AN EXPERIMENTAL APPLICATION OF THE DELPHI METHOD TO THE USE OF EXPERTS *†

NORMAN DALKEY AND OLAF HELMER

The RAND Corporation, Santa Monica, California

This paper gives an account of an experiment in the use of the so-called DELPHI method, which was devised in order to obtain the most reliable opinion consensus of a group of experts by subjecting them to a series of questionnaires in depth interspersed with controlled opinion feedback.

1. Introduction

"Project DELPHI" is the name for a study of the use of expert opinion that has been intermittently conducted at The RAND Corporation. The technique employed is called the DELPHI method. Its object is to obtain the most reliable consensus of opinion of a group of experts. It attempts to achieve this by a series of intensive questionnaires interspersed with controlled opinion feedback.

The present paper gives an account of an experiment conducted about ten years ago. The content of the paper has, for security reasons, only now been released for open publication.

The experiment was designed to apply expert opinion to the selection, from the viewpoint of a Soviet strategic planner, of an optimal U. S. industrial target system and to the estimation of the number of A-bombs required to reduce the munitions output by a prescribed amount.

The technique employed involves the repeated individual questioning of the experts (by interview or questionnaire) and avoids direct confrontation of the experts with one another.

The questions, which are all centered around some central problem (in our present case, an estimate of bombing requirements), are designed to bring out the respondent's reasoning that went into his reply to the primary question, the factors he considers relevant to the problem, his own estimate of these factors, and information as to the kind of data that he feels would enable him to arrive at a better appraisal of these factors and, thereby, at a more confident answer to the primary question. The information fed to the experts between rounds of questioning is generally of two kinds: It consists either of available data previously requested by some one of the experts (e.g., output statistics for steel mills), or of factors and considerations suggested as potentially relevant by one or another respondent (e.g., the extent to which power transmission facilities

* Received July 1962.

† This research was sponsored by the United States Air Force under Project RAND—Contract No. AF 49(638)-700—monitored by the Directorate of Development Planning, Deputy Chief of Staff, Research and Technology, Hq USAF. Views or conclusions contained in this paper should not be interpreted as representing the official opinion or policy of the United States Air Force.

permit reallocation of electric power). With respect to the latter type of information, an attempt was made (not always successfully) to conceal the actual opinion of other respondents and merely to present the factor for consideration without introducing unnecessary bias.

This mode of controlled interaction among the respondents represents a deliberate attempt to avoid the disadvantages associated with more conventional uses of experts, such as round-table discussions or other milder forms of confrontation with opposing views. The method employed in the experiment appears to be more conducive to independent thought on the part of the experts and to aid them in the gradual formation of a considered opinion. Direct confrontation, on the other hand, all too often induces the hasty formulation of preconceived notions, an inclination to close one's mind to novel ideas, a tendency to defend a stand once taken or, alternatively and sometimes alternately, a predisposition to be swayed by persuasively stated opinions of others.

By systematically exploring the factors which influence the judgment of the individual expert, it becomes possible to correct any misconceptions that he may have harbored regarding empirical factors or theoretical assumptions underlying those factors, and to draw his attention to other factors which he may have overlooked in his first analysis of the situation. Needless to say, considerable discretion has to be exercised by the experimenters in any efforts designed to make an expert change his mind, in order to obtain results which are free of any bias on the experimenters' part. A device for helping to assure this is to feed in only such data as have been asked for by at least one respondent and are obtainable from reliable sources, and to suggest only such theoretical assumptions as seem to represent a consensus of a majority of respondents.

If the purpose of the experiment is the estimation of a numerical quantity (in our case the number of bombs required to do a certain job), it may be expected that, even if the views expressed initially are widely divergent, the individual estimates will show a tendency to converge as the experiment continues. This is almost inevitable in view of the progressively more penetrating analysis of the problem, achieved partly by means of the procedural feedback described above.

