AD 728741

ARPA ORDER NO. 189-1

R-678-ARPA

May 1971

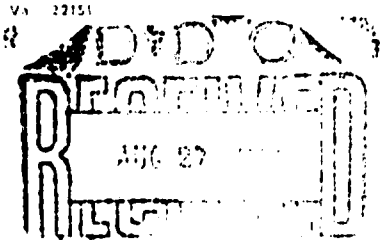
COMPARISON OF GROUP JUDGMENT TECHNIQUES WITH SHORT-RANGE PREDICTIONS AND ALMANAC QUESTIONS

Norman Dalkey and Bernice Brown

A Report prepared for

ADVANCED RESEARCH PROJECTS AGENCY

Reproduced by
NATIONAL TECHNICAL
INFORMATION SERVICE
U.S. Department of Commerce
Springfield VA 22151



Rand
SANTA MONICA, CA 90406

DOCUMENT CONTROL DATA

1. ORIGINATING ACTIVITY The Rand Corporation	2a. REPORT SECURITY CLASSIFICATION UNCLASSIFIED 2b. GROUP

3. REPORT TITLE

COMPARISON OF GROUP JUDGMENT TECHNIQUES WITH SHORT-RANGE PREDICITONS AND ALMANAC QUESTIONS

4. AUTHOR(S) (Last name, first name, initial)

Dalkey, Norman, Bernice Brown

5. REPORT DATE	6a. TOTAL NO OF PAGES	6b. NO. OF REFS.
May 71	34	9

7. CONTRACT OR GRANT NO.	8. ORIGINATOR'S REPORT NO.
DAHC15 67 C 0141	R-678-ARPA

9a. AVAILABILITY/LIMITATION NOTICES	9b. SPONSORING AGENCY
DDC-1	Advanced Research Projects Agency

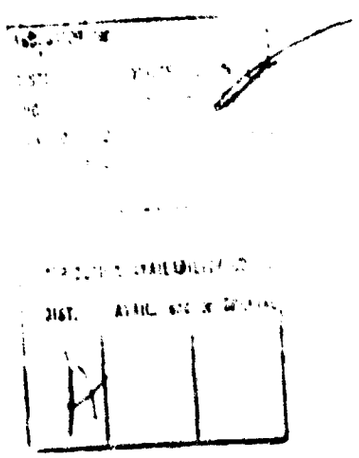
10. ABSTRACT

An experiment designed to discover whether the results of laboratory studies dealing with general (almanac) information are relevant to the applied case when the true answer is unknown. Using short-range prediction questions as subject matter, the experiment indicates that, in general, Delphi procedures are at least as effective with short-range prediction as they have been for almanac material. Eight groups, of about 20 each, of upperclassmen and college graduates were given short-range prediction questions to answer in a 2-round Delphi exercise. Satisfactory answers were obtained for 32 of the 40 questions. Correlations between standard deviation and accuracy, and between group self-rating and accuracy, were significantly higher for the prediction than for the almanac questions. Half the groups generated estimates of the 3 quartiles of the distribution; the other half generated point estimates. No significant difference was observed between these two kinds of estimates.

11. KEY WORDS

Delphi
Forecasting

This research is supported by the Advanced Research Projects Agency under Contract No. DAHC 15 67 C 0141. Views or conclusions contained in this study should not be interpreted as representing the official opinion or policy of Rand or of ARPA.



ARPA ORDER NO. 189-1

R-678-ARPA

May 1971

COMPARISON OF GROUP JUDGMENT TECHNIQUES WITH SHORT-RANGE PREDICTIONS AND ALMANAC QUESTIONS

Norman Dalkey and Bernice Brown

A Report prepared for

ADVANCED RESEARCH PROJECTS AGENCY

Rand

SANTA MONICA, CA. 90406

Rand maintains a number of special subject bibliographies containing abstracts of Rand publications in fields of wide current interest. The following bibliographies are available upon request:

Africa • Arms Control • Civil Defense • Combinatorics
Communication Satellites • Communication Systems • Communist China
Computing Technology • Decisionmaking • Delphi • East-West Trade
Education • Foreign Aid • Foreign Policy Issues • Game Theory
Health-related Research • Latin America • Linguistics • Maintenance
Mathematical Modeling of Physiological Processes • Middle East
Policy Sciences • Pollution • Program Budgeting
SIMSCRIPT and Its Applications • Southeast Asia • Systems Analysis
Television • Transportation • Urban Problems • USSR/East Europe
Water Resources • Weapon Systems Acquisition
Weather Forecasting and Control

To obtain copies of these bibliographies, and to receive information on how to obtain copies of individual publications, write to: Communications Department, Rand, 1700 Main Street, Santa Monica, California 90406.

Published by The Rand Corporation

PREFACE

The Group Judgment Technology project at Rand is a continuing research activity concerned with developing improved procedures for formulating expert opinion. The application of group judgment techniques (Delphi) to decisionmaking in both military and nonmilitary governmental agencies and in industry is increasing rapidly; accordingly, the design of more effective procedures is of increasing practical importance.

