

MARVIN A. JOLSON and GERALD L. ROSSOW*

This article describes a field experiment which uses the Delphi Process to assign subjective probabilities to the set of possible demand levels for a company's product. The method is validated by the use of questions to which the "correct" answers are known.

# The Delphi Process in Marketing Decision Making

## *INTRODUCTION*

When assessing probabilities in making any decision, the marketing manager can seldom rely on historical frequencies alone because of the preponderance of behavioral variables. Faced with such uncertainty, he can choose among alternate actions only when he can prejudge their success or failure. The Bayesian approach to decision making is useful in such situations because in it, "it is legitimate to quantify one's feelings about uncertainty in terms of assessed numerical probabilities, even when confronted by a single unique decision and when there is no extensive past history on which to base the assessment of probabilities to an anticipated range of states of nature" [11].

There is little interest at the theoretical level in how well subjective probability assessments by the decision maker correspond to reality [14]. Instead, the interest is in how well the assessments correspond to the assessor's judgments [7, 12, 14]. Such an approach tacitly assumes that the decision maker is a statistician, which is seldom the case [13].

When the marketing manager assigns *subjective probabilities* his a priori knowledge is an indispensable input to both prior and posterior methods of analysis. However, in the real world, business executives are reluctant to use conventional Bayesian devices largely because of the difficulties of eliciting the required prior distribution [3].

### *Assessment of Prior Probabilities*

Whether the marketing manager has limited or even substantial knowledge about the variables or parameters of interest, many statistical and psychological factors complicate his task of assigning subjective probabilities. He may require advice from co-workers or consultants who may be considered reasonably expert with regard to the uncertain variables.

In one case where others within the firm were consulted, four alternatives covered the range of feasible prices for an industrial product. Introductory sessions were held with the firm's sales managers in order to develop a set of states of nature large enough to represent an adequate description of the real problem yet small enough to be comprehended by participating sales personnel. Then, through separate interview sessions with two groups of sales personnel, subjective probabilities of the occurrence of alternative states of nature were developed. Finally, all contributing personnel discussed each projection or subjective probability in detail. The pricing action was determined using only prior analysis [8].

Traditional open discussion approaches have been hampered by psychological factors, such as a dominant pressure for conformity and an unwillingness to change a publicly expressed opinion [2, 6]. A number of mathematical and behavioral approaches to achieving group consensus without face-to-face discussion have been described [15], but few have been used by business executives.

Assignment of prior probabilities to states of nature without historical data can be done through many varieties of individual and group opinion measurement. The methods differ in the number of judges employed and the degree of anonymity allowed for individual responses. The technique usually followed is a function of the time available, the importance of the decision, the number of qualified judges available, and the relative merit placed on the alternative procedures by the decision maker.

### *The Delphi Method*

The purpose of this article is to test the feasibility of a specific group consensus method as a source of sub-

* Marvin A. Jolson is Assistant Professor of Marketing and Gerald L. Rossow is a doctoral student at the University of Maryland. They thank Professor Harry V. Roberts of the University of Chicago for his critical review of the manuscript.



*Journal of Marketing Research*, Vol. VIII (November 1971), 443-8

jective probabilities for the various levels of demand for a company's product. This method is the Delphi Process (which was developed at The Rand Corporation by Olaf Helmer, Norman C. Dalkey, and Theodore J. Gordon).

The Delphi Process is a unique method of eliciting and refining group judgment, based on the rationale that *n* heads are better than one when exact knowledge is not available. Delphi uses a panel of experts and repeated measurement and controlled feedback and replaces direct confrontation and debate with a planned program of sequential, individual interrogations usually conducted by questionnaire [3].

The salient features of the process are: (1) anonymity, (2) controlled feedback, and (3) group response. Anonymity, ensured by the questionnaires, reduces the effect of dominant individuals. Feedback, controlled by conducting the experiment in a series of rounds and notifying judges of the previous rounds' results, reduces noise. Group response ensures that the opinion of every member of the panel is taken into consideration.

Within the three features there can be considerable variation, particularly in the controlled feedback system. The results returned to respondents may be, at one extreme, the median response of the previous round, or at the other extreme, the 25th, 50th, and 75th percentiles plus a list of respondents' reasons for responses. The purpose of this feedback is to encourage informed judgment and eliminate pressure to conform to the opinions of prestigious judges.