On the other hand, it cannot even ideally be expected that the final responses will coincide, since the uncertainties of the future call for intuitive probability estimates on the part of each respondent. To some extent this terminal disagreement can sometimes be decreased by applying justifiable corrections to the final answers. Such corrections are in fact an integral part of the procedure; they must, however, be based on a careful analysis of the responses, taking into account whatever can be learned regarding (i) a consensus as to basic assumptions, (ii) the sensitivity of the individuals' responses to changes in these basic assumptions, and (iii) their estimates of functional dependencies rather than mere point estimates. Essentially, the resulting corrections amount to a replacement of the individual expert's estimates concerning some of the components of the main problem by a consensus of estimates by all the experts. For example, in the experiment of this report, the problem of estimating the total number of bombs

was factorable into that of determining, for each of several industries, what percentage of each industry must be destroyed and the average number of bombs per plant needed to do so. Each respondent made estimates of both these quantities. For the first, which involved the selection of the industries to be bombed, the choices made were too divergent to permit the taking of a concensus. The second estimate, however, was a perfect example of a case wherein a consensus would seem to yield more reliable results; accordingly we corrected the respondents' final answers by replacing their own numbers for bombs per plant by the median of all seven estimates. Table 1 reflects the resulting trend. It will be noted that the ratio between the largest and smallest response, which was initially 100 to 1, dropped finally to about 3 to 1, and upon correction was ultimately reduced to only about 2 to 1.

2. Description of the Experiment

The experiment was conducted with a panel of seven experts. Four of these were economists, one was a physical-vulnerability specialist, one a systems analyst, and one an electronics engineer.

There were altogether five questionnaires, submitted at approximately weekly intervals. The first and third of these were followed up by interviews with each of the respondents. We present here a condensed log of the proceedings.

Questionnaire 1

This is part of a continuing study to arrive at improved methods of making use of the opinions of experts regarding uncertain events.

The particular problem to be studied in this experiment is concerned with the effects of strategic bombing of industrial targets in the U. S. . . .

Please do not discuss this study with others while this experiment is in progress, especially not with the other subject experts. You are at liberty, though, to consult whatever data you feel might help you in forming an opinion.

The problem with which we will be concerned is the following:

Let us assume that a war between the U. S. and the S. U. breaks out on 1 July 1953. Assume also that the rate of our total military production (defined as munitions output plus investment) at that time is 100 billion dollars and that, on the assumption of no damage to our industry, under mobilization it would rise to 150 billion dollars by 1 July 1954 and to 200 billion dollars by 1 July 1955, resulting in a cumulative production over that two-year period of 300 billion dollars. Now assume further that the enemy during the first month of the war (and only during that period) carries out a strategic A-bombing campaign against

TABLE 1

Estimated Number of Bombs

Response	Answer		
	Smallest	Median	Largest
Initial	50	200	5000
Final	159	255	494
Corrected final	167	276	360

U. S. industrial targets, employing 20-KT bombs. Within each industry selected by the enemy for bombardment, assume that the bombs delivered *on target* succeed in hitting always the most important targets in that industry. What is the least number of bombs that will have to be delivered on target for which you would estimate the chances to be even that the cumulative munitions output (exclusive of investment) during the two-year period under consideration would be held to no more than one quarter of what it otherwise would have been?

This question will be referred to below as the "primary question."

In a follow-up interview to the first questionnaire, each respondent was asked to provide a breakdown by industries of the number of bombs specified by him and to reproduce some of the reasoning that went into his estimate. He was further asked to estimate the number of bombs needed to do the job with 10% and with 90% confidence of success, and to indicate what kind of data he would consider most helpful in arriving at a better appraisal.

The total numbers of bombs were estimated as shown in Table 2. (The respondents have been ordered according to the numerical order of their corrected final responses.) The choices of target systems were quite distinct, the only common feature being the inclusion of the steel industry in each.

Questionnaire 2

As the result of the first round of interviews, it appears that the problem for which we are trying with your help to arrive at an estimated answer breaks down in the following manner.

There seem to be four major items to be taken into consideration, namely:

A. The vulnerability of various potential target systems,

B. The recuperability of various industries and combinations of industries,

C. The expected initial stockpiles and inventories, and

D. Complementarities among industries.

Taking all these into account, we have to

(1) Determine the optimal target system for reducing munitions output . . . to one fourth . . . ,

(2) Estimate for this target system the minimum number of bombs on target required to create 50% confidence of accomplishing that aim.

. . . We would like to establish the background material consisting of A, B, C, D more firmly. . . . With regard to A and B, the interviews have suggested the following tentative breakdown of possibly relevant factors: . . . (here, two lists of factors were given, related to vulnerability and recuperability respectively).

