SYSTEMATIC USE OF EXPERT OPINIONS

Olaf Helmer

November 1967

P-3721

SYSTEMATIC USE OF EXPERT OPINIONS

Olaf Helmer*

The RAND Corporation, Santa Monica, California

When we talk about systems analysis, we tend to think of an orderly analytical process, usually involving a mathematical model, by which the available alternatives in a decision-making situation are systematically compared with regard to their costs and benefits. There are many cases, however, where decisions have to be based, not on the results of theoretical analysis, but on the intuitive judgment of whatever experts on the particular issue are at hand. This may be so simply because no satisfactory theory has as yet been devised; it may also be because judgment must be sought as a matter of principle, because the issue under consideration may involve moral in addition to factual aspects, and thus preferences in addition to data.

Once we have recognized that reliance on expert opinion at times is unavoidable, we need to give some thought to the question of how to obtain such opinions most efficiently and most reliably.

The method I want to report on very briefly is a systematic procedure for obtaining the opinions and, if possible,

---

*Any views expressed in this paper are those of the author. They should not be interpreted as reflecting the views of The RAND Corporation or the official opinion or policy of any of its governmental or private research sponsors. Papers are reproduced by The RAND Corporation as a courtesy to members of its staff.

This paper was prepared for presentation at the AFAG Board Meeting, November 1967.

a consensus from a panel of experts on a particular issue. It was first designed many years ago here at RAND* and has since aroused a good deal of interest both in this country and abroad and has been applied in universities as well as in government and industry. This method has become known, somewhat inadvertently, as the Delphi Technique. Like the ancient oracle, it has been employed to obtain opinions on what the future holds. But its scope is more general, in that it can be used in any context where it is appropriate to seek a consensus of opinions among experts on a particular subject.

It may perhaps be easiest to describe the basic principle involved in this technique by reference to a particular example. I use as an illustration one based more or less on an actual case that arose in the course of an inquiry into the future of automation. Each member of a panel of experts on this subject was asked, among other things, to estimate the year when a machine would become available that would comprehend standard IQ tests and score above 150. The chart (p. 3) illustrates the sort of thing that happened, and I think it is to some extent selfexplanatory. The procedure consists in four rounds of questionnaires, asking for successive refinements in the estimated answer to the given question. In each round after the first, some information feedback on the outcome of the previous round is provided. In particular, the respondents are informed of the previous response distribution in terms of its median and its so-called interquartile range (that is, the interval containing the middle 50 percent of the responses).

A further characteristic of the procedure is that it is designed to generate what may be thought of as an anonymous debate. Thus, in the second round, if a respondent's

*An Experimental Application of the Delphi Method to the Use of Experts, RM-727, by Norman Dalkey and Olaf Helmer.