The basic theory upon which the Delphi method is based is two-fold: (1) that with repeated measurement the range of responses will decrease and converge toward the midrange of the distribution and (2) that the total group response, or median, will successively move toward the "correct" or "true" answer. The basic characteristics of Delphi can be maintained, while varying panel size and the feedback system to satisfy a wide range of decision-making restrictions. Of course, the validity and reliability of the method for obtaining subjective judgments must be evaluated against available alternatives.

### INITIAL QUESTIONNAIRE

I. Please make a subjective estimate of the various demand levels for Pace Corporation courses in the New York City area during the following year. The sum of the probability estimates should equal one. In estimating these probabilities make the following assumptions:
   1. Staff limitations will restrict the number of courses taught to 12 or less.
   2. Course length is not a factor.

| *Course demand in the New York City area next year* | *Subjective probability* |
|---|---|
| 0 | |
| 7–9 | |
| 1–3 | |
| 4–6 | |
| 10–12 | |

II. What is your estimate, in minutes and seconds, of the time it would take to run (compile, link edit, and execute) a source program which will perform a regression analysis (fit a linear trend line) on 20 points of data and then graph this data and the calculated regression line?
_____MIN _____SEC

III. In your estimation, what was the total enlisted strength of the U.S. Navy (including USNR) on 1 January, 1968?
_____PEOPLE

Name_____

### *Previous Applications of Delphi*

The Delphi technique has been largely employed in technological and environmental forecasting. It has been applied in the selection of an optimal industrial target system and the estimation of the number of enemy A-bombs required to reduce the United States' munitions output by a prescribed amount [5]. In 1954, The Rand Corporation used several panels of experts to make contingency forecasts regarding the state of the world 25 to 50 years in the future [5]. Delphi was also used by the Charles F. Kettering Foundation to generate some perspectives on possible changes in American education [1].

## *AN ILLUSTRATIVE EXAMPLE*

Our experiment was conducted at the Pace Computing Corporation,[1] Arlington, Virginia. Pace, which provides programming analysis and consulting and educational services, was considering opening a branch in or near New York City if demand for its educational services was sufficient. The lack of data and precedent for this decision made it suitable for application of the Delphi method.

The 11-member Pace staff, actively aware of the corporation's goals and performance, was assumed to have a feel for probable demand for classes in New York. The staff served as the panel of experts, and with their help a payoff table was constructed showing five different levels of demand for class instruction. A required input for the determination of the optimal strategy was the assignment of probabilities of occurrence to these five demand levels (states of nature).

Questionnaires were delivered to the panel with cover letters briefly explaining the Delphi method and asking that respondents not discuss the questions among them-

[1] Pace is one of the few competitors with IBM in the instruction of the organization, facilities, control language, programming, and coding of the 360 operating system. At the time of the study, classes were available at Pace's Arlington office, at the site of the individual contracting company, or at an otherwise mutually agreeable location.

selves. The questionnaires were printed by computer and the order of the categories randomized to eliminate any ordering effect (see the figure).

The median response was then determined for each category and responses were normalized so that the sum of the probability medians totalled one:

| *Number of classes demanded* | *Probability median* | *Normalized probability median* |
|---|---|---|
| 0 | .05 | .07 |
| 1–3 | .15 | .20 |
| 4–6 | .25 | .33 |
| 7–9 | .20 | .27 |
| 10–12 | .10 | .13 |
| | .75 | 1.0 |

A second questionnaire was then submitted to the Pace staff. This instrument was identical to the figure except that it reported the first-round results including the adjusted probability medians for each demand level. After the second questionnaire was completed, the entire process was repeated in order to develop the final estimates.

*Delphi Validity Test*

With the iteration and feedback process, changes in probability estimates tended to be in the same direction, that is, the distribution of individual responses progressively converged. To be of value in decision making, the group response (defined as the median of each round's individual responses) should move in the direction of the "correct" answer, although in this study there were correct responses only to two questions, on naval manning and computer running time. For the first, there was an exact answer, and a panel of 14 officers and civilians from the Enlisted Plans Section of the Bureau of Naval Personnel was added to the study as experts in this area. For the second, the Pace personnel were considered experts, and the correct answer could vary within a reasonable range. These two questions thus clearly separated expert and nonexpert respondents and helped in: (1) examining the movement toward accuracy or validity and (2) verifying the expertness of respondents.

## *FINDINGS*

The adjusted probability estimates for all levels of demand for the New York project, by rounds, are given in Table 1. The optimal company strategy was computed using the figures from Round 3.