The experiment described in this Report was designed to shed light on the question of whether the results of laboratory studies dealing with general information (almanac) subject matter are relevant to the applied case where the true answer is unknown. The experiment used short-range prediction questions as subject matter. In general, the experiment indicates that Delphi procedures are at least as effective with short-range prediction material as they have been for almanac material.

The Group Judgment Technology project is being conducted for the Advanced Research Projects Agency. For those interested in reports of project activity see list of references, p. 27.

SUMMARY

In the experiment described in this Report, 8 groups of upper-class and graduate college students of about 20 subjects each were given 40 short-range prediction questions to answer in a 2-round Delphi exercise; satisfactory answers were later obtained for 32 of these questions. The proportion of questions on which groups improved their answers between round 1 and round 2 was about the same as for similar exercises with almanac questions; the proportion of questions on which answers became less accurate was about half that for almanac questions. Correlations between standard deviation and accuracy, and between group self-ratings and accuracy were significantly higher for the prediction questions. Half of the groups generated estimates of the three quartiles of the distribution; the other half generated point estimates. No significant difference was observed between these two kinds of estimates.

CONTENTS

COMPARISON OF GROUP JUDGMENT TECHNIQUES WITH SHORT-RANGE PREDICTIONS AND ALMANAC QUESTIONS

Norman Dalkey and Bernice Brown

I. INTRODUCTION

An extensive series of experiments has been conducted at Rand to assess the effectiveness of a set of systematic procedures (Delphi) for the formulation of group judgment. The general outcome of these experiments has been that the systematic techniques show distinct advantages over traditional, less formal ways of pooling the judgments of group members. Most of these experiments have been conducted using general information (almanac) questions, where the subjects did not know the answers to the questions, but the answers were available in some reference work. Of course, in applications the interest is in the case where the answer is not known, and where the best information available is the judgment of knowledgeable individuals.

A few experiments have been conducted at Rand and elsewhere [4-7] which have dealt with forecasts, usually of short-range economic and social events expected to occur within a year or less, where the answers were unknown at the time of the experiment. The results of these experiments have been compatible with the results of the experiments

dealing with almanac material, but the data generated were in a form that makes direct comparison with the almanac experiments difficult. The question thus remained rather open as to whether the results obtained with almanac subject matter are applicable to situations involving "objective uncertainty," i.e., where the answers to questions do not already exist in some form.

The experiment described in this Report was intended to cast some additional light on this question. The experiment did not simulate applied studies in their entirety; subjects were college students, and the questions dealt with simple forecasts of items of general interest -- demographic, economic, and political events. However, they did involve the element of "objective uncertainty." It was necessary to wait until the events had transpired to evaluate the forecasts. In the experiment, 151 upper-class and graduate students from UCLA were divided into 8 groups (4 experimental and 4 comparison groups) and each group made 20 forecasts. In all, 40 forecasts were made; 4 of the 8 groups answered one set of 20 questions, and the other 4 answered the remaining 20. Satisfactory answers were obtained for 32 of the 40 questions.

In addition to gathering data on forecast material, the experiment included a secondary purpose, namely, to compare performance using distributional estimates rather than point estimates. The four comparison groups made point estimates of the forecast quantities; the four "experimental" groups made distributional estimates--a Low, Mid, and High estimate--defined as the three quartiles of the estimated probability distributions.

In general, the outcome of the experiment was that the Delphi procedures were at least as effective with short-range forecasts as with almanac material. The proportion of cases in which median estimates changed as a consequence of feedback was somewhat lower for the forecast questions; but for medians that did change, the proportion of cases in which the estimates improved was somewhat higher. Perhaps more significant, the correlations between standard deviation and accuracy and between a group self-rating index and accuracy were distinctly higher for the prediction questions. The experiment gives no basis for expecting that questions involving "objective uncertainty" are inappropriate for Delphi treatment.

As to the comparison of performance using distributional estimates and point estimates, the results were

negative. There was no clear distinction between the groups generating point estimates and those generating distributional estimates, either in terms of accuracy, amount of change on feedback, or in shape of distributions of answers.

II. PURPOSE

The purpose of the experiment was to compare group performance using Delphi procedures on short-range predictions with results obtained previously using almanac questions.

An additional purpose was to test the hypothesis that groups of respondents would show greater accuracy when making distributional estimates (three quartiles) than when making single (point) estimates. A correlative hypothesis to be tested was that groups which were given feedback of the medians of Low, Mid, and High estimates would exhibit more individual changes and more changes of group median than those which were given the quartiles of point estimates.