TABLE 2

Confidence-of-Destruction Estimates

Response	Respondent						
	1	2	3	4	5	6	7
Primary (50% confidence)	125	50	150	300	200	1000	5000
10% and 90% confidence	75–200	25–150	100–175	250–800	70–500	—	2500–10000

Question 1. Does the preceding breakdown of the problem agree with your intuitive approach to a solution? If not, explain in detail; in particular, are there major items in addition to A, B, C, D which should be taken into consideration?
Question 2. What additional factors, if any, do you consider relevant to the problem of vulnerability? Which of the factors listed do you consider irrelevant?
Question 3. What additional factors, if any, do you consider relevant to the problem of recuperability? Which of the factors listed do you consider irrelevant?
Question 4. What factors should be taken into account for our problem in assessing the size and role of initial stockpiles?
Question 5. What factors should be taken into account in our problem as regards determining complementarities among industries?
Question 6. Are there any general comments which you wish to make?

The response consisted of a large volume of informal comments. The most significant among these pointed out the difference between economic and physical vulnerability, the influence of the planned munitions product mix, the importance of substitutabilities of plants and materials, and the dependence of the lead times of components on the damage done to the industries producing these. Only some of this material was reflected in the later phases of the experiment.

Questionnaire 3

You are being asked today for a reconsideration of your original estimate. The question is restated below, together with a few explanatory comments. We are also listing a few facts and estimates, which you may wish to take into consideration in forming a revised opinion.

Restatement of primary question: . . .

Comments: (Clarification of the terms "industrial target" and "bomb on target" and of some assumptions to be made by the respondents in forming their estimates.)

Data on U. S. economy:

(a) Number of plants presently (i.e., in 1951) accounting for indicated percentages of various industries' outputs: . . .
(b) Percentages of metals output going into munitions, consumption, and gross investment: . . .
(c) Percentages of munitions value constituted by value of metals inputs: . . .

Data on structural vulnerability:

(d) Examples of damage with 20-KT bomb obtained from Japanese bombings: . . .
(e) Vulnerability estimates for specific industries: . . .

Question 1. What is your revised answer to the primary question of Questionnaire 1?
Question 2. Do you consider the tabulation of industrial plants given under (a) above reasonably correct? (If not, please specify.)
Question 3. What changes, if any, in that tabulation do you expect by mid-1953?
Question 4. Do you roughly agree with the estimates of physical vulnerability expressed under (e) above? (If not, please specify.)
Question 5. For the following industries, how would you allot the minimum number of bombs on target called for in the primary question?

Steel	Heavy steel fabrication
Petroleum refining	Machine tools
Aluminum	Electron tubes
Copper	Aviation fuel
Power	Anti-friction bearings
A-bombs	Other industries
Aircraft engines	

The follow-up interviews served to clarify a few uncertainties and produced

TABLE 3

Revised Estimates

Response	Respondent						
	1	2	3	4	5	6	7
To question 1	158	89	200	250	256	800	450
To interview	158	106	184	250	256	525	450

further minor revisions. The responses to the primary question are given in Table 3.

Questionnaire 4

. . . The principal purpose of this questionnaire is again to obtain from you revised answers as to the numbers of bombs allotted to various industries—the revisions to be based upon consideration of the information supplied below as well as any further thought you may have given to the matter. In addition you will be asked to make certain recuperation forecasts . . . and to make a critical comparison between your own bombing schedule and two others to be specified below. (These two bombing schedules, labeled 'A' and 'B' below, had been obtained as follows: the seven bombing schedules obtained previously were roughly ordered cyclically in such a manner that each was as similar as possible to its two neighboring schedules; according to the numbering of the respondents, the following cyclical order was obtained: 1234765. Each expert was then confronted with the bombing schedules of his two neighbors in this ordering, e.g., 1 with 2 and 5.)

Additional information on the target system: . . .

Information on stockpiles: . . .

Information on the power system: . . .

Information on the uses of steel: . . .

Information on the bombing of Europe in World War II: . . .

Information on Japanese recuperation: . . .

Question 1. In the last column of the following table, indicate your revised bombing schedule:

Plants producing			Industry	Bombing schedules			
50%	75%	100%		A	B	Your former figures	Your revised figures
17	37	215	Steel				
25	85	437	Petroleum				
2	5	12	Aluminum				
4	6	12	Copper				
125	325	3700	Power				
		7	A-bombs				
4	8	21	A/C engines				
3	6	9	Steel fabric.				
20	55	316	Mach. tools				
8	17	53	Electron tubes				
		14	Aviation fuel				
3	6	19	Ball bearings				
			Other				
			Total:				

TABLE 4

Revised Total Number of Bombs

	Respondents						
	1	2	3	4	5	6	7
No. of bombs	166	153	200	250	300	332	500

Questions 2 and 3. Draw graphs indicating the estimated progress of steel and of munitions output recuperation after bombing according to your revised schedule. (Coordinate systems were provided.)