The range and variance of replies for each round are summarized in Tables 2 and 3. The initial round showed widespread individual answers, but with iteration and feedback, the distribution of individual responses narrowed. In comparing the variance between rounds, the sample space may be considered to be the set of respondents' possible probability estimates. If the variance values of the three rounds are expressed by $s_1^2$, $s_2^2$, and $s_3^2$, the observed differences ($F$-tests) between Rounds 1 and 3 are significant in all cases.[2]

**Table 1**

**ADJUSTED PROBABILITY ESTIMATES FOR ALL ROUNDS**

| *Round* | *Probability medians for various levels of demand* | | | | | |
|---|---|---|---|---|---|---|
| | *0* | *1–3* | *4–6* | *7–9* | *10–12* | *Total* |
| Round 1 | .07 | .20 | .33 | .27 | .13 | 1.0 |
| Round 2 | .05 | .18 | .31 | .31 | .15 | 1.0 |
| Round 3 | .00 | .16 | .36 | .36 | .12 | 1.0 |

Postexperimental questioning of Pace panel members indicated that the assumptions underlying their forecasts were cautiously optimistic based on favorable primary demand for classes, inexpensive promotional requirements, limited competition except for IBM, and expected favorable word-of-mouth recommendations by students who completed the course. In fact, the data disclosed an apparently increasing optimism with iteration and feedback, as evidenced by an increase in the expected level of demand from 5.64 classes following Round 1 to 6.32 classes following Round 3.[3]

Convergence of the responses away from a correct answer would surely not be desirable.[4] An analysis of the two supplementary questions investigates such a possibility.

The estimates of naval strength made by the Bureau personnel were quite accurate throughout the experiment. As illustrated by Table 4, the range of responses decreased with each round and the final median response was within 3,000 of the correct answer. On the other hand, the median responses of the Pace personnel range estimate narrowed but remained at a very wide 380,000 people.

Conversely, since the Pace staff could be classified as experts in the area of estimating computer program running time, this group came within 9 seconds of the exact answer, despite a slight widening of the range of estimates on the third round. Responses of the naval personnel to the computer time question revealed a clear lack of confidence. Afterwards, many respondents remarked that they were not sure what was being requested for the range was 25 times that for the Pace staff's responses. The Navy personnel's median responses remained quite stable around the one-minute point.

It was hoped that the data would furnish some insights into the mechanics of the basic estimation proc-

[2] A possible exception is the natural state of zero demand for classes, where the results are significant at the .07 level.

[3] These calculations were made using the midpoint of each level of demand.

[4] The personalist view is not really presumed to generate correct assessments. All self-consistent or coherent assessments are admissible if an expert feels they correpond to his judgments [13].

**Table 2**
RANGE OF ANSWERS ON EACH ROUND

| *Demand for classes* | *Round* | | |
|---|---|---|---|
| | *1* | *2* | *3* |
| 0 | .20 | .15 | .10 |
| | (0, .20)[a] | (0, .15) | (0, .10) |
| 1–3 | .70 | .40 | .28 |
| | (0, .70) | (0, .40) | (.02, .30) |
| 4–6 | .55 | .30 | .30 |
| | (0, .55) | (.15, .45) | (.20, .50) |
| 7–9 | .45 | .40 | .40 |
| | (.05, .50) | (.10, .50) | (.15, .55) |
| 10–12 | .95 | .20 | .25 |
| | (0, .95) | (.05, .25) | (.05, .30) |

[a] The largest and smallest numbers in the set of answers are given in parentheses beneath the range.

ess. Since a respondent must rationalize any difference between his personal opinion and the consensus of the group, his judgment is dependent on both the degree of disparity in responses and his self-confidence. It is likely that a knowledgeable, confident panel member will probably not revise his predictions over iterations. This was supported by a significant correlation between the individuals' first-round accuracy and variance in estimates by round for naval personnel in response to the naval strength question.

In addition, the assumption that the correct answer will continue to be an underlying force, even under the obvious attraction of the group median, was supported (although no significance test could be performed). Half the responses to the second- and third-round questionnaires for the validation study were closer to the true answer than to the median returned from the previous round. Also, the attraction of the true answer seemed to vary directly with the distance between the group median and true value. In the case of the computer time question, the responses of 75% of the Pace staff were closer to the correct answer when the returned median was off by 21 seconds, whereas only 42% of the responses were closer when the median was off by only 9 seconds.

Analysis of the validation questions also revealed that experts could consider possible responses in terms of a ratio scale. That is, they were concerned with the relationship between estimates in addition to their magnitude, as indicated by the roughly log-normal shape of the frequency distribution of normalized logarithms of responses. This log-normal shape and ratio-scale hypothesis was not supported, however, by the random appearance of the nonexpert frequency distribution.