Method

One hundred fifty-one students from UCLA were paid to serve as respondents. Of these, 71 were male, 80 were female, 21 were graduate students, and 130 were upper-division students. Eight groups of about twenty respondents each were formed; four of these were designated as comparison groups (17, 19, 21, 23) and four experimental

(18, 20, 22, 24).* In May 1969, on each of four days, one comparison and one experimental group were used as respondents. The design of the experiment was as follows:

	Comparison Group	Experimental Group
Round 1	Give self-rating for each question. Answer each of 20 questions with a point estimate.	Give self-rating for each question. Answer each of 20 questions with a Low, Mid, and High estimate.**
	Keep separate record of answers for round 2.	Keep separate record of answers for round 2.
Interim Period	Take Terman's Concept Mastery Test.***	Take Terman's Concept Mastery Test.***
Round 2	Feedback three group quartiles for each question.	Feedback medians of Low, Mid, and High estimates for each question.
	Revise answers to 20 questions.	Revise answers to 20 questions giving Low, Mid, and High estimates of each.

*The group numbers derive from consecutive numbering of groups involved in the 1969 series of experiments.

**The Low estimate is defined as the number that the subject thought has about a 25 percent chance of being larger than the true answer; the Mid estimate is the number that has about an even chance of being larger than the true one; and the High estimate is the one that has a 75 percent chance of being larger than the true answer.

***Terman's Concept Mastery Test, Form T, was used as an interim task while statistics on round 1 answers were being computed. Analysis of the data relating CMT scores and performance will be reported in a subsequent publication.

Questions

A list of the questions used is included in the Appendix. Also shown are the true answers and the average group errors on round 1. There were 40 questions in the experiment. The first 20 were used for groups 17 and 18 and groups 21 and 22. The second set (questions 21-40) was used for groups 19 and 20 and for groups 23 and 24. The period of projection into the future varied from a little less than 1 month to about 6 months.

For eight questions, either the process of getting the answer presented too many complications or the questions had been formulated in such a way that a meaningful answer did not exist in the standard statistical summaries. The questions for which we failed to get answers were 1, 3, 5, 14, 16, 21, 35, and 38.

III. RESULTS

Table 1 summarizes the effect of iteration and feedback for each of the eight groups. As in previous analyses of group judgments [1, pp. 25-26], the measure of group error is defined by

$$E = \left| \ln \frac{\text{median}}{\text{true}} \right|$$

For groups giving Low, Mid, and High estimates (labeled D in the table), the group response was defined as the median of the Mid responses. "Improved" means that the round 2 error was smaller than the round 1 error; "became less accurate" means that the round 2 error was greater than the round 1 error; "remained same" designates no change. Fifteen answers were available for groups 17, 18, 21, and 22; seventeen answers were available for groups 19, 20, 23, and 24.

Table 2 compares the changes between round 1 and round 2 for a set of almanac questions [3] with the present results for prediction questions. The proportion of questions on which improvement occurred was about the same for the two types; the proportion remaining unchanged was higher for the prediction questions; whereas the proportion which became less accurate was distinctly lower

Table 1

EFFECT OF ITERATION AND FEEDBACK

	Change on Interation (Number of Questions)		
Group	Improved	Remained Same	Became Less Accurate
17-P[a]	3	11	1
18-D[b]	3	11	1
19-P	6	9	2
20-D	6	9	2
21-P	6	6	3
22-D	9	3	3
23-P	9	6	2
24-D	3	14	0
Total P	24	32	8
Total D	21	37	6
Total 8 groups	45	69	14

[a]Point estimate.

[b]Distributional estimate.

Table 2

COMPARISON OF CHANGES BETWEEN ROUND 1 AND ROUND 2 BY TYPE OF QUESTION

	Change on Iteration (Percent)		
	Improved	Remained Same	Became Less Accurate
Almanac questions	36	39	25
Prediction questions	35	54	11

for the prediction questions. For the questions on which there was a change, the proportion improved was .60 for almanac questions and .76 for prediction questions.

Table 3 displays average errors for a series of experiments using almanac questions and the present experiment using prediction questions. Since the experiments with almanac questions involved differing task conditions on round 2 for the experimental groups, the more meaningful comparison is among the control groups, and round 1 for the experimental groups. The large error reduction between round 1 and round 2 for experimental groups in the set labeled 9-16 is due to the input of an additional hard fact on round 2. The table indicates that the performance of our subjects on the prediction questions was quite similar to their performance on the almanac questions. It will be noted that the error reduction between round 1 and round 2 is approximately the same for all the control groups--between 4 and 5 percent.

Table 3

AVERAGE ERROR FOR SEVERAL EXPERIMENTAL SERIES

Group Data	Control		Experimental		Number of Questions
	round 1	round 2	round 1	round 2	
1968 (8)	1.04	1.00	1.08	.97	160 Almanac
1969 (1-8)	.84	.80	1.01	1.00	80 Almanac
1969 (9-16)	1.20	1.14	1.28	.81	80 Almanac
1969 (17-24)	1.00	.96	.92	.89	32 Prediction

The average round 1 standard deviation for the 160 questions answered by groups 1-16 (almanac) was 1.9; the average round 1 standard deviation for the 40 questions in the prediction experiment was 1.4. The difference between these two is statistically significant ($t = 3.2$, $p < .001$ on a two-tailed test). Since the average errors are about the same for the almanac and prediction questions, the smaller average standard deviation for the latter would indicate a slightly higher bias [1, p. 12].

