DELPHI

Norman C. Dalkey

October 1967

P–3704

12

DELPHI

Norman C. Dalkey[*]

The RAND Corporation, Santa Monica, California

1. INTRODUCTION

Delphi is the name of a set of procedures for eliciting
and refining the opinions of a group of people. In practice,
the procedures would be used with a group of experts or
especially knowledgeable individuals.

The significance of the Delphi technique should be
examined in the context of what I call the Advice Community.
Both industry and government are served by a large group of
consultants who purvey information, predictions, and analy-
ses to aid the formation of policy and making decisions.
The community is a highly miscellaneous assortment of "in-
house" advisors, and external consultants from academia,
other industries, nonprofit corporations, and, of course,
any other walk of life that appears relevant to the problem
facing the decisionmaker. Some of this advice is based on
solid generalizations from observation, either of the
"crude" empirical variety or somewhat more prestigious de-
ductions from established scientific principles. A great
deal of it is "opinion."

The notion of opinion is extremely fuzzy, but with your indulgence, I would prefer not to try to make it precise. With respect to the interests of this conference, I believe you will agree that in the area of long range forecasting of technological and social developments there is an especially large admixture of opinion. For this area, the creation of techniques for refining opinion is of particular interest.

Pragmatically, a basic characteristic of opinion as opposed to more solid knowledge is the fact that if you interrogate several equally competent individuals, you are likely to get a divergence of answers. This is obviously not a defining characteristic, since uniformity of response does not guarantee the solidity of that response. From the standpoint of the decisionmaker, a divergence of estimates creates a problem of how to use the estimates in fashioning his policies. There are several heuristic devices that are traditional in the advice community. One is to select a single advisor on some grounds (ranging all the way from personal friendship to lustre within the community). This usually guarantees a certain uniformity. Another is to involve several knowledgeable individuals and employ some method of group interaction to arrive at a common opinion. The most popular of such methods is that of the committee, or commission, with a variety of informal ways to arrive at the "sense of the committee."

Selection of a single advisor in "soft" areas is clearly fraught with danger, on the other hand, committees have certain drawbacks which have been dramatized by a large number of investigations by psychologists and small-group sociologists over the past two decades (1). One major drawback is the influence of the dominant individual. A quite convincing group of studies have shown that the group opinion is likely to be highly influenced, if not determined, by the views of the member of the group who does the

most talking, and that there is no significant correlation between success in influencing the group and competence in the problem being discussed. Another difficulty which has not received as much attention in the literature is "noise"— irrevelant or redundant material that obscures the directly relevant material offered by participants. A third diffi- culty is group pressure that puts a premium on compromise.

2. DELPHI PROCEDURES

The Delphi procedures have been designed to reduce the effects of these undesirable aspects of group interaction. The procedure has three distinctive characteristics:

1. Anonymity.
2. Controlled feedback.
3. Statistical "group response."

Anonymity is a device to reduce the effect of the socially dominant individual. It is maintained by eliciting separate and private answers to prepared questions. Ordi- narily, the procedure is carried out by written question- naire; on-line computers have been used for some exercises. All other interactions between respondents is through for- mal communication channels controlled by experimenters.

Controlled feedback is a device to reduce noise (among other things). A Delphi exercise will usually consist of several iterations where the results of the previous itera- tion are "fed back" to the respondents, normally in sum- marized form.

As a representative of the group opinion, some form of statistical index is reported. For cases where the group task is to estimate a numerical quantity, the median of in- dividual estimates has turned out to be the most useful in- dex tried to date. Thus, there is no particular attempt to arrive at unanimity among the respondents, and a spread of opinions on the final round is the normal outcome. This is a further device to reduce group pressure toward conformity.

A typical exercise is initiated by a questionnaire which requests estimates of a set of numerical quantities, e.g., dates at which technological possibilities will be realized, or probabilities of realization by given dates, levels of performance, and the like. The results of the first round will be summarized, e.g., as the median and inter-quartile range of the responses, and fed back with a request to revise the first estimates where appropriate. On succeeding rounds, those individuals whose answers deviate markedly from the median (e.g., outside the inter-quartile range) are requested to justify their estimates. These justifications are summarized, fed back, and counter-arguments elicited. The counter-arguments are in turn fed back and additional reappraisals collected. This basic pattern has, of course, many possible variants, only a few of which have been tried.

