THE DELPHI TECHNIQUE AND JUDGMENTAL FORECASTING

THOMAS R. STEWART†

Environmental and Societal Impacts Group

National Center for Atmospheric Research, P. O. Box 3000, Boulder, CO 80307.

Abstract. The Delphi technique for judgmental forecasting by expert groups is described and the controversy surrounding its use is summarized. The technique clearly does not eliminate all unwanted psychological effects on group judgment. Furthermore, the design of most Delphi studies makes it impossible to separate the signal from the noise in expert judgment. A methodological standard for evaluating judgmental forecasts is proposed.

1. Introduction

The Delphi technique was developed at the Rand Corporation around 1950 as a method for eliciting expert opinion. Its purpose is to provide a practical means for obtaining the opinion of a group while avoiding the "biasing effects of dominant individuals, of irrelevant communications and of group pressure toward conformity" (Dalkey, 1969, p. 408). Long-range forecasting studies using some variation of the Delphi technique probably number in the thousands. Recent applications include studies of the future of fringe benefits (Baldwin, 1982), predicted breakthroughs in economics (Cicarelli, 1984), forecasting the recreational catch of spiny lobsters off Florida (Zuboy, 1981), the future of affirmative action (Fry, 1980) and forecasts of international trade and industrial cooperation (Paliwoda, 1983).

This chapter will briefly describe the Delphi technique and the controversy surrounding its use and will summarize the research on judgment which suggests that the results of any Delphi study are influenced by the particular procedures used. Improvement of judgmental forecasts through the use of several methods, including methods for making the basis for judgment explicit, is then discussed. The use of a multimethod approach to forecasting both exposes method-related errors and challenges the participants to explain and eliminate them.

---

† Current address: Center for Research on Judgment and Policy, University of Colorado, Campus Box 344, Boulder, Colorado, 80309.

Climatic Change 11 (1987) 97–113.

2. The Delphi method

Linstone and Turoff (1975) describe the Delphi technique as "... a method for structuring a group communication process so that the process is effective in allowing a group of individuals, as a whole, to deal with a complex problem" (p. 3). They go on to describe "structured communication" as involving

1. "some *feedback* of individual contributions of information and knowledge".
2. "some *assessment* of the group judgment or view".
3. "some opportunity for individuals to *revise views*".
4. "some degree of *anonymity* for the individual responses" (p. 3, emphasis added).

The "classical" Delphi involves the following steps (Martino, 1983):

1. A group of respondents, who have been selected for their expertise in a particular area, are asked individually to forecast important events or trends in that area.
2. The moderator combines the forecasts into a single set of events and the combined set is sent to each respondent.
3. Respondents are asked to forecast the date of occurrence of each event in the list.
4. The moderator summarizes the individual forecasts by computing the median date and the upper and lower quartiles for each forecast.
5. Respondents receive a new questionnaire listing the events and the medians and quartiles of the forecasts from the previous round. They are asked to prepare new forecasts, and, if their forecasts fall outside the upper or lower quartile boundaries, to provide reasons for their forecasts.
6. The moderator summarizes the forecasts and the reasons provided and prepares a revised questionnaire.
7. Respondents receive the revised questionnaire and are asked to take the reasons into account in preparing a new forecast. The median forecasts from this round for each event are the final forecasts.

A Delphi exercise is generally conducted by mail or by a computerized conferencing system in order to preserve the anonymity of the respondents. There have been numerous variations on the classical procedure (Martino, 1983; Preble, 1983), including providing an initial list of events to be forecast, changing the number of rounds, and requesting respondents to assign probabilities to their forecasts.

3. The Delphi dispute

Until the early 1970s, almost all the literature about the Delphi method was written by its proponents. Then a Rand-supported critique of Delphi by Harold Sackman (1975) set off an intense debate about the merits of the technique. The major objections to the Delphi technique were: a) the sloppy execution of many Delphi studies, b) pressure to conform and undue emphasis on consensus, c) the tendency to promote and legitimize shallow thinking and d) the unscientific nature of the technique. These will be discussed briefly.