Figures 1 and 2 display the scatter diagrams and least-squares estimates of group error on standard deviation. The increase in slope of the estimation line on round 2 represents a much larger change than was reported for almanac questions in [1]. This may be due to the fact that the regression was computed for grouped data in [1] and for ungrouped data in Fig. 2. In any event, Fig. 2 confirms for prediction questions the conclusion previously drawn for almanac data--that the reduction in dispersion between round 1 and round 2 as a result of feedback represents overconvergence. The reduction in dispersion is much greater than the reduction in error. This conclusion is strengthened by examining the bias, error/standard deviation, for individual questions. Of the 32 questions for which

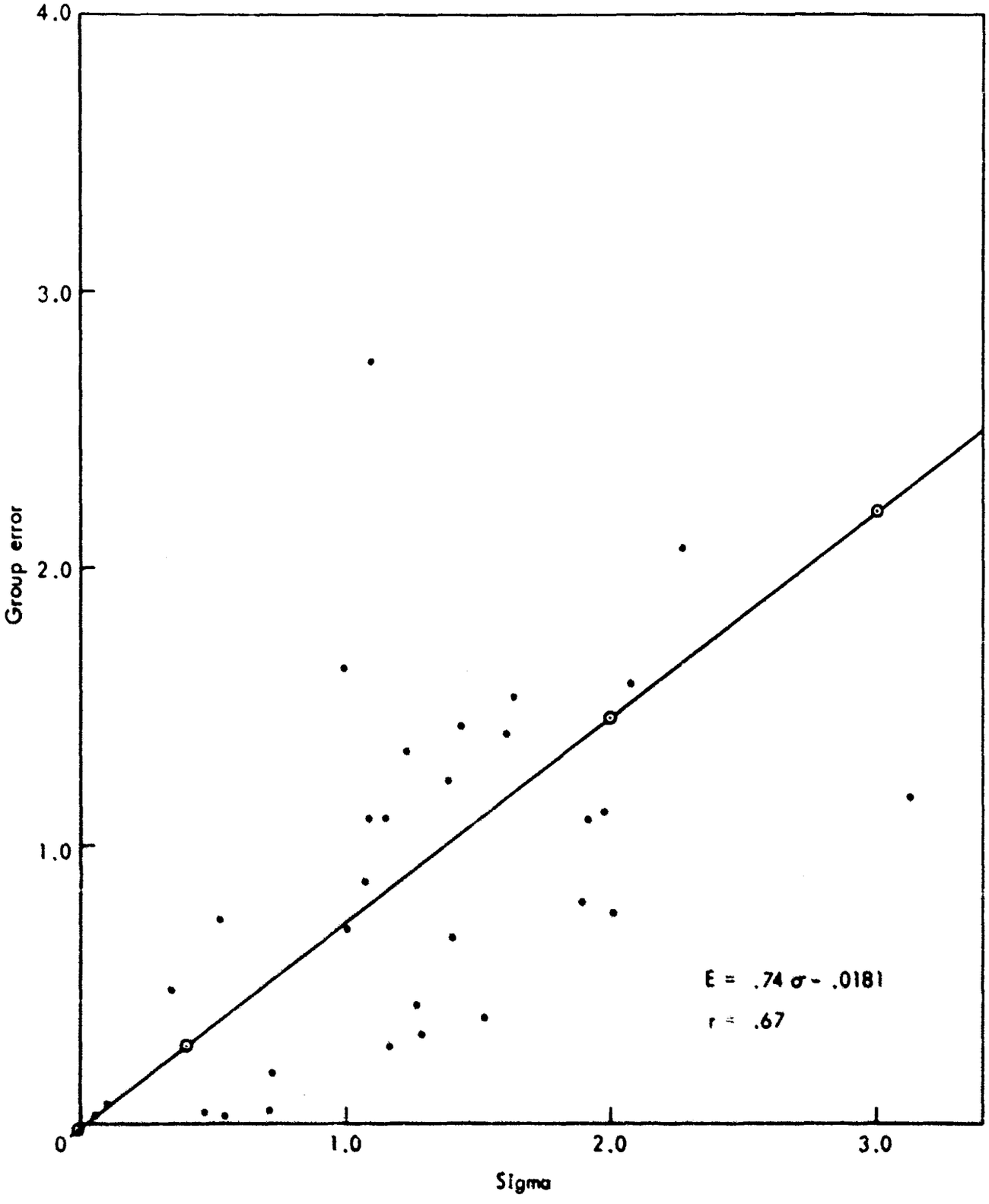


Fig. 1—Round 1 group error versus standard deviation

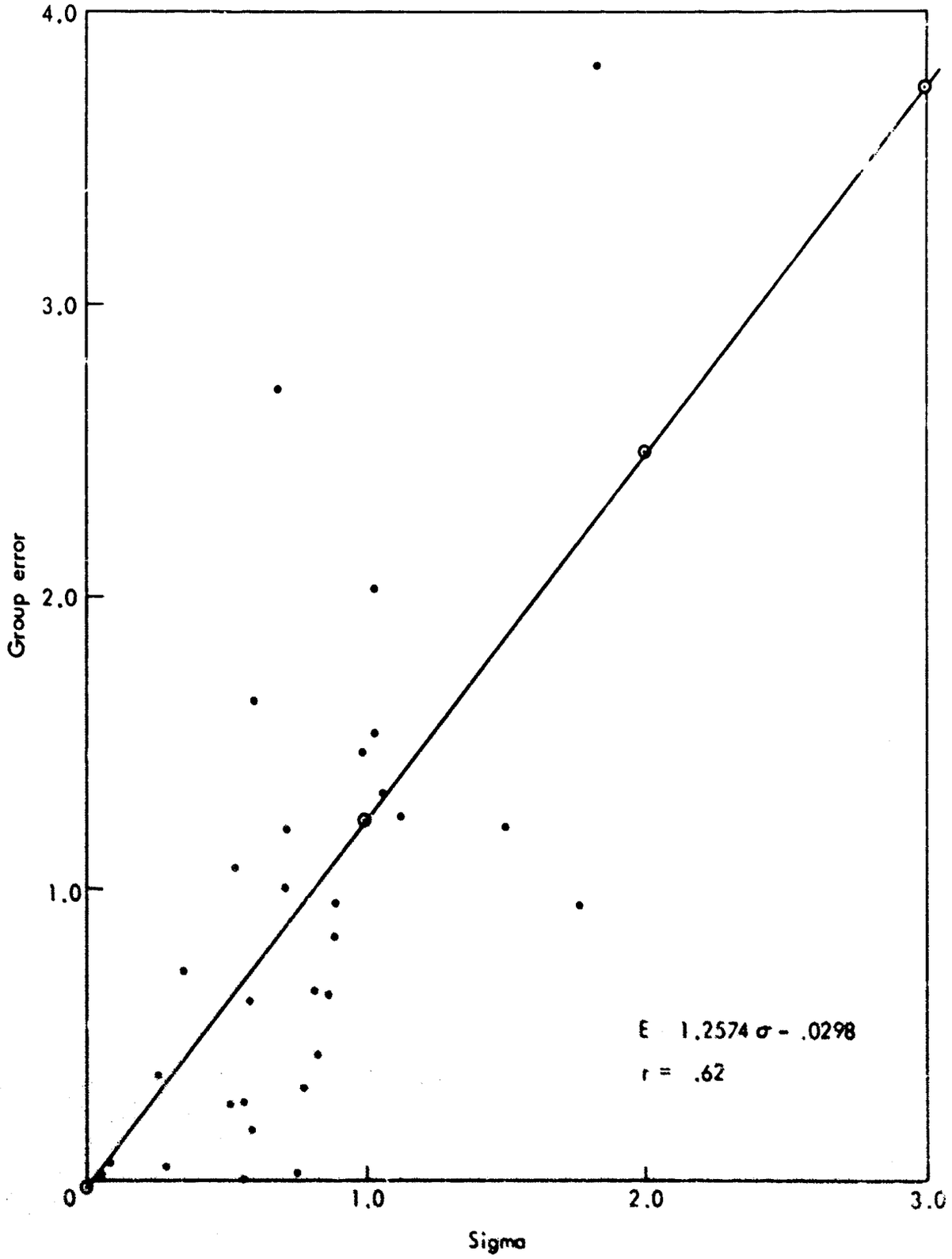


Fig. 2 — Round 2 error versus standard deviation

we had answers, the bias decreased between round 1 and round 2 on only three questions. For two of these, the bias was very small (.05 and .03) and the change was slight. For the remaining questions, the bias increased between round 1 and round 2 by a median factor of 1.8.

The most marked difference between the prediction and almanac results concerns the relations between self-ratings and accuracy and standard deviation and accuracy. In previous experiments with almanac questions, both self-rating and standard deviation have shown significant correlation with accuracy, leading to the conclusion that they are useful indices of the "excellence" of the group's judgments [1, pp. 68ff]. Table 4 compares the almanac and prediction tasks in this regard, where the numbers shown are correlations taken over 16 groups for almanac questions, and over 8 groups for the prediction questions.

Table 4

ROUND 1 CORRELATIONS AMONG VARIABLES FOR ALMANAC AND PREDICTION QUESTIONS

	Correlation			
Type of Question	GSR and Std Dev[a]	GSR and Error	Std Dev and Error	GSR and Std Dev and Error (multiple)
Almanac	-.55	-.46	.39	.49
Prediction	-.67	-.60	.63	.67

[a]Group Self-Rating and Standard Deviation.