The procedure has been exercised with material where there is no immediate way to evaluate the results—e.g., long range technological and social developments—and also with material where there is the possibility of checking, such as short range economic predictions and estimates of quantities where the actual figures are obtainable, typically "almanac type" material. For material where confirmation is possible, typical outcomes are that opinions tend to converge during the experiment, and more frequently than not, the median response moves in the direction of the true answer. In the case of material where confirmation is not possible, all we can say is that opinions do converge during the exercise (2) (3).

One additional feature of present Delphi procedures should be mentioned. Respondents are requested to make some form of self-rating with respect to the questions. Several different kinds of self-ratings have been tried—ranking the questions in the order of the respondents judgment as to his competence to answer them; furnishing an

absolute estime of the respondent's confidence in his
answer; estimating a relative self-confidence with respect
to some reference group. In general there has been no
significant correlation discovered between such self-
ratings and <u>individual</u> performance for confirmable esti-
mates. However, it has usually been possible to use the
self-ratings to select a subgroup of relatively more con-
fident individuals where the performance of the subgroup
has been slightly, but consistently better than the group
as a whole. In one very thorough study. the improvement
was obtained only by combining two self-rating indeces—
ranking of questions, and absolute estimates of
confidence (4).

3. RESULTS OF EXPERIMENTS

There are many things we do not understand as yet about
the information processing going on during a Delphi exer-
cise. Thus, we cannot as yet determine how much of the
convergence is due to three different factors which are
clearly at work: (1) social pressure, (2) "rethinking"
the problem, (3) transfer of information during feedback.
Several exercises have been conducted that throw some light
on this. In one (5), a set of twenty almanac type ques-
tions were posed to a group of 23 respondents. A control
group of 11 respondents were given the same questions,
but on the second round were simply asked to reassess
their answers, with no feedback whatsoever. They were not
even told what their previous responses had been. In
general, except for two questions, the amount of conver-
gence was comparable for the two groups, and the accuracy
of responses for the control group was as good as for the
second-round responses of the experimental group. This
would appear to indicate that a major factor in this
exercise was "rethinking." However, the effect of social

pressure and/or information transfer is also indicated by
the fact that for the experimental group the interquartile
ranges of the second round responses were uniformly con-
tained in the interquartile ranges of the initial responses,
whereas for the control group the second-round ranges were
contained in the initial ranges for only thirteen out of
the twenty questions.

To try to pin down a little more the factors involved,
we conducted an experiment, this summer comparing the per-
formance of structured face-to-face discussion groups and
the anonymous questionnaire technique. The experiment was
guided by two presumptions. (Hypotheses is too pretentious
a notion in this rather unstructured subject.) The first
presumption was that in a face-to-face situation, informa-
tion transfer is likely to be much greater than in the
anonymous controlled communication situation. This would
presumably tend toward greater accuracy on the part of
the conference estimates. The second presumption was that
the effect of undesirable social interactions could be
meliorated by imposing a specific format for the discus-
sions. The format employed was: for each question a new
discussion leader was selected by chance; the leader listed
on a blackboard all relevant information (including
"opinions") suggested by members of the group; he then
listed as many different approaches (little "models") for
answering the question as the group could devise; esti-
mates were made by each approach; and finally a group con-
sensus was arrived at by informal agreement.

The presumption to be tested was that a structured
conference of this sort would produce more accurate esti-
mates than the questionnaire technique. The experiment
was performed using a group of graduate students engaged
in summer consultant activities at RAND. There were ten
participants, divided into two groups of five. There were
twenty questions, of the almanac sort. divided into four sets

of five. Each participant group answered ten of the ques-
tions by questionnaire, and ten by structured discussion.
The only innovation in the Delphi procedure was to inter-
pose a pure information round between the first and second
estimation round. Each respondent was allowed to ask the
group two questions and the group replies were fed back
before the second estimate was made.

The major outcome of the experiment was that the pre-
sumption that the structured discussion would turn in a
better performance was not born out; in fact, the question-
naire responses were, if anything, somewhat more accurate
than the structured conference responses. The difference
was not significant except for one measure, namely the
sums of ranks of standard scores,[*] in which the question-
naire technique showed up as better.

For the discussion groups, no adequate measure was ob-
tained for the role of dominant members, noise, and pres-
sure for consensus; but it was clear from observation of
the discussions that the structure imposed was inadequate
to eliminate these effects.