3.1 Execution of the study

Both critics and proponents of Delphi have deplored the sloppy conduct of many Delphi studies. A few of the problems cited are: expert respondents that have been poorly chosen (Hill and Fowles, 1975; Preble, 1983); questionnaires that have been poorly worded, ambiguous, and too long (Hill and Fowles, 1975; Scheele, 1975; Linstone, 1975; Preble, 1983); and superficial analysis of responses (Linstone, 1975).

Proponents of Delphi readily admit that some studies have been poorly conducted. Linstone and Turoff (1975) describe the design of a Delphi study as an "art" and speculate that the apparent simplicity of the method itself tempts people to use it, even though they lack the appropriate skills and have an inadequate understanding of the problems involved. Sackman (1975) and Hill and Fowles (1975) lay much of the blame for poor technique on the poor background of most Delphi researchers in social science research.

In defending Delphi, proponents emphasize the distinction between technique and execution (Goldschmidt, 1975) and argue that problems in execution are largely avoidable (Linstone, 1975). They also argue that similar problems will be experienced with any technique.

3.2 Emphasis on consensus

Criticism of the Delphi method extended beyond problems of execution to the fundamental nature of the technique itself. A major criticism was its tendency to produce a false appearance of consensus among the respondents. Sackman (1975) describes social-psychological factors that may operate to produce pressure to conform. Bardecki (1984) reports evidence that social-psychological factors can influence Delphi results. He points out that measures of central tendency (*e.g.*, medians) provided in a Delphi exercise are powerful anchors that tend to narrow the range of responses on future rounds. He concludes that "unless the individual has great assurance and the issue is of considerable importance, there is reason to believe that any consensus will be at least in part a result of assimilative pressure rather than of any true education" (p. 283). His conclusion is supported by Rohrbaugh's (1979) study which found that, although Delphi groups produced a "consensus" result, a poll of individual respondents conducted after the Delphi study was completed showed no greater agreement than before it began.

Bardecki also investigated psychological influences on dropout rates among Delphi respondents and found relations between respondents' beliefs and their dropping out; respondents holding extreme views were more likely to drop out. These results suggest that the respondents who complete a Delphi exercise may not represent those who began it and that the impression of consensus may be partly due to attrition.

A related criticism of the Delphi method is that it tends to overemphasize consensus (Coates, 1975) and to equate consensus with validity. As Chan (1982) notes, consensus may represent "collective bias rather than wisdom" (p. 440). Although proponents are careful to distinguish between consensus and validity (*e.g.*, Linstone, 1975; Mitroff and

Turoff, 1975), it remains true that "*the validity of the resulting judgment of the entire group is typically measured in terms of the explicit 'degree of consensus' among the experts*" (Mitroff and Turoff, 1975, p. 22, emphasis original).

Critics and proponents of the Delphi method are widely separated on the issue of consensus. One of the original purposes of the technique was to reduce group pressure to conform, and this is often cited as one of its advantages (*e.g.*, Dalkey, 1969; Morgan *et al.*, 1979; Martino, 1983). At the same time, the convergence of opinion observed in many Delphi studies is often taken as an indication of the value of the method. Delphi critics, on the other hand, argue that it simply replaces the pressures of a conventional meeting with other, perhaps more subtle, pressures. They attribute the observed convergence of opinion to spurious influences.

3.3 Shallow thinking

Disagreement also exists regarding the kind of thinking that is promoted by a Delphi exercise. Proponents argue that it challenges respondents to think more deeply about problems (Coates, 1975; Goldschmidt, 1975; Paliwoda, 1983), makes their assumptions accessible (Scheele, 1975), exposes uncertainty and divergent views (Linstone, 1975), and facilitates effective communication among respondents (Linstone and Turoff, 1975; Martino, 1983). Critics argue that the method generates "snap answers to ambiguous questions" (Sackman, 1975, p. 73), simply reflects, and canonizes, conventional wisdom (Spinelli, 1983), suppresses uncertainty and divergent views, leading to a narrowing of concerns (Hill and Fowles, 1975), and has demonstrated no superiority over other communication techniques (Sackman, 1975; Rohrbaugh, 1979).

