A CRITIQUE OF THE DELPHI TECHNIQUE

Gordon Welty, American University

The Delphi technique is a recent RAND Corporation development in long-range forecasting. A group of experts are polled for their opinions on a given forecasting problem. The opinions are aggregated and fed back for a second round of opinion formation. Hence, the exercise is iterative, and polling, aggregation and feedback continue until consensus develops /cf. 3/.

In particular, a percentage of the respondents giving the most extreme individual forecasts each round, usually the upper and lower quartiles, are requested to reconsider the forecast they gave in a given round, in succeeding rounds, in light of their deviance from the group norm. This labelling or selection procedure supposedly hastens the development of consensus.

In several papers, we have examined a number of problems associated with structural aspects of the Delphi technique, such as aggregation of expert opinions /9/ and the selection of experts /10/. In this paper we shall study a further problem related to Delphi, namely that of the robustness of the Delphic exercise in withstanding deliberate manipulation of judgment and deceitful opinion formation.

It is important for our purposes to differentiate risk bearing from confronting uncertainty. If instead of a known world, we are talking about "unique events," etc., then known frequencies will not apply, since there won't be any sequences upon which the frequency can be based. Consider the distinction between a mechanical "one-armed bandit" with its known risk and pari-mutuel gambling schemes with its uncertainty. In the case of the mechanical schemes, there are fixed odds (known perhaps only to the house), while in the case of pari-mutuel betting, the odds change constantly, as a function of the social psychologically determined behavior of the bettors.

Following Frank Knight /5/, the relative frequencies in the known or measurable case is called risk and one's judgments in the case of unmeasurable circumstances will be judgments of uncertainty. For the latter, Knight notes that there are two fundamental methods of dealing with uncertainty, based respectively upon "reduction by grouping" and upon "selection of men" to control uncertainty /5, p. 239/. Grouping consists in categorizing the world (or its attributes) which confront men, while selection consists of categorizing the men who confront the world. For grouping, Knight has emphasized that nothing in the universe of experience is absolutely unique any more than any two things are absolutely alike. Consequently, it is always possible, for Knight, to form classes if a loose enough interpretation of similarity is accepted /5, p. 227/.

Knight goes on to point out several illustrative social institutions which deal with uncertainty. For grouping, the best-known sort of institution is that of insurance. The best-known institutions which select men to control uncertainty are those of speculation. These include future markets, produce and security exchange, etc. In elaborating upon the distinction between grouping and selecting men to control uncertainty, Knight points out that in the former case, the institutions don't serve to lessen "real risk" but merely spread it around. In the second case, however, the institutions do, by a process of elimination, lessen the real risk. "There is better management, greater economy in the use of economic resources, as well as a mere transformation of uncertainty into certainty" /5, p. 259/.

As Knight points out, "The problem of meeting uncertainty . . . passes inevitably into the general problem of management, of economic control" /5, p. 259/. This is the point at which we seek to specify a problem to be expected in a Delphi exercise. Since, in Delphi, the control of the exercise is centralized, in the person of the Delphi manager, we have the institutional analogue of the monopolistic institution rather than the competitive (decentralized) institution. As is well-known in elementary economic analysis, a monopolistic institution frees one parameter to function variably (e.g. "price"). Since this parameter can function variably, it can be "fixed" at a socially suboptimal level by the monopolist. The same, we will suggest, may be the case for the institution of the Delphic exercise.

The Delphi manager may deliberately misrepresent the outcome of one Delphi round to the participating experts in the next round, in an attempt to influence the outcome of the entire exercise. This influence would operate along the same lines as the group influence observed in Sherif-Asch type social psychological experiments /1;8, Ch. 7/. Being an "expert" and supposing the rest of the experts disagree with oneself may be sufficient for opinion change. Evidence for such an hypothesis can be provided by an experiment with a group of Delphi participants, wherein an attempt is made deliberately to mislead them. Are the Delphi participants susceptible to such influence and deception?

The Greek historians tell the story

of King Athanus of Alus in Thessaly who had two wives, first Nephele and then Ino. Ino was jealous of her step-children, and planned their death. A famine occurred after Ino convinced the local women to secretly roast the seed grain, and no crops grew. Athanus sent a messanger to the Oracle of Delphi to find the cause of the famine. Ino bribed the messenger to lie on his return. The messenger consented and said falsely the Oracle proclaimed the famine would cease only when Nephele's children were sacrificed to Zeus. Thus, we find deception.