An interesting anomaly appeared in the performance of
the questionnaire groups; namely, the responses on the
second round were more accurate than the responses on the
fourth (and final) round. Whether this was due to fatigue—
for each set of five questions, the entire set of responses
was obtained in one afternoon session—or due to a satura-
tion effect (all of the relevant information elicited by
the second round, and simply "wandering" estimates from
then on) cannot be determined from the data.

---

[*] Standard scores were computed by dividing the group
estimate by the true answer. The 40 responses were ranked
in order of accuracy, and the sums of these ranks taken
for each configuration (group, method, question set). The
analysis of variance for the sums of ranks indicated a
difference between the two methods significant at the
.05 level.

Perhaps the significance of the experiment can be most sharply summed up by the following conclusion: if the conference groups had been requested to open their session with anonymous individual "guestimates" of the answer to each question, the median of these off-the-cuff guesses would have been more accurate than the group consensus obtained after a more or less thorough discussion of the subject.

4. DISCUSSION

Delphi procedures are still in an experimental stage with regard to applications to the advice process. The evidence is mounting that systematic processing of expert opinion can produce significant improvements both in accuracy and reliability (using the notion of reliability to refer to the range of estimates). However, the role of Delphi procedures within the corpus of forecasting techniques—extrapolation, simulation, demand analysis, gaming, etc.—has not been established. In particular, there are no cases that I know of where Delphi procedures have been explicitly employed to support specific policy decisions. Hence, there are no direct comparisons of the relative effectiveness of the procedures vs. other more traditional forms of advice. The studies that I am familiar with in areas relevant to policy have been more like exploratory exercises to test the feasibility and "manageability" of the procedures with extensive subject matters and geographically scattered experts. In this respect, the procedures have turned out to be manageable, but often rather cumbersome.

A common reaction is to imagine Delphi as a method of obtaining inputs for some kind of formal estimating structure—e.g., inputs for a simulation model. I must confess that at times I find this an appealing notion, but it cannot be the full story. Most often, for those areas where data

is lacking, a formal model is lacking as well. As a matter
of fact, the Delphi procedure is one of the most efficient
I know for "uncovering" the implicit models that lie behind
opinions in the "soft" areas. One of the most valuable
side—products of a Delphi exercise concerned with strategic
bombing was the skeleton of a model which was later fleshed
out in great detail (2).

There are several tautologies which are directly rele-
vant to the group estimation process: (a) The total amount
of information available to a group is at least as great
as that available to any member. (b) The median response
to a numerical estimate is at least as good as that of
one half of the respondents. (c) The amount of misinformation
available to the group is at least as great as that avail-
able to any member. (This one is usually overlooked in
discussions of the advantages of groups vs. individuals.)
(d) The number of approaches (or informal models) for
arriving at an estimate is at least as great for the group
as for any member. (e) Corresponding item for approaches
is as (c) for information. For simplicity I have included
noise in misinformation and poor approaches.

These tautologies do not add up to anything like a
"theory" of the group estimation process, but they are
suggestive. For example, (c) and (e) hint that there may
be an optimal size of group for a given kind of estimation.
This would be in accordance with some experimental results
with small discussion groups. They also suggest that part
of the group estimation process should be concerned with
active suppression of misinformation as well as "filling
voids" in information.

We have no way at present of determining whether the
questionnaire—feedback procedure is anything like an optimal
use of the information available to a group, or whether it
includes a mechanism for reducing the effect of misinforma-
tion. Nor can we say that it is most effectively used in

isolation, or within the context of other methodologies. In short, there is a very large field waiting for the plough.

REFERENCES

(1)   Maier, Norman R. F., "Assets and Liabilities in Group
      Problem Solving: The Need for an Integrative
      Function," Psychological Review, Vol. 74, No. 4,
      July 1967, pp. 239–249.

(2)   Dalkey, N., and O. Helmer, " An Experimental Applica—
      tion of the Delphi Method to the Use of Experts,"
      Management Science, 1963, 9, 458–467.

(3)   Gordon, T. J., and O. Helmer, Report on a Long—Range
      Forecasting Study, The RAND Corporation, P—2982,
      September 1964.

(4)   Campbell, Robert, "A Methodological Study of the
      Utilization of Experts in Business Forecasting,"
      unpublished Ph.D. dissertation, UCLA, 1966.

(5)   Brown, B. and O. Helmer, Improving the Reliability of
      Estimates Obtained from a Consensus of Experts,
      The RAND Corporation, P—2986, September 1964.