Sackman (1975) describes an "expert halo effect" which surrounds Delphi studies and which could both give excessive credence to the Delphi output and, at the same time, permit shallow thinking:

> The result of the expert halo effect for Delphi is to make no one accountable. The director merely reports expert opinion objectively according to prescribed procedure; he is not responsible or liable for outcomes. The panelist obligingly follows the ritual, protected at all points by faceless anonymity. The user can always claim that he was simply following the best advice available, and that he is not responsible for what the experts say. Everyone has an out, no one needs to take any serious risks, and no one is ultimately accountable (p. 36).

The differences between the two views could hardly be greater. One side sees the Delphi technique as the antidote to shallow, narrow, conventional thinking and the other side sees it as promoting and legitimizing just such thinking.

8.4 Unscientific technique

One way to settle the dispute surrounding the Delphi method would be to compare the performance of a large number of Delphi forecasts to that of alternative forecasting methods. Unfortunately, this has not been done and is not likely to be done soon because Delphi technique is used primarily for long-range forecasting. Furthermore, attempts to validate the predictions of older Delphi studies have been thwarted by the problem of confirmation (Hill and Fowles, 1975, p.185). The original questions were so ambiguous that it was impossible to determine whether the predicted events had occurred or not.

Experimental studies meant to validate the Delphi technique (*e.g.*, Dalkey, 1969; Riggs, 1983; Parente *et al.*, 1984) have inherent weaknesses that severely limit their usefulness. For obvious practical reasons, these studies use relatively short-term forecasts, such as the outcomes of upcoming sports events or political elections, or they use "almanac questions"–questions that have numerical answers that are unknown to the respondents. These studies also typically use non-expert respondents, generally college students. The usefulness of such studies for the evaluation of a method for making long-range forecasts by experts is questionable.

Several kinds of empirical evidence have been cited for the validity of the Delphi technique (Martino, 1983). For example, there is evidence that the arguments that are made during the rounds of a Delphi exercise have an effect on the final result. In addition, the distribution of first round responses exhibits some regularity. Furthermore, Delphi studies consistently attribute more uncertainty to predictions that are more remote in time. Critics argue that these kinds of evidence for the internal consistency of Delphi studies bear only indirectly, if at all, on the external validity of forecasts produced by the Delphi method (Pill, 1971; Hill and Fowles, 1975; Sahal and Yee, 1975).

The weakness of empirical evidence for evaluating the Delphi method is just one point made in support of the broader charge that Delphi is "unscientific." Sackman (1975) used standards developed by the American Psychological Association for educational and psychological tests and manuals as a basis for his indictment of this technique. He concludes, among other things that the Delphi method is "virtually oblivious to reliability measurement and scientific validation of findings" (p. 73), and that Delphi research has produced "virtually no serious critical literature to test basic assumptions and alternative hypotheses" (p. 74).

Although some of Sackman's points are conceded by Delphi proponents, their response to his critique is that he has missed the point. Delphi cannot be evaluated by scientific criteria because it is to be used only as a "method of last resort" when no adequate models exist (Linstone, 1975, p. 573). Coates (1975) states "Sackman ignores the crucial point that Delphi is not a scientific tool, nor is scientific experiment or a scientifically structured activity" (p. 193).

Mitroff and Turoff (1975) argue that those who accuse the Delphi method of being unscientific are arguing that knowledge can only be gained through a Leibnizian

inquiring system, that is, through development of formal, symbolic systems by purely rational, deductive logic. They argue that the Delphi method should be evaluated from the standpoint of other epistomological systems, in particular, the Lockean inquiring system which is based on the development of experimental, consensual systems through objective data and inductive logic. In other words, critics dismiss the Delphi method because it is not scientific while proponents reply that, scientific or not, we can gain knowledge by using it.