The children, however, escaped to Colchis in Asia Minor, on the Golden Ram whose fleece was later retrieved by Jason and the Argonauts. Then an Oracle proclaimed that Athanus must be sacrificed for the country. The wicked Ino and her children met unhappy ends, and the king went insane and left the country. Thereafter the eldest male heir in each generation of the family of Athanus was sacrificed, since Athanus' sacrifice never properly occurred /4, pp. 161-163/. The point we would like to emphasize is that the Oracle, once deceived, was apparently unable to rectify its utterances. While this tale has an element of myth about it, nevertheless a basic point we wish to address is clear.

The anonymity of the experts who participate in a Delphi exercise may have a unique effect, precisely the same as we saw in the Greek tale of the deception of the Oracle, that warrants consideration. With the Delphi technique, not only is rectification of erroneous assumptions perhaps not possible, as we have seen, but the mechanism may actually facilitate reaching erroneous conclusions. Indeed, Dean Cyphert and Dr. Gant have presented experimental evidence that suggests that this is the case /2/.

In an attempt further to examine this circumstance, we replicated a substantial portion of Professor Rescher's Delphi-like study of anticipated changes in American values by the year 2000 A.D. /7/. While the questions posed were identical to those of Rescher's generic or "primary" Question 2, and comparable procedures were used, the respondents were radically different. Instead of high status scientists, such as used by Rescher /6, p. 21/, 192 sophomore engineering students were selected.

We were able to gather demographic information from 168 of the student participants. 78% reported they were sophomores, and 61% were 19 years of age. 24% reported that they resided within cities of 50,000 or more persons; 32% resided in suburban areas; 35% resided within towns of 2,000 to 50,000 persons; the remainder of the participants reported they resided in rural areas. Father's occupation provides a rough and ready measure of socio-economic status.

8% of the student participants reported their father's occupation as "professional," 10% as "managerial," 13% as technical and engineering, and 12% reported their father's occupation as sales, services, or bureaucratic. In addition, 21% reported their father's occupation as a craft (skilled blue collar) and 25% as unskilled labor. The remainder of the responses were scattered across several self-employment categories.

A preliminary null hypothesis was that there is no difference in the ability to forecast changes in values by high status scientists and university sophomores.

For each of Rescher's 37 items or "secondary questions," representing a value of American society in the year 2000 A.D. /cf. 7, p. 140, p. 145/, an opinion of the probable change in emphasis was elicited on a five point scale (ranging from 1 = greatly increased emphasis to 5 = greatly decreased emphasis). Each item mean was computed and compared by means of the F-test with the (rescaled) item means reported in Rescher's study. Since the covariance structure of Rescher's data was unknown, it was not possible to compute a single multivariate F ratio for this comparison. Instead, we computed a univariate F for each of the 37 items. For 21 of the items, we found no significant difference at $p = 0.05$ between the means of Rescher's distribution and our distribution. 8 items showed that Rescher's subjects expected more emphasis in the future upon the value itemized, and the other 8 items showed that the students expected more emphasis on the value itemized. A tabular display of this analysis, including item means and standard deviations, is available in Table I following.

Hence, we concluded that there was no significant difference in the ability to forecast values and their changes of the two groups of respondents. Thus, we felt confident, at least for forecasting subject-matter of values and their changes, that we could use the student Delphi participants as the subjects of a further study of the Delphi technique.

We can then turn to our major hypothesis. Let us suppose of a Delphi exercise that the median of the first round of forecasts indicates a specific central tendency in the expert's judgments. If the manager of the Delphi exercise chooses arbitrarily and deceitfully to change the forecast, as did Ino, and feeds back a median value for the second round substantially different from the "true" value, then there are two interesting alternatives to consider. On the one hand, it might be supposed that the expertise of the respondents would permit their immediate recognition of the deception, in which case they would seek to reestablish the "true" value, refuse to participate in