Higher correlations are evident for prediction questions across all categories.

One interesting difference between prediction and almanac results concerns improvement depending on initial overestimation or underestimation. For the almanac questions, initial underestimates tend to improve on feedback; overestimates tend to become less accurate. Table 5 shows the results for 148 almanac questions.

Table 5

CHANGE ON ITERATION AS A FUNCTION OF OVERESTIMATION OR UNDERESTIMATION ON ROUND 1, ALMANAC QUESTIONS[a]

	Better on Round 2	Worse on Round 2
Round 1 overestimate	10	28
Round 1 underestimate	78	32

[a] Chi-square for 1 d.f. is 23.3, $p << .001$.

Table 6 displays the same information for prediction questions where there is no significant difference between initial overestimation and underestimation results. At present we have no explanation for this difference in performance on the two types of questions.

Table 6

CHANGE ON ITERATION AS A FUNCTION OF OVERESTIMATION OR UNDERESTIMATION ON ROUND 1, PREDICTION QUESTIONS[a]

	Better on Round 2	Worse on Round 2
Round 1 overestimate	14	4
Round 1 underestimate	31	10

[a] Chi-square for 1 d.f. is .0325, $p > .50$.

Finally, with regard to the shape of distributions, in previous experiments with almanac questions the distributions have tended to be log normal [1, p. 25]. Fig. 3 is the summed distribution of round 1 log responses on all 40 of the prediction questions. The abscissa is in intervals of 0.4 on σ; 7 is the interval $-0.2\ \sigma$ to $+0.2\ \sigma$. A normal curve is shown for comparison. The log normal approximation of the distribution is at least as good as that previously observed for almanac questions.

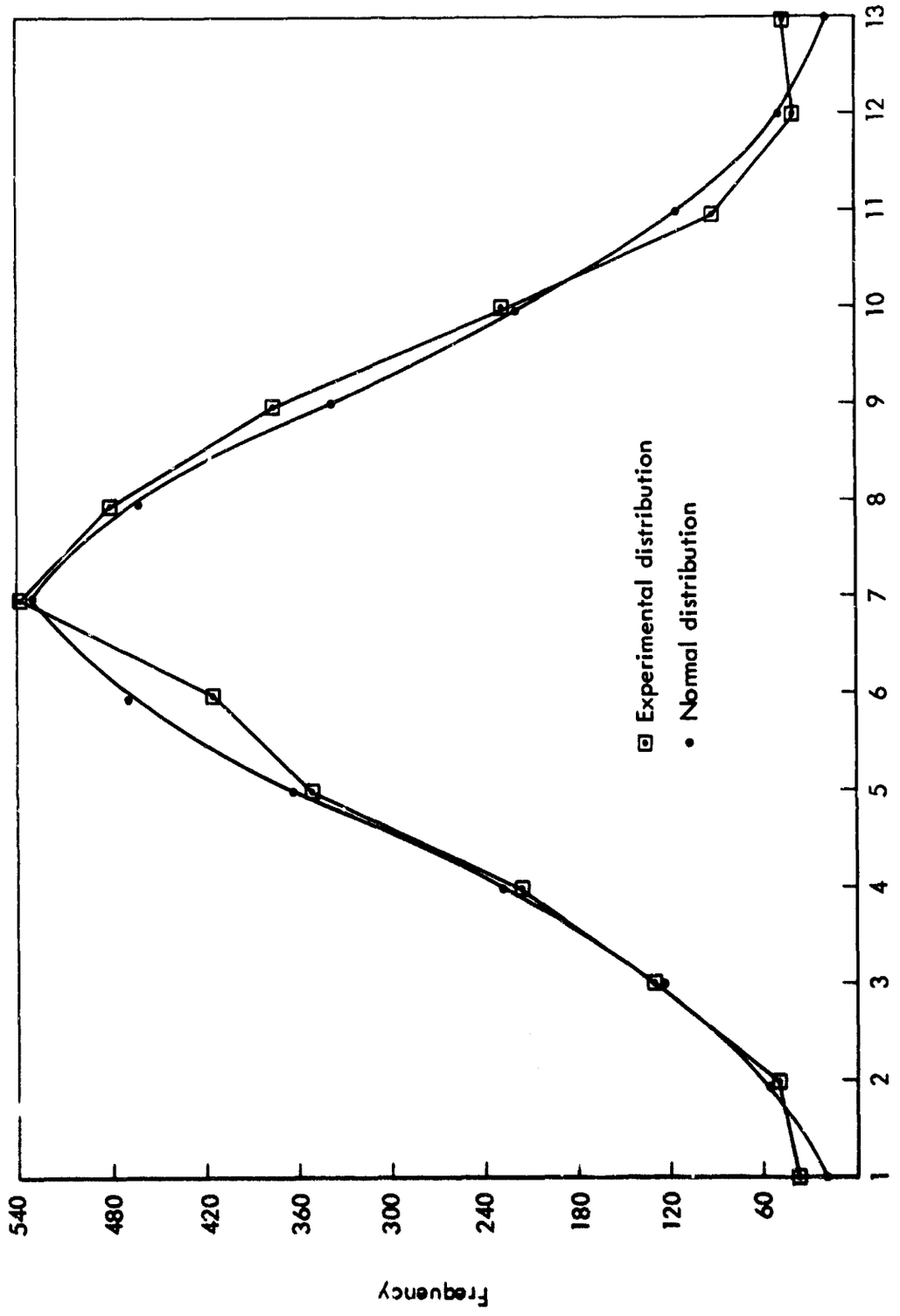


Fig. 3—Distribution of log responses ($\ln \frac{\text{answer}}{\text{true}}$) on round 1