4. Evaluating the Delphi method

The lack of strong empirical evidence for validity and the failure to conform to the traditional requirements of scientific inquiry are not sufficient grounds for abandoning the Delphi technique, as Sackman suggests. The Delphi method was created, and survives, because important decisions depend on judgments about the future, and those decisions will be made whether science is ready or not. Decision makers cannot avoid judgmental forecasts on the grounds that they are unscientific. They can, however, choose how to forecast. They can decide to obtain an explicit forecast or to allow an implicit forecast to be embedded in the decision process. If they decide to obtain an explicit forecast, they can decide whom to consult–other decision makers, the public, or experts. They can allow the forecasters to deliberate in a conventional meeting, in a Delphi exercise, or in any of a number of other structured formats (see Armstrong, 1978; and Delbecq *et al.*, 1975; for examples).

Delphi is a method for obtaining an explicit judgmental forecast and can be evaluated only by comparing it to other judgmental forecasting methods that can be used when a forecast must be made despite insufficient data, inadequate models, and lack of time and resources for thorough scientific study.

5. The problem of judgmental forecasting

Before making a judgmental forecast, it is necessary to determine whose judgment should be used in the forecast. In some forecasting areas, however, particularly in the social sciences, there is so much uncertainty about the future that it is not clear that there are any real experts who have knowledge or experience to cope with that uncertainty. In other cases, it is difficult to determine what kinds of expertise are appropriate. For example, specialists may be so absorbed in their own subject matter that they do not foresee important changes in technology. Linstone (1975) cites the example of the experts on reciprocating engines who forecast in the 1930's that propeller aircraft would be standard until 1980.

As is the case with most procedural decisions in judgmental forecasting, selection of experts is context-dependent. It is likely that the "best" experts for a long-range forecasting study will combine the minimum level of training and experience needed to understand the problem and place reasonable limits on the forecast with the breadth necessary to contemplate the forecast in the context of other developments that might affect it. It is possible that a panel that included both specialists and generalists could

produce a good forecast, but that depends upon the problem and on how the communication among the panelists is structured.

Any judgmental forecast can be thought of as a combination of signal and noise. The signal is based on experience and knowledge and the appropriate use of that experience and knowledge to forecast the future and to accurately assess the inherent uncertainty about the future. The noise, or error, results from mistaken beliefs, misleading or irrelevant experience, inconsistencies, biases, psychological factors, social pressure, and other unwanted influences on both individual and group judgment that reduce forecast accuracy. The challenge for judgmental forecasting is first to choose experts to maximize the potential signal and then to enhance that signal, if possible, and separate it from the noise. An understanding of the judgmental process is indispensible for doing this successfully.

Research on the judgment process has increased rapidly during the last 15 years. The literature provides a catalog of biases and limitations which could be sources of noise in judgmental forecasts. Much of the research, typified by Tversky and Kaheneman's (1974) classic paper, presented such a dismal view of human cognitive capacity that some people began to wonder how humanity ever progressed as far as it did[1]. Research results repeatedly recount systematic biases in judgment, the inconsistent and poorly controlled nature of the judgment process, the pervasiveness of cognitive limitations that can reduce the validity of judgments, and the difficulty of overcoming those limitations. Irrelevant features of judgment tasks can strongly influence judgment while relevant information is often ignored or used inappropriately.

Partly because the results of judgment studies seemed at odds with the reality of human achievement and partly because it was recognized that most studies were conducted in the laboratory by confronting college students with contrived judgment tasks, some researchers began to question the generalizability of research on the judgment process (Ebbesen and Konecni, 1980) and to present a more positive view of human capabilities (Nisbet *et al.*, 1983). Kruglanski *et al.*, (1984) found that some of the biases identified by Tversky and Kahneman disappeared with slight changes in the instructions given to the subjects. Levin *et al.*, (1983), on the other hand, argue that many of the results of laboratory research do generalize to situations outside the laboratory.