Question 4. Compare your proposed bombing schedule with that given under A above. While you estimate your own schedule to reduce munitions output over two years to 25%, a reduction to how many percent do you expect from Schedule A?____Briefly, why is your proposal superior to Schedule A?

Question 5. The same for Schedule B.

The revised total numbers of bombs, obtained in response to Question 1, are shown in Table 4. The comparison with other bombing schedules brought out a number of interesting points, the most important of which were brought to the group's attention in the subsequent questionnaire.

Questionnaire 5

In this final questionnaire you will have a last opportunity to revise once more your earlier estimates if you should feel so inclined. The possibility of such a further revision suggests itself in view of (i) a piece of information, given below, on World War II munitions expenditures . . . , (ii) certain considerations emphasized by the respondents themselves in their replies to the preceding questionnaire, and (iii) a possible discrepancy, in some cases, between the prescribed bombing goal and the accomplished munitions-output reduction as indicated by your graph (response to Question 3 of the preceding questionnaire).

Attached you will find your previous response sheet. On the graph which represents the answer to Question 3, the munitions output under normal wartime expansion without bombing has been indicated by a dotted line; this corresponds to the assumptions stated in our original formulation of the problem in the first questionnaire. Also indicated, in red, is the approximate munitions output, in percent of the normal output, computed from your graph. If this number differs substantially from 25, this may of course be due to your having drawn the graph free-hand, or to a difference of opinion as to the amount of munitions output under normal expansion. If, however, the difference is due to your having attempted to reduce munitions output to 25% of what it would have been without expansion, you have in fact overbombed and may wish to revise your estimates accordingly.

Distribution of munitions expenditures in 1944: . . .

Considerations emphasized by respondents in preceding questionnaire:

1. The effect of industrial expansion on the number of plants producing 75%.
2. Use of the principle of equal marginal utility in assigning bombs to industries.
3. Observation of intra-industry complementarities (e.g., alumina and aluminum).
4. Observation of inter-industry complementarities (e.g., aluminum and aircraft engines).
5. The possibility that concentrating the attack allows concentration of the recuperation effort.

Question: Please fill in the blank columns in the following table (here the table of the

preceding questionnaire was reproduced, with the left-hand half as before, and the right-hand half replaced by columns with the following headings):

Estimate number of plants in mid-1953 producing		If this industry were to be bombed, estimate no. of bombs on target needed to destroy		Give your finally revised bombing schedule
75%	100%	75%	100%	

The respondents' final bombing estimates are shown in Table 5.

3. Correction of The Final Responses

As indicated at the end of Section 1, the final responses given above are capable of correction on the basis of replacing some of the individual component estimates by a consensus of estimates. Whenever this was done, the median of the responses was taken as the consensus. Our procedure was, first of all, to tabulate for each of the industries considered the medians of (i) the expected numbers of plants respectively producing 50%, 75%, and 100% of the total output in mid-1953, and (ii) the number of plants requiring two rather than one bomb on target for destruction.

We then listed (iii) the percent of damage to each industry that each expert intended as indicated from the figure *he* gave for the numbers of plants in mid-1953, the number of bombs needed to destroy 75% and 100%, and of bombs to be allocated to each industry, and (iv) the corresponding numbers of bombs as computed with the aid of the tabulation obtained under (iii). The total of these latter numbers, for each respondent, was taken as his corrected final answer, as shown in Table 6. The five successive sets of responses, plus the corrected totals, are shown in Fig. 1, which brings out very clearly the gradual convergence of

TABLE 5

Final Bombing Estimates

	Respondents						
	1	2	3	4	5	6	7
No. of bombs	177	159	200	255	312	314	494

TABLE 6

Corrected Final Estimates

Estimate	Respondent						
	1	2	3	4	5	6	7
Final	177	159	200	255	312	314	494
Corrected final	167	179	206	276	292	349	360