## *POSSIBLE MODIFICATIONS OF THE PROCESS*

The specific purpose of any given marketing project and the complex interactions associated with any communication process open the door to a number of modifications and refinements of the basic Delphi Process used in this experiment. The following observations and recommendations are based on our experience in the study.

First, it is clear from the two check questions that knowledge of the subject area is necessary for the proper operation of the Delphi Process. Only when respondents were experts was the basic process of convergence upon the correct answer supported.

If the panel members' expertise is directly measurable, greater or even exclusive weight can be given to the opinions of the most knowledgeable participants. In one study without such measurements, each panelist ranked his own relative competence in estimating ten economic and business indices. After four rounds of Delphi arguments, only the responses of those who had ranked themselves most competent for particular indices were used. This *select* median was closer than that of all respondents to the true value in 13 out of 20 cases [10.]

Second, although in this experiment only the median values of each round were fed back, other information could also be returned. For example, the interquartile range (containing the middle 50% of responses) could be fed back, and if a panel member's new response lay outside it, he could be required to explain why his answer should be that much lower, or that much higher, than the group's majority judgment. Such judgment may stimulate thinking, clarify poorly articulated problem parameters, and influence panelists to give new weight to factors which may have been overlooked or inappropriately rated. However, those without strong convic-

**Table 3**
ANALYSIS OF VARIANCE BETWEEN ROUNDS

| *Demand for classes* | *Variance* | | | *Rounds compared* | *F-value*[a] | *Level of significance* |
|---|---|---|---|---|---|---|
| | *1* | *2* | *3* | | | |
| 0 | .0040 | .0024 | .0014 | 1 and 2 | 1.667 | n.s. |
| | | | | 1 and 3 | 2.857 | .07 |
| | | | | 2 and 3 | 1.714 | n.s. |
| 1–3 | .0426 | .0123 | .0072 | 1 and 2 | 3.463 | .03 |
| | | | | 1 and 3 | 5.917 | .005 |
| | | | | 2 and 3 | 1.708 | n.s. |
| 4–6 | .0362 | .0307 | .0077 | 1 and 2 | 1.179 | n.s. |
| | | | | 1 and 3 | 3.987 | .01 |
| | | | | 2 and 3 | 3.987 | .02 |
| 7–9 | .0225 | .0206 | .0066 | 1 and 2 | 1.092 | n.s. |
| | | | | 1 and 3 | 3.409 | .04 |
| | | | | 2 and 3 | 3.121 | .05 |
| 10–12 | .0735 | .0039 | .0082 | 1 and 2 | 18.846 | .0005 |
| | | | | 1 and 3 | 8.963 | .001 |
| | | | | 2 and 3 | 0.476 | n.s. |

[a] All calculations are based on 10 and 10 d.f.

Table 4
ESTIMATES OF NAVAL STRENGTH AND COMPUTER RUNNING TIME

| | *Bureau personnel* | | | *Pace personnel* | | |
|---|---|---|---|---|---|---|
| | *1* | *2* | *3* | *1* | *2* | *3* |
| *Naval strength* | | | | | | |
| Median | 651,000 | 652,000 | 654,500 | 725,000 | 750,000 | 763,000 |
| Range | 162,599 | 105,000 | 33,000 | 3,300,000 | 400,000 | 380,000 |
| Lowest estimate | 617,401 | 635,000 | 647,000 | 200,000 | 600,000 | 620,000 |
| Highest estimate | 780,000 | 740,000 | 680,000 | 3,500,000 | 1,000,000 | 1,000,000 |
| Correct answer | 657,533 people | | | | | |
| *Computer running time* | | | | | | |
| Median | 60 | 55 | 57 | 150 | 120 | 120 |
| Range | 22,495 | 1,068 | 1,068 | 890 | 90 | 113 |
| Lowest estimate | 5 | 15 | 15 | 10 | 90 | 67 |
| Highest estimate | 22,500 | 1,083 | 1,083 | 900 | 180 | 180 |
| Correct answer[a] | 129 seconds | | | | | |

[a] Running time varies as a function of a multiprocessing environment. Repeated runs of the program yielded an average time of 129 seconds.

tions may move their estimates closer to the median while those who are more confident may tend to defend their original estimate [10].

Third, what is the optimal number of iterations? Termination could take place at a predetermined level of interround variance or when the variance changes are no longer significant.