Seemingly conflicting results in judgment research can be understood if one is careful to distinguish between competence and performance and to recognize that performance depends on many context-specific factors. It appears that performance is not governed by a single judgment process that can be described by a few basic principles

[1] For reviews of the literature on judgment, see Slovic and Lichtenstein (1971), Slovic Fischhoff and Lichtenstein (1977), Hammond *et al.*, (1980), Hogarth (1980), Einhorn and Hogarth (1981), and Kahneman *et al.*, (1982). For reviews of judgment research specifically applied to forecasting, see Kaheneman and Tversky (1979), Hogarth and Makridakis (1981), Einhorn and Hogarth (1982), Evans (1982), and Sjoberg (1982).

and laws (Ebbesen and Konecni, 1980). Instead, particular tasks evoke different judgment processes, or strategies, depending upon the judge and the context for judgment. In some cases the strategies evoked are appropriate; in others they are not. "In cognition, as in perception, the same mechanisms produce both valid and invalid judgments" (Tversky and Kahneman, 1983, p. 313).

Unfortunately, we do not yet understand the process which governs the evoking of judgment strategies in specific situations (although Hammond and his colleagues at the Center for Research on Judgment and Policy are working toward a theoretical framework which may help in this regard, *e.g.*, Hammond, 1981). Although some judgmental biases seem robust (Fischhoff, 1982; Tversky and Kahneman, 1983), the research needed to predict accurately when and where judgmental limitations and biases will appear in a practical situation has not progressed very far (see Fischhoff, 1983, for an example of such research that met with limited success). We know that seemingly inconsequential changes in the wording or the order of the questions, in the mode of response, and mode of presenting information can affect judgment (see Stewart and Glantz, 1985, for example), but we cannot predict accurately when those effects will occur or how strong they will be. In other words, we know that method can have powerful effects on a judgmental forecast, but we do not yet know how to predict or minimize these effects.

5.1 Implications for the Delphi method

Inconsistencies and biases in judgment are known to exist, but the conditions that produce them are not well understood. Furthermore, judgments are often sensitive to irrelevant and seemingly inconsequential features of the method used to elicit them. As a result, it is not possible to examine a judgmental forecasting method and predict the effect of the method on the forecast in a particular context. Two important conclusions can, however, be drawn about Delphi studies. First, the results of a Delphi study are probably highly method-specific. Another method used on the same problem would probably produce different results. Second, it is not possible to ascertain whether the method used in a particular Delphi study has enhanced the signal and reduced the noise, or whether the reverse is true. The methods used in a particular Delphi study for eliciting judgments, providing feedback, and structuring communication can introduce noise into the results, and the amount of noise introduced, relative to the signal, is unknown, but is potentially large.

Delphi proponents might argue that the noise in a Delphi forecast can be minimized by careful design and conduct of the study by an experienced investigator who is schooled in the "art" of Delphi studies and is aware of its many pitfalls (Linstone, 1975). This may be true, but, since the level of noise depends on the complex and poorly understood interactions among properties of the forecasting context, characteristics of the respondents, and formal and substantive properties of the forecasting problem, even the most professionally designed and managed Delphi forecast may produce results that are largely artifacts of the method.

Furthermore, the lack of adequate standards for conducting a Delphi exercise, the apparent simplicity of the method, and the track record of misuse remain serious concerns to potential Delphi users. The results of a Delphi study may be the product of the creativity and ingenuity of a skilled practitioner or of the misconceptions and stumbling of an ill-informed novice, but there is no easy way to tell the difference.

6. Improving judgmental forecasts

The Delphi technique or any other judgmental forecasting method may produce results that reflect errors introduced by the method. As a result, decision makers should not rely on any single method to produce judgmental forecasts. Instead, they should insist on studies that employ a combination of methods.