TABLE I

Item No.	Mean (Rescher)	S. D. (Rescher)	Mean (Student)	S. D. (Student)	Significance of F (df = 1,246; p = 0.05)
1	2.02	0.94	1.79	0.86	
2	1.80	0.80	1.65	0.80	
3	1.96	1.01	2.39	1.08	R
4	1.70	0.69	2.00	0.94	R
5	2.68	1.01	2.65	0.99	
6	2.27	1.12	2.19	1.12	
7	3.59	1.08	3.16	1.15	S
8	2.70	0.97	2.48	1.10	
9	2.95	1.12	2.81	1.19	
10	2.41	1.14	2.54	1.09	
11	2.68	1.18	2.73	1.09	
12	2.70	1.16	3.02	1.21	
13	3.27	1.20	2.21	1.24	S
14	3.27	1.45	2.66	1.30	S
15	1.71	0.80	1.67	0.76	
16	2.91	0.94	2.50	1.00	S
17	2.05	0.86	2.43	3.02	
18	2.43	0.89	2.27	0.93	
19	3.02	0.96	3.34	1.13	R
20	3.43	1.01	3.43	0.98	
21	3.32	0.96	2.69	1.02	S
22	2.54	0.99	2.29	1.15	
23	2.59	0.93	2.69	1.17	
24	2.79	1.16	2.66	1.10	
25	2.27	1.07	1.69	1.04	S
26	2.29	0.95	2.73	1.09	R
27	2.04	0.95	2.38	0.97	R
28	2.59	0.87	2.92	1.01	R
29	2.23	1.01	2.65	1.01	R
30	1.84	0.65	1.88	0.92	
31	3.48	1.04	3.42	0.99	
32	2.52	1.10	2.79	0.98	
33	2.09	0.64	2.10	0.88	
34	1.95	0.96	1.53	0.86	S
35	1.70	0.83	2.35	1.02	R
36	2.43	1.08	2.23	0.95	
37	2.86	1.09	2.18	1.15	S

Legend: We have included above the 37 item means and standard deviations for the responses as given by Rescher's experts and the student participants. Also, we indicate that the differences were not significant (blank), that Rescher's experts expected more emphasis on the item in the future (R), or that the student participants expected more emphasis (S). Recall that the <u>lower</u> the mean value, the higher the expected emphasis on that value. The items are listed in 7, p. 140.

an obviously corrupt exercise, etc. In the Greek tale we gave, this would have been illustrated by the Oracle's refusal to prescribe the sacrifice of Athanus.

On the other hand, it might also be expected that the experts would not return, in subsequent rounds, to the "true" value. Under this alternative, they would more likely reflect the arbitrarily chosen median in the second, etc. rounds, as the Greeks relate the Oracle in fact did, and even rationalize their first round "deviance" from what they suppose to be a group judgment. Hence, on the second alternative, the experts in the Delphi exercise are deceived.

The second alternative might be expected because the expert who has been deceived in this fashion will not have prepared a rationale for his deviance and status of "minority of one," as would have an "intellectual maverick." After all, the deceived was (and still is) part of the deceived majority. Hence, he will be inclined to change his "deviant" judgment to accord with what he supposes to be the group judgment, rather than generate a rationalization for an unanticipated iconoclasm. Once he has made this accommodation, he can then rationalize his new estimate by denouncing his earlier assumptions.

As we have noted, Cyphert and Gant conducted an experiment which bore on this problem. While undertaking a Delphic exercise on the goals for the School of Education of the University of Virginia at Charlottesville, they introduced a "bogus goal" which was initially rated as having a low priority among all goals considered by the Delphi participants. The consensus was distorted and reported in later rounds as positive, and the final consensus showed the bogus goal rated considerably above the average /2, p. 13/. They concluded that "the hypothesis that the /Delphi/ technique can be used to mold opinion as well as to collect it was supported" /2, p. 14/.

We varied Cyphert and Gant's experimental procedure somewhat in our examination of the effectiveness of influence processes in causing shifts in group opinion. We fed back information to the subjects in the second round that labelled, for each of the 37 items noted above, various percentages of the respondents deviant. The range of percentages was from seven and a half percent, as a lower limit, up to eighty-six percent, as an upper limit. Thus, percentage labelled deviant per item was the independent variable, and would appear on the face of it to be a somewhat weaker intervention than that given in Cyphert and Gant's experiment. The dependent variable was the distance moved from the first to the second round. The mode of analysis was the product moment correlation.