Fourth, by *credence decomposition,* the target variable can be expressed as a function of two or more components [3]. For example, demand for Pace's classes could be expressed as a product of total primary demand and Pace's expected share of the New York market. Each component could then be assessed probabilistically on available evidence.

Finally, if panel members are separated by distance and since mail responses may introduce delays, each expert could be provided with a console to feed his responses to a computer, which would feed back some measures of group response and other relevant information. At The Rand Corporation, small Delphi experiments are being conducted that use personal, electric typewriter consoles connected through an on-line time-sharing computer system.

## *SUMMARY AND CONCLUSIONS*

The Delphi Process is very useful in assigning prior probabilities in a marketing decision to be made under uncertainty. The method is preferred to the hunch of a single decision maker or the consensus of a group following face-to-face discussion. In direct comparisons of face-to-face and Delphi procedures (although all experiments did not yield clear outcomes) discussion was not better than anonymous interaction [4, 6]. Obviously, however, there are numerous situations where one well informed person or a cohesive committee could be more effective than a moderately informed Delphi panel.

Also, the Delphi Process eliminates time consuming and sometimes argumentative meetings within the firm. A Delphi exercise, properly managed, can be highly motivating, since its systematic procedures lend an air of objectivity to the outcomes [6]. The convergence of the group response to this study's problem and the validity test offered by the two supplementary questions provided convincing evidence of the usefulness of the process.

A major advantage of Delphi lies in compelling each judge to make explicit what elements of a situation he takes into consideration and clarify the concepts he uses. The technique thus establishes unambiguous intersubject communication.

When dealing with multifaceted marketing problems, each expert is likely to be a specialist, so it may not be easy to achieve agreement on the identity of the problem. A common frame of reference is necessary in order to promote a unified, collaborative effort. Delphi helps each participant acquire an integrated overview of the problem area and acts as a catalyst in crystallizing the reasoning process, even in the absence of a group consensus [9, 10].

This experiment was intended merely to illustrate the potential of the Delphi technique. Further studies can test the extent of its validity and refine it so that it will become a standard tool for generating prior probabilities in business decisions. Because of the limited panel size and the simplicity of the decision at hand, our findings are not conclusive; rather, they suggest ways of experimentation and implementation in varying marketing frameworks by proponents of Bayesian decision theory.

## REFERENCES

1. Adelson, Marvin, *et al.* "Planning Education for the Future: Comments on a Pilot Study," *American Behavioral Scientist,* 10 (March 1967), 1–12; 21–31.

2. Brown, Bernice B. *Delphi Process: A Methodology Used for the Elicitation of Opinions of Experts*. Santa Monica: The Rand Corporation, 1968.
3. Brown, Rex V. "Assessing the Accuracy of Estimates: A Personalist Approach," paper presented at the Conference of the American Statistical Association, 1969.
4. Campbell, R. "A Methodological Study of the Utilization of Experts in Business Forecasting," unpublished doctoral dissertation, University of California at Los Angeles, 1966.
5. Dalkey, Norman C. *The Delphi Method: An Experimental Study of Group Opinion*. Santa Monica: The Rand Corporation, 1969.
6. ——— and O. Helmer. *An Experimental Application of the Delphi Method to the Use of Experts*. Santa Monica: The Rand Corporation, 1962.
7. De Finetti, Bruno. "Foresight: Its Logical Laws, Its Subjective Sources," translated by Henry E. Kybrug, Jr., in Henry E. Kybrug and Howard E. Smokler, eds., *Studies in Subjective Probability*. New York: John Wiley & Sons, 1964.
8. Green, Paul E. "Bayesian Decision Theory in Pricing Strategy," *Journal of Marketing*, 27 (January 1963), 5–14.
9. Helmer, Olaf. *Social Technology*. Santa Monica: The Rand Corporation, 1965.
10. ———. *Analyses of the Future*. Santa Monica: The Rand Corporation, 1967.
11. Roberts, Harry V. "Bayesian Statistics in Marketing," *Journal of Marketing*, 27 (January 1963), 1–4.
12. ———. "Statistical Inference and Decision," unpublished paper, Graduate School of Business, University of Chicago, 1966.
13. Winkler, Robert L. "The Assessment of Prior Distributions in Bayesian Analysis," *Journal of the American Statistical Association*, 62 (September 1967), 776–81.
14. ———. "The Quantification of Judgment: Some Methodological Suggestions," *Journal of the American Statistical Association*, 62 (December 1967), 1105–6.
15. ———. "The Consenses of Subjective Probability Distributions" *Management Science*, 15 (October 1968), 61–3.