A wide range of methods that could be used in judgmental forecasting exist or can be designed. Some methods may elicit highly analytic thinking while others may be conducive to an intuitive approach (Hammond, 1981). Methods vary in the amount of formal structure imposed on the expert and in the amount and kind of explicit justification required for judgments. Questions can be worded or "framed" (Tversky and Kahneman, 1981) in different ways. Responses can be made in different formats. Feedback and communication among respondents can be structured in different ways. All these variations can affect results.

The extent to which a forecast can be improved by the use of multiple methods depends upon which methods are chosen and how the conflicting results of different methods are reconciled. The most diverse possible set of methods should be used, subject to constraints imposed by the problem context and available resources. Then, differences in results produced by different methods must be reconciled as much as possible. Although the methods may be applied iteratively, a "method-Delphi" which simply reported the median of the results of several methods would not be an acceptable procedure. The reconciliation process will be highly context-specific and will require the effort of experts in both the subject matter and the methods themselves. In many cases, total reconciliation of conflicting results obtained by multiple methods will be impossible and a range of results, reflecting inherent uncertainty in the problem, will have to be reported.

7. Externalization

An additional requirement for a methodologically acceptable judgmental forecasting study is that it include methods designed to make the reasons for judgments explicit. The basis for such methods is decomposition and explicit structuring of the forecasting problem and externalization of the process of arriving at a judgment. Methods developed through judgment research have been found useful in describing, or "externalizing," the process of making a judgment, thus exposing that process to critical examination (Hammond, 1976). These methods are the basis for judgment aids which can be used to improve forecasts. Descriptions of such judgment aids can be found in Hammond *et al.*, (1980) and Hogarth (1980).

Different externalization methods influence the judgment process in different ways, tend to externalize different aspects of the judgment process (Einhorn *et al.*, 1979), decompose the problem in different ways and have different strengths and limitations. Methods for externalizing the judgment process can be grouped into two general categories—descriptive methods and prescriptive methods. Descriptive methods are intended to develop models of the process that a judge used to make a judgment. Prescriptive methods are used to develop models of how, according to some normative theory, a person should make a judgment.

Many forecasters and Delphi investigators seem unaware of methods developed in judgment and decision research that can be used in forecasting. Although it is not possible here to fully describe and illustrate such methods, it is hoped that the following brief outline of some prominent externalization methods and sources of further information will help forecasters make more informed choices.

7.1 Descriptive methods based on statistical analysis

These methods use statistical procedures to analyze the relation between the judgment and the information, or "cues," on which the judgment is based. The analysis requires data in the form of a sample of judgments made in response to a number of scenarios containing cues which are known or can be measured. Multiple regression analysis is a common statistical procedure for judgment modeling (Hammond *et al.*, 1975). The judgment is regressed on a set of cues and the resulting model describes the relative importance of each cue, the functional relation between the cue and the judgment, and (in some cases) the principle by which the cues are organized into a judgment (*e.g.*, additive or multiplicative).

7.2 Descriptive methods based on verbal reports

The judge's own verbal description of his or her judgment process can provide data for modeling the process (Ericsson and Simon, 1980). A prominent example of the use of verbal reports to develop judgment models is the "process tracing" approach of Kleinmuntz (1968). The "cognitive mapping" method of Axelrod (1976) is a related approach which employs a content analysis of text or verbal statements to develop a qualitative model of the judgment process.

7.3 Prescriptive methods

These methods do not claim to describe how people naturally make judgments or decisions. Their intent is to replace the natural process with a better process. In doing so, however, a model of the better process is developed and that model externalizes the basis for the better judgment. Both prescriptive and descriptive approaches are highly reactive, that is, they change the judgment process being studied (unless, in the descriptive approach, data can be obtained from the judge unobtrusively so that he or she does not know that the judgments are being analyzed). In the descriptive approach, such changes are an undesirable side effect. In the prescriptive approach, the change is intended.

Perhaps the best known prescriptive approach is decision theory (see Raiffa, 1968, and Keeney, 1982, for overviews). Decision theory is based on a well developed theory of rationality which says that decisions should be based on the maximization of expected utility. Decision analysts have developed elaborate procedures for decomposing decision problems into probabilities and utilities, assessing the decision-makers' subjective probabilities and utilities, and recombining them to evaluate decision options. For a description of decision analysis methods, see Keeney and Raiffa (1976).