If the correlation was negative, we could suppose that the first alternative, that the participants were not susceptible to influence and opinion formation, was true. If the correlation had been positive, we could suppose that the second alternative, that for Delphi participants to be arbitrarily labelled "deviant" could cause substantial shifts in the group opinion being formed by a Delphi exercise, was true. In fact, the correlation was a healthy -0.53; hence, we found evidence (at $p < 0.01$ for $df = 35$) that arbitrary labelling of deviants did not have an effect upon opinion formation to be expected on the basis of Cyphert and Gant's experiment; cf. scattergram in Fig. I.

Figure I.

PERCENT LABELLED DEVIANT

380
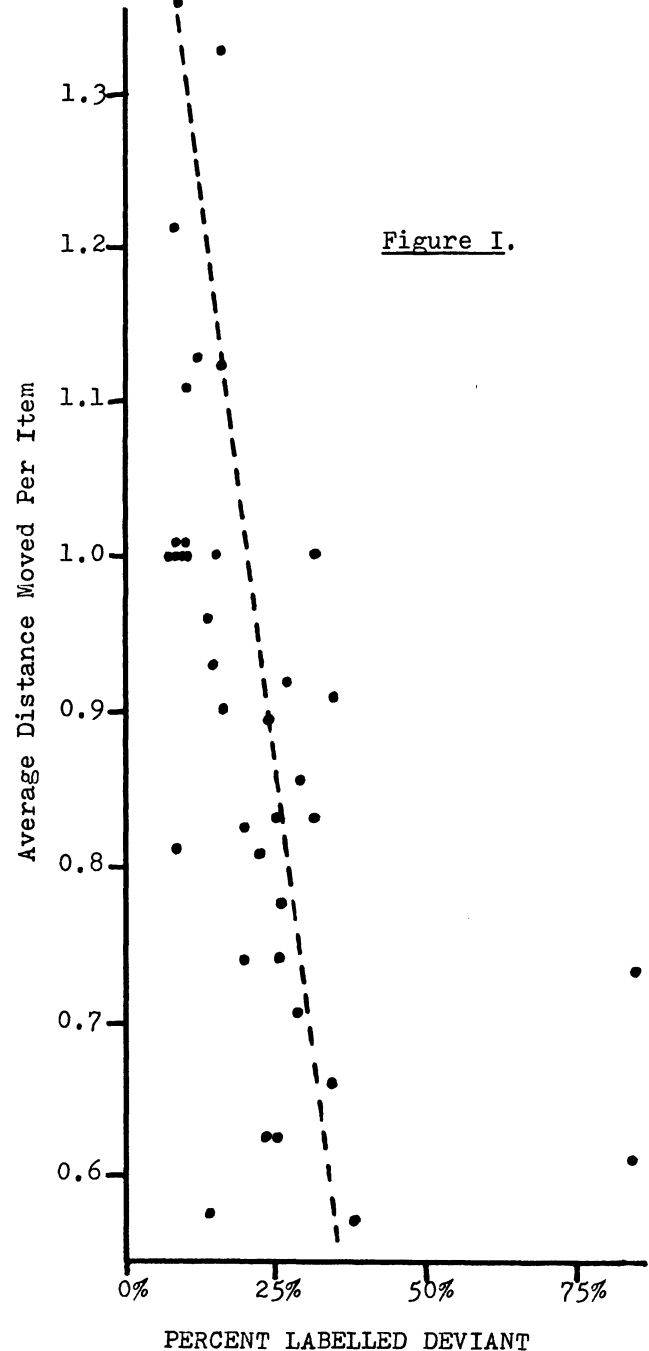
The data from our experiment is available in Table II below.

The Delphi technique appears, on the basis of our research, to be more powerful an institution in resisting wilful and arbitrary manipulation than we might have been lead to believe on the basis of Cyphert and Gant's work. At least one point is clear: further research on the structure of the Delphi exercise is called for.

SUMMARY

The Delphi exercise as a long-range forecasting technique can be considered, in Frank Knight's terms, a mechanism for coping with uncertainty. This permits us to bring to bear on Delphi the corpus of microeconomic theory. As such, the exercise can be treated as an institutional analogue to the monopolistic market of economic analysis. Such an anology allows us to note that, as in the case of the monopoly, the institution of Delphi and its manager can fix one variable of the analytical scheme at a socially suboptimal level.