IV. DISTRIBUTIONAL ESTIMATES

Based on some earlier experiments, we formulated the hypothesis that distributional estimates would lead to more accurate responses than point estimates. In addition, since the medians of the Low and High estimates could be expected to exhibit a narrower range than the lower and upper quartiles of point estimates, it was expected that the group making distributional estimates would show more changes and greater convergence than the group making point estimates. Neither expectation was fulfilled.

Table 7 displays the comparison between groups making point estimates and those making Low, Mid, and High estimates in terms of the average group error. There is a slight tendency for groups making Low, Mid, and High estimates to be more accurate, but the effect is not significant.

Table 7

AVERAGE ERROR FOR GROUPS MAKING POINT ESTIMATES AND GROUPS MAKING DISTRIBUTIONAL ESTIMATES

Group	Point Estimate	Group	Distributional Estimate
17	1.23	18	.97
19	.64	20	.71
21	.97	22	.97
23	1.05	24	.94
All groups	.96	All groups	.89

As feedback on round 2, the control groups received the three quartiles of the distribution of their (point) estimates on round 1. The comparison groups received the medians of the individual Low, Mid, and High estimates on round 1 as their feedback for round 2.

The expectation that the medians of the Low and High estimate would exhibit a narrower range than the lower and upper quartile of point estimates was fulfilled. For the 80 comparisons on round 1,* lower and upper quartiles of the point estimate distributions were farther apart on 66, closer on 10 and the same on 4 as compared with the medians of Low and High estimates. However, individuals making point estimates changed their answers in 58 percent of the opportunities, whereas individuals making distributional estimates changed their answers in only 55 percent of the opportunities.

A glance at Table 1 indicates that there is no significant difference in the number of changes between round 1 and round 2 for point and distributional estimate groups.

---

*The 40 questions were answered twice by the comparison groups and also twice by the experimental groups.

V. DISCUSSION

The present experiment was intended to furnish additional data concerning the properties of the Delphi procedures when applied to subject matter with "objective uncertainty." In general, the results are reassuring, but of course they do not deal with all the differences between laboratory and applied exercises. In particular, the subjects were college students and not mature experts (although the subjects were by no means naive with respect to the task).

Of most interest to applications is the definitely higher correlations between standard deviation and error, and group self-rating and error for the prediction questions. Considering these two indices as measures of the excellence of the answers to individual questions, their value appears to be enhanced in the short-range prediction situation. This is true, despite the fact that the average error on the prediction questions was about the same as the average error we obtained on the almanac questions. It seems likely that making short-range predictions was a more meaningful task for the subjects.

One of the interesting and suggestive results of the experiment is the similarity between the point

estimates and the Mid estimate for the distributional answers. One of the questions that has concerned us is the nature of the point estimates furnished by subjects. Since the subjects appear to be able to generate distributions, the point estimate is presumably related in some fairly direct way to these distributions. The similarity between the Mid estimates and the point estimates suggests that in a large number of cases what the subjects are reporting for point estimates are, in fact, the medians of their subjective probability distributions.

Appendix

QUESTIONNAIRE FOR ROUND 1

Listed below are the questions used in the prediction experiment, along with the true answers and the average group error on round 1. The group error is defined as the absolute value of the natural logarithm of the median divided by the true answer. The listed errors are the average of the errors for four groups. The top number to the right of each question is the true answer, the lower number is the average error on round 1.

Self-Rating		Question	Answer/Error
[ ]	1.	How many of the new Ford Maverick cars will have been sold by the end of September?	N/A*
[ ]	2.	How many national communist parties will be represented at the International Communist Congress in June?	75 nations 1.109
[ ]	3.	How many college students will be arrested as the result of disturbances on campus in the U.S. during the month of July?	N/A
[ ]	4.	Assuming a moon landing is successfully accomplished this summer, how many minutes will the first U.S. astronaut leaving the landing module spend on the surface of the moon?	135 min 1.651

*Not available.