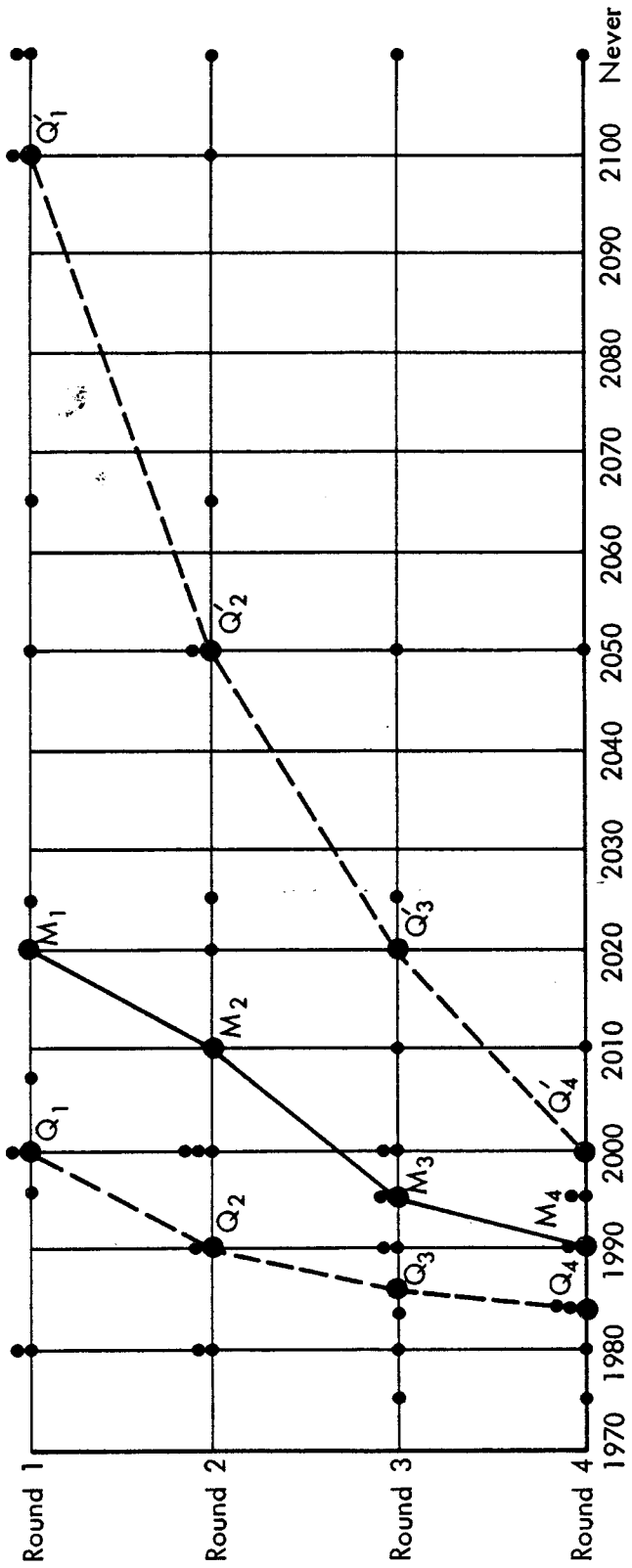

Chart 1—When will a machine become available that can comprehend standard IQ tests and score above 150 ?

revised answer falls outside the interquartile range, he is required to state briefly why he thought that the event would occur that much earlier (or that much later) than the majority seemed to think. The effect of thus placing the onus of having to justify relatively extreme opinions on the respondents typically has the effect shown in the illustration: Those without strong convictions tended to move their estimates closer to the median, while those who felt they had a good argument for a deviant opinion tended to retain their original estimate and defend it.

In the third round the respondents were given a concise summary of such reasons in support of extreme positions, and asked to base any revision of their estimates on due consideration of these reasons, giving them whatever weight they thought was justified. Moreover, if a respondent's revised answer fell outside the (new) interquartile range, he was now required to state why he was unpersuaded by the opposing argument.

In the fourth round, similarly, these counter-arguments were fed back to the entire group, giving rise to one last chance for reestimating the date of occurrence.

As is seen from the chart, in this example the procedure--presumably under the influence of convincing arguments--caused the median (which originally had been the year 2020) to move to a much earlier date and the interquartile range (originally extending from 2000 to 2100) to shrink considerably. It seems fair to accept the median of the final responses, that is the year 1990, as representing the nearest thing to a group consensus.

A convergence of opinions has been observed in the majority of cases where the Delphi approach has been used. Sometimes, when no convergence toward a narrow interval of values takes place, opinions are seen to polarize around two distinct values, so that two schools of thought regarding a particular issue seem to emerge; this may be an

indication that opinions are based on different sets of data or on different interpretations of the same data. In such cases, it is conceivable that a continuation of the Delphi process through several more rounds of anonymous debate might eventually track down and eliminate the basic cause of disagreement and thus lead to a true consensus. But even if this does not happen, it should be realized that the Delphi technique serves to crystallize the reasoning process that might lead to one or several positions on an issue and thus helps to clarify the issue even in the absence of a group consensus.

The method which I described should still be considered very much in the experimental stage. Little is known about its validity, in the sense of yielding more reliable results than rival methods that are more traditional. One reason for our relative ignorance in this regard is that experts are a scarce commodity, who cannot easily be persuaded to give up valuable time to act as guinea pigs in an experiment. Some limited experimentation with students as subjects has been carried out, but there is always some doubt as to whether favorable results obtained under such conditions are transferable to the case of high-level specialists for whom the Delphi technique is intended. Therefore the validity of the procedure may be considered established only in an intuitive sense, in that the participants themselves in a Delphi inquiry generally appear satisfied that the method is both fair and efficient in extracting whatever information the group collectively possesses.

Experiments continue in an effort to invent further improvements in the technique. A few such possibilities are listed in this chart:

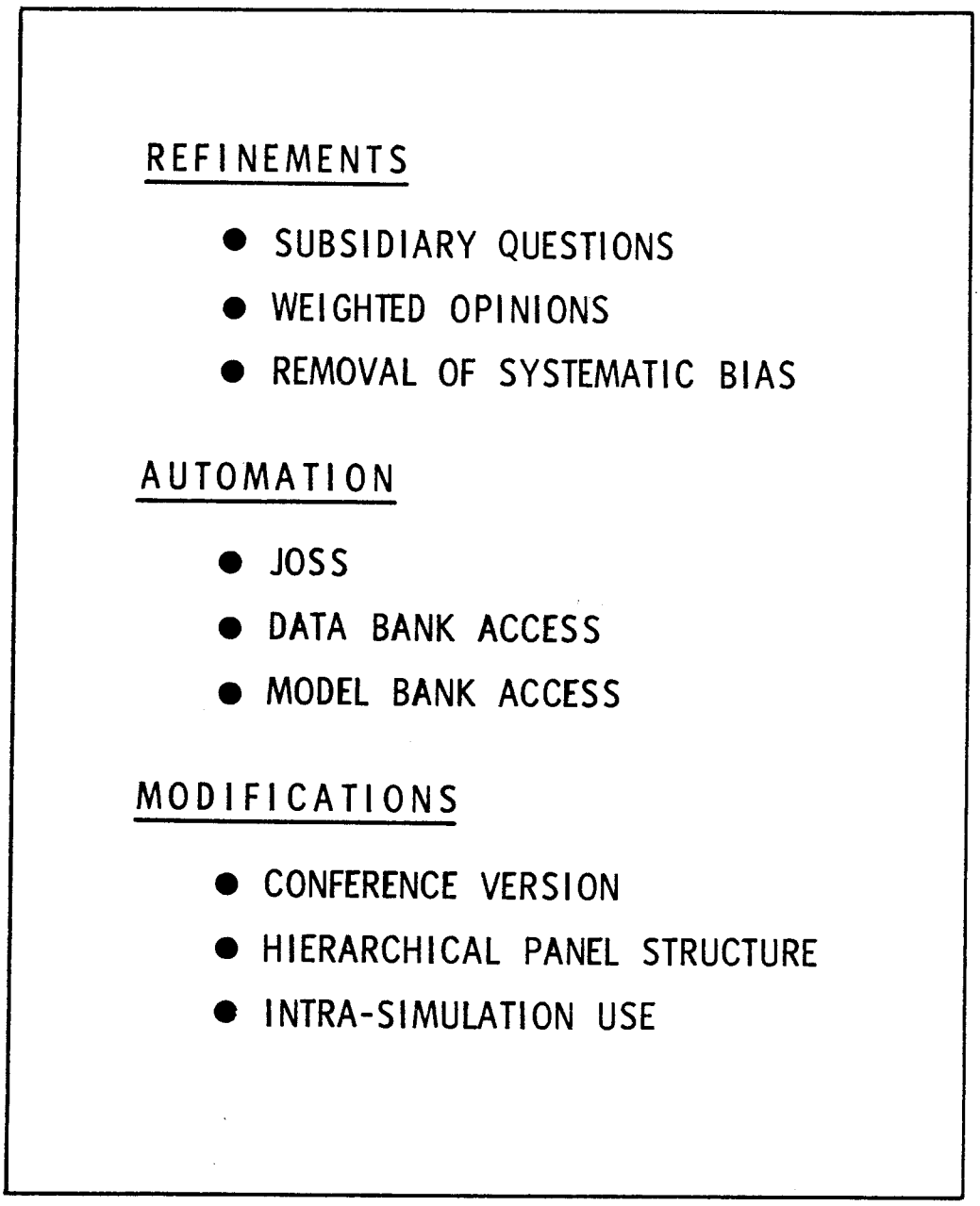

Chart 2—Possible further developments in the Delphi technique

By "subsidiary questions" I refer to the possibility of inviting the respondents, as part of the first-round questionnaire, to state subsidiary questions the answers to which might help them arrive at a more reliable answer to the primary question. If such questions refer, say, to existing statistical data, these can then be supplied along with the second questionnaire. Otherwise they may give rise to a subsidiary Delphi inquiry among the panel of participants. For example, in the illustration I gave, a respondent may well want to know the other experts' opinions as to when they expect computer speed to go up by another factor of 1000 before he would want to estimate the date of arrival of the 150-IQ computer.

A refinement which has already been successfully tested is that of attributing differential weights to the opinions of different experts. Clearly, if it were easy to measure the relative trustworthiness of different experts, we would give greatest weight to the opinions of those who are most trustworthy. In the absence of objective measurements to this effect, we have examined the possibility of relying instead on the experts' subjective self-appraisal of their own competence, and found this quite promising. For example, this idea was used not so long ago when twenty faculty members of UCLA's Graduate School of Business Administration made forecasts for the following year of certain economic and business indices. The procedure was as follows: In addition to going through four rounds of Delphi arguments, the respondents were asked to rank the indices under consideration with respect to the competence they felt in estimating their future values. Then, instead of using for each index the median of all twenty final responses as the group consensus, and thus as the group's prediction for the following year, only the responses of those individuals were taken who had ranked their competence regarding that index relatively most highly,

and the median of just these forecasts was then used as the group consensus. It subsequently turned out that this select median, compared to the median of all responses, was closer to the true value in two thirds of the cases.

With regard to the removal of systematic bias, it has recently been suggested that there may be a negative correlation between a forecaster's age and the date of occurrence of an event predicted by him. This, and other similar correlations, ought to be looked at, and if found true we would be in a position to correct for such systematic bias.

