyder
Department of Industrial Engineering
Virginia Polytechnic Institute and State University
Blacksburg, VA 24061

Dr. Amos Tversky
Department of Psychology
Stanford University
Stanford, CA 94305

Dr. W. S. Vaughan
Oceanautics, Inc.
422 6th Street
Annapolis, MD 21403

Dr. Richard W. Pew
Information Sciences Division
Bolt Beranek & Newman, Inc.
50 Moulton Street
Cambridge, MA 02238

Dr. Hillel Einhorn
University of Chicago
Graduate School of Business
1101 E. 58th Street
Chicago, IL 60637

Dr. John Payne
Duke University
Graduate School of Business Administration
Durham, NC 27706

Dr. Baruch Fischhoff
Decision Research
1201 Oak Street
Eugene, OR 97401

Dr. Andrew P. Sage
University of Virginia
School of Engineering and
  Applied Science
Charlottesville, VA 22901

Dr. Leonard Adelman
Decisions and Designs, Inc.
8400 Westpark Drive, Suite 600
P.O. Box 907
McLean, VA 22101

Dr. Lola Lopes
Department of Psychology
University of Wisconsin
Madison, WI 53706

Mr. Joseph G. Wohl
Alphatech, Inc.
3 New England Industrial Park
Burlington, MA 01803

Dr. Rex Brown
Decision Science Consortium
Suite 721
7700 Leesburg Pike
Falls Church, VA 22043

Dr. Wayne Zachary
Analytics, Inc.
2500 Maryland Road
Willow Grove, PA 19090

DATE
FILMED
-'8