Self-Rating	Question	Answer/Error
☐	5. How many murders will be reported in the U.S. during the month of August?	N/A
☐	6. How many inches of rain will fall in Hawaii during the three summer months June through August?	.97 inches 2.745
☐	7. What will be the total enrollment at UCLA for the summer quarter?	8,171 students .352
☐	8. How many U.S. Armed Forces personnel will be in South Korea on October 1, 1969?	59,878 personnel .800
☐	9. How many moderate earthquakes (registering more than 4.5 on the Richter Scale) will occur in California during the months of June, July, and August?	3 earthquakes .703
☐	10. How many people will be killed in the U.S. in motor vehicle accidents during the next July 4 weekend?	578 deaths .224
☐	11. What will be the value of the French franc (in U.S. dollars) on October 1?	$.18 1.432
☐	12. What will be the total vote cast for Samuel Yorty in the runoff of the Los Angeles city election on May 27?	447,030 votes 1.227
☐	13. How many games will the St. Louis Cardinals lose in the National League this season?	74 games 1.106
☐	14. How many Israeli aircraft (including helicopters) will be lost as a result of incidents in the Middle East during the months of June through September?	N/A

Self-Rating		Question	Answer/Error
☐	15.	How many marriage licenses will be issued in Los Angeles during the month of June?	7,421 licenses 1.089
☐	16.	How many color TV sets will be sold in the U.S. during the months of June through August?	N/A
☐	17.	What will be the highest temperature recorded during June in California?	114° F .068
☐	18.	How many out-of-state passenger cars will enter California at Needles during the month of July?	81,946 cars 2.077
☐	19.	How many rescues will be made on California State Beaches on July 4, 1969?	229 rescues .387
☐	20.	What will be the total number of votes cast in the French elections in June?	22,898,656 votes 1.131
☐	21.	How many cars will be stolen from Los Angeles International Airport parking lots during the month of July 1969?	N/A
☐	22.	How many Ph.D. degrees will be awarded by UCLA at the close of the present quarter?	287 degrees .283
☐	23.	How many votes will Pompidou receive in the French Presidential election on June 1?	10,151,804 votes .762
☐	24.	How many games will the Detroit Tigers lose in the American League this season?	71 games .871

Self-Rating		Question	Answer/Error
☐	25.	What will be the amount of the U.S. defense budget approved by Congress for fiscal year 1970? (July 1969 to July 1970)	$69.6 billion 1.180
☐	26.	How many babies will be born in Los Angeles during the month of August?	13,685 births 1.527
☐	27.	What will be the value of the British pound (in U.S. dollars) on October 1?	$2.39 .026
☐	28.	How many Soviet soldiers will be stationed in Czechoslovakia on September 1?	70,000 soldiers 1.395
☐	29.	On how many days during June and July will public peace negotiations take place in Paris?	8 days .733
☐	30.	How many members of the U.S. Armed Forces will be killed in action in South Vietnam during the months July through September?	1,876 deaths .525
☐	31.	What will be the average miles per hour of the winning automobile at the Indianapolis 500-mile race, May 30?	156.867 mph .033
☐	32.	How many new housing units will be started in the U.S. in July?	1,358,000 units 3.880
☐	33.	How many incidents of hijacking to Cuba will be recorded during June, July, and August 1969?	12 incidents .177
☐	34.	How many votes will be cast by absentee ballot in the Los Angeles city election for mayor on May 27?	18,026 votes 1.578

Self-Rating		Question	Answer/Error
☐	35.	What will be the number of reported suicides in the U.S. during the month of July?	N/A
☐	36.	How many U.S. aircraft (including helicopters) will de destroyed in Vietnam during June, July, and August?	263 aircraft 1.347
☐	37.	How many deaths from motor vehicle accidents will be reported for California the weekend of July 4?	56 deaths .669
☐	38.	How many heart transplants will be performed in the U.S. the months June through September?	N/A
☐	39.	In how many U.S. cities will major riots (estimated damage over one million dollars) occur during the summer months of June, July, and August?	5 cities .024
☐	40.	What will be the highest recorded temperature in the U.S. during the month of September?	118° F .017

REFERENCES

1. Dalkey, N., The Delphi Method: An Experimental Study of Group Opinion, The Rand Corporation, RM-5888-PR, June 1969.

2. Brown, B., S. Cochran, and N. Dalkey, The Delphi Method II: Structure of Experiments, The Rand Corporation, RM-5957-PR, June 1969.

3. Dalkey, N., B. Brown, and S. Cochran, The Delphi Method III: Use of Self-Ratings To Improve Group Estimates, The Rand Corporation, RM-6115-PR, November 1969.

4. McGregor, D., "The Major Determinants of the Prediction of Social Events," J. Abn. Psychol., 33, 1938, pp. 179-204.

5. Cantril, H., "The Prediction of Social Events," J. Abn. Psychol., 33, 1938, pp. 364-389.

6. Kaplan, A., A. L. Skogstad, and M. A. Girshick, "The Prediction of Social and Technological Events," Public Opn. Quart., Spring 1950, pp. 93-110.

7. Campbell, R. M., "A Methodological Study of the Utilization of Experts in Business Forecasting," Unpublished Ph.D. Dissertation, University of California, Los Angeles, 1966.