The Greek historians relate an event in the kingdom of Alus where such a monopolistic practice occurred. The Oracle of Delphi's pronouncements were deliberately distorted, and the oracles were unable to recti y this distortion. The socially suboptimal level of functioning of the oracles was maintained. Some recent evidence suggests that

TABLE II

Item No.	Deviations 0	1	2	3	4	Average Distance	No. Deviants
1	5	6	3	1	0	1.000	15
2	4	7	2	1	0	1.000	14
3	19	35	9	2	0	.908	65
4	13	16	3	2	0	.824	34
5	22	14	10	0	0	.739	46
6	8	5	6	2	0	1.095	21
7	12	8	6	3	0	1.000	29
8	32	23	10	0	0	.662	65
9	12	9	9	0	0	.900	30
10	86	44	20	7	3	.731	160
11	10	10	8	0	0	.929	28
12	8	7	13	1	1	1.333	30
13	14	4	11	2	1	1.125	32
14	27	8	6	7	1	.918	49
15	3	6	4	1	0	1.214	14
16	27	12	12	0	0	.706	51
17	18	13	6	1	0	.737	38
18	22	18	5	0	0	.622	45
19	23	19	12	4	1	1.000	59
20	20	9	12	0	0	.805	41
21	6	7	3	0	0	.813	16
22	10	6	9	0	0	.960	25
23	98	37	16	5	3	.604	159
24	20	23	10	1	0	.852	54
25	6	7	4	2	0	1.105	19
26	9	6	6	0	2	1.130	23
27	24	15	5	1	0	.622	45
28	4	2	7	1	0	1.357	14
29	18	15	9	2	0	.886	44
30	14	9	3	0	0	.577	26
31	23	11	11	2	0	.830	47
32	7	4	5	1	0	1.000	17
33	36	32	3	1	0	.569	72
34	8	8	1	2	1	1.000	20
35	24	22	14	0	0	.833	60
36	22	18	8	0	1	.776	49
37	5	8	2	2	0	1.059	17

Legend: We have included above the 37 items and the frequency distribution of distances moved between rounds one and two of the exercise. The last two columns give the average distance moved per item, and the number of participants who were labelled deviant and thus were respondents for a given item. The data of the last two columns were the basis of the correlation of -0.53, as well as the scattergram displayed overleaf.

similar distorting effects can be induced in a Delphic exercise and are not amenable to rectification within the exercise. We have undertaken a study which bears on the problem of the deliberate distortion of the responses of a Delphi exercise, and conclude on the basis of our data that the structure of the Delphi technique is more robust than one might have concluded on the basis of the earlier work.

ACKNOWLEDGMENT

We would like to thank Mr. Harris N. Miller for his helpful comments and criticisms of this study. Of course, the author alone is responsible for any errors or opinions remaining.

REFERENCES

/1/ Asch, Solomon, "Studies of Independence and Conformity: I," Psychological Monographs, Vol. 70, Whole Number 416 (1956).

/2/ Cyphert, Frederick and William Gant, "The Delphi Technique." Presented to the American Educational Research Association, Annual Meeting, Minneapolis, Minnesota (February, 1970), Mimeo.

/3/ Dalkey, Norman, "The Delphi Method." Presented to the American Statistical Association, 131st Annual Meeting, Fort Collins, Colorado (August 23, 1971).

/4/ Frazer, James, The Golden Bough: Part III, The Dying God, London: Macmillan (1923).

/5/ Knight, Frank, Risk, Uncertainty and Profit, Boston: Houghton, Mifflin (1921).

/6/ Rescher, Nicholas, "Study of Value Change," Journal of Value Inquiry, Vol. 1 (1967).

/7/ Rescher, Nicholas, "A Questionnaire Study of American Values by 2000 A.D." in K. Baier and N. Rescher eds. Values and the Future, New York: Free Press (1969).

/8/ Sherif, Muzafer, An Outline of Social Psychology, New York: Harper and Brothers (1948).

/9/ Welty, Gordon, "Note sur l'optimisme des futurologues," Revue internationale des sciences sociales, Vol. 22 (1970).

/10/ Welty, Gordon, "A Critique of Some Long-range Forecasting Developments," Proceedings of the 38th Session of the International Statistical Institute (1971).