As for automating part of the Delphi process, I may mention that we have begun to experiment with the use of JOSS on-line computer consoles. By having each participating expert in a Delphi inquiry give his responses on a JOSS console, we can process a group's responses automatically and immediately feed back the information and instructions that make up the next questionnaire. Thus an effort which might otherwise take weeks or months can be carried through in an hour or less. Moreover, once the interaction among the respondents is via machine, it would be relatively easy to enrich the process by providing on demand automated access to existing data banks and eventually even to banks of mathematical models that might aid the expert in the analysis of the situation under consideration. Once this process has been perfected, it is easy to imagine that for important decisions simultaneous consultation with experts in geographically distinct locations via a nation-wide or even a world-wide computer network may become a matter of routine.

To return from these rather grandiose future applications involving panelists remote from one another, let me go in the opposite direction and say a few words about modifications of the Delphi technique that make it of use in situations where face-to-face confrontation of the

experts is unavoidable or even expedient.

Consider the simplest and most common case of a small committee having to arrive at a joint determination of some quantity, say, an estimate of next year's GNP, or a decision on how much to invest in a new military procurement item. Round-table discussions for such purposes have certain psychological drawbacks, in that the outcome is apt to be a compromise between divergent views, arrived at all too often under the undue influence of certain factors inherent in the face-to-face situation. These may include such things as the purely specious persuasion of others by the member with the greatest supposed authority or even merely the loudest voice, an unwillingness to abandon publicly expressed opinions, and the bandwagon effect of majority opinion. These psychological shortcomings gave rise to the development of Delphi in the first place, the contention being that the price paid for the anonymity of the Delphi procedure in terms of reduced ease of communication is well worth it if thereby the substance of what is being communicated is improved. However, some of the inadequacies of the face-to-face debate can be removed without going all the way to the full anonymity guaranteed by the Delphi procedure.

A simplified version of Delphi, which has been found useful in such situations, goes something like this: Have each member of the panel independently write down his own estimate, reveal the set of estimates but without identifying which was made by whom, and debate openly the pros and cons of various estimates; then have each person once more independently write down his own (possibly revised) estimate, and accept the median of these as the group's decision.

From this simplest case, where the outcome is the determination of a single number, one can proceed to more sophisticated ones, such as simulated planning exercises or, for that matter, actual planning. A conceptually

simple example, but one already requiring a much more elaborate procedure than that involved in determining a single number, is that of allocating a given budget among a number of projects. The difficulty of course is that the allocation to a particular project cannot be handled in isolation but that an integrated view of the entire menu of projects and their interrelations must be taken. Even more difficult is the problem involved in program-budgeting, where the group of experts entrusted with the planning task, prior to the budget allocation process, must invent possible measures, compose coherent programs of such measures, estimate their costs, decide how to assess their benefits, and having done so estimate their benefits. This entire process, so important in governmental decision-making, is replete with occasions for expert judgment and for the reconciliation of judgments at variance with one another. It invites experimentation with such methods as the Delphi technique in order to achieve the necessary systematic use of expertise in this area.

An example among the many possibilities that might be experimented with in this regard is the use of a hierarchical panel structure. This would be a Delphi procedure where responses are collected, not from individuals, but from sub-panels of experts, and where each such sub-panel in turn arrives at its group response by a Delphi opinion survey among its individuals.

As for other experimentation with the systematic use of group expertise, some has already been carried out here at RAND. In the military area, I can refer to the SAFE game, which is concerned with the simulation of military procurement decisions. More recently, we carried out a simple experiment in the societal area with the help of some of our graduate summer students. A group of them was asked to engage in a simulated planning exercise concerned with anti-crime measures. They were, in fact, given a

budget of so many billion dollars and asked to come up with a program of measures, and an allocation of the given budget among these measures, that would reduce the rate of crimes of violence in this country. The result of this little experiment was quite encouraging. Not only did the students produce a respectable menu of anti-crime measures, but the frequent use of miniature Delphi routines (of the kind I discussed) seemed to be quite helpful in program selection and in the budget allocation process, and thus promising enough to suggest further experimentation along this line.

Let me say by way of summary that progress in this general area of systematizing the use of judgment, though slow, is steady. This is encouraging. For regardless of whether we are concerned with military or with societal matters, we must realize that we are living in an era where the extremely rapid advances in hardware technology are accelerating the pace of change in our world and where, in order to cope with the consequences of these changes, every effort must be made to close the gap between hard and soft technology. And it is here that the systematic use of intuitive judgment plays a vital role. For, the lagging progress in the soft sciences is merely an indication that much of our understanding of what goes on is still on an intuitive level rather than of a form where it can be articulated and abstracted into theory; it does not mean that such understanding is absent. To tap this knowledge, because it is not neatly formalized but distributed in the minds of many people, it is necessary to develop methods, of which the Delphi technique is one, for collecting th opinions of individual experts and combining them into judgments that have operational utility to policy maker.