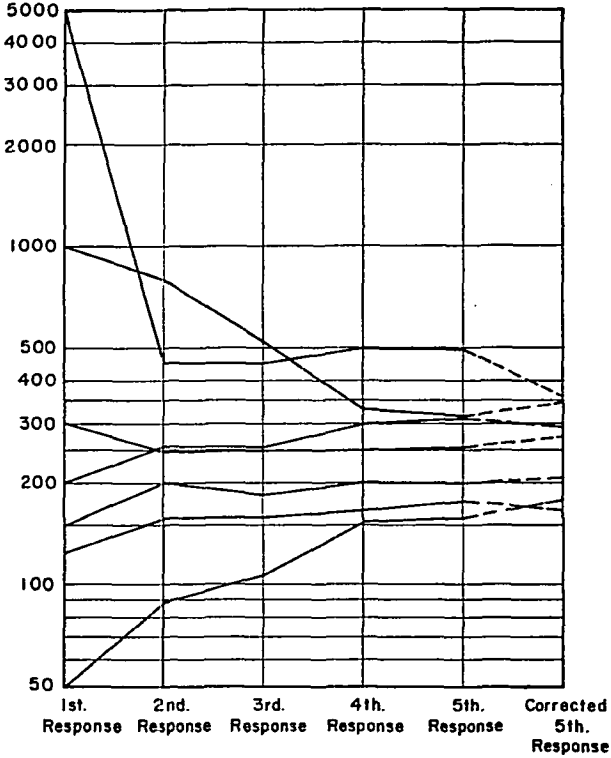

FIG. 1. Successive estimates of bomb requirements

the answers. The smallest answer is seen to have increased monotonically from 50 to 167, while the largest decreased from 5000 to 360. The median advanced slightly from 200 to 276. There are strong indications that, if the experiment had been continued through a few more rounds of questionnaires, the median would have shown a downward trend and the ratio of the largest to the smallest answer would have shrunk to 2 or less.

4. Critique of the Experimental Procedure

The following points represent a summary of the items for which the experimenters are conscious of the need for apology:

(i) The experts' responses were not strictly independent. Although the respondents on the whole complied with the initial cautioning not to discuss the experiment with one another while it was in progress, their other working assignments on related subjects required some contact among several of them.

(ii) At least one of the respondents was also used by the experimenters as a consultant on one aspect of the subject matter of the experiment.

(iii) Some "leading" by the experimenters inevitably resulted from the selection of the information supplied by the experts.

(iv) The experiment was terminated prematurely, before it was possible to give as much emphasis to complementarities and recuperation as had been given, say, to vulnerability.

(v) The comparison of two "neighboring" bombing strategies, called for in Questionnaire 4, was a shortcut necessitated by the pressure of time; it was intended to throw some light on the sensitivity of the bombing figures given by each respondent. This purpose would have been served better by a less biased but more time-consuming approach.

(vi) Vague questions inviting general critical comment, such as were presented in Questionnaire 2, produce literary outpourings of little value for the analysis and should either be omitted or replaced by an interview.

(vii) The correction of the final responses, carried out above in view of certain median considerations, may seem plausible but nevertheless should be given a firmer theoretical foundation.

The authors are convinced that most of these shortcomings can gradually be eliminated by further experimentation in this area. Even as it stands, the method exemplified by the experiment reported here is highly conducive to producing preliminary insights into the subject matter at hand on which a more effective research program may be based, even though the predictions obtained in the form of an opinion consensus may be lacking in reliability. But with further progress in the methodology of the efficient use of experts, it may be hoped that a carefully contrived opinion consensus would often turn out to be an acceptable substitute for direct empirical evidence when the latter is unavailable.[1]

[1] For a further discussion of the methodology of the use of expert opinion, see *On the Epistemology of the Inexact Sciences*, by O. Helmer and N. Rescher, The RAND Corporation, Report R-353; also published in *Management Science*, Vol. 6, No. 1, 1959. For reports on two experiments in the use of expert opinions regarding the qualitative ranking of given alternatives, see Thrall, Coombs and Caldwell, "Linear Model for Evaluating Comples Systems," *Naval Research Logistics Quarterly*, Vol. 5 (1958), and Calawell, Coombs, Schoeffler and Thrall, "A Model for Evaluating the Output of Intelligence Systems," *Naval Research Logistics Quarterly*, Vol. 8 (1961). An earlier experiment in the use of an expert consensus for predictive purposes was reported by Kaplan, Skogstad and Girshick in their paper, "The Prediction of Social and Technological Events," *Public Opinion Quarterly*, Spring 1950.