While decision theory is based on a theory of rationality, another prescriptive approach has developed from descriptive research on the judgment process. A number of judgment studies using the regression approach described above have found that complex judgments can be described by simple linear models involving only a few variables. Several studies also showed that the linear models are more accurate than the original judgments from which they were derived, and Camerer (1981) described the conditions under which this can occur. Still other studies showed that simple weighted sums of variables generally outperform judges and that, in many cases, it does not matter much what the weights are–equal weights for the variables will do nicely. Dawes and Corrigan (1974) turned this finding into a prescription for decision-making: "the whole trick is to decide what variables to look at and then to know how to add" (p. 105). (It should be obvious that this or any other method described in this paper is not universally applicable.)

8. Critical examination of judgmental models

All externalization methods are simply methods for developing explicit models of the judgment process. Modeling the judgment process is similar to modeling the environment: both require decomposing the problem into elements and developing an explicit structure that describes the relations among the elements. Similar modeling methods (*e.g.*, regression analysis) may be used to develop both types of models. Both types of models are models of how people think about the environment. In the development of environmental models, however, the emphasis is on modeling the environment rather than modeling the thinking about it, and the data for the model is drawn, as much as possible, from objective data about the environment. In the development of judgmental models, the emphasis is on modeling the thinking, and the data for the model is provided subjectively by the judge (objective data about the environment influence the judgment model indirectly through their influence on the judge).

Once externalized judgment models have been developed, they can be critically examined–an important step that is missing from most Delphi studies. Models can be compared with relevant data and theory in order to identify substantive errors or omissions. Judgment models can also be compared to relevant normative models (*e.g.*, logic, probability, statistics, decision theory, linear models) to identify inconsistencies, biases, or other errors in the use of information to make a judgment.

Examination of externalized judgment models can be undertaken by the judges themselves, by other substantive experts, or by normative experts such as judgment

researchers, philosophers, statisticians, logicians, or social scientists. Ideally, the result of this step is a set of corrected models free of unwanted influences and omissions.

Once known biases and substantive errors have been eliminated, remaining differences between models for different judges and methods are due to differing interpretations of the evidence, different ways of using the same evidence to make an inference, different ways of decomposing and structuring the problem, different judgment processes evoked by various methods, or different values. Although judgment research was useful in externalizing these differences, reconciliation of models must be left to the expert judges, the forecaster, and, ideally, also to the potential user of the forecast. Psychological research on group processes (*e.g.*, Steiner, 1972) and on group judgment and decision-making (*e.g.*, Rohrbaugh, 1979; Fischer, 1981) may provide some helpful guidance in conducting the deliberations necessary to reconcile the differences.

The approach described above for improving judgment is similar to the approach used by Hammond *et al.*, (1983) to improve scientists' judgments of cancer risk. Hammond and Adelman (1976) used judgment research methods to externalize differences among ballistics experts with regard to judgments of the safety, injury, and threat to bystanders of handgun ammunition. Some applications of decision theory are similar in spirit to this approach. The use of several externalization methods and reconciliation of the differences among models elicited by different methods has not, to my knowledge, been used in a practical application. Einhorn *et al.*, (1979) did compare two externalization methods in a laboratory study.

9. Conclusion

The Delphi method was designed to structure communication in order to eliminate some of the psychological effects that limit the effectiveness of traditional discussion groups. It clearly does not eliminate all psychological effects on group judgment. Some psychological characteristics of interacting groups remain and some new ones are introduced. Furthermore, the Delphi technique is subject to many unwanted influences on individual judgments.

The extent of the influence of unwanted psychological effects on Delphi results is dependent on the respondents, the problem, and the context. Our understanding of the judgment process is not sufficient to detect and eliminate all unwanted effects. Delphi studies can be misleading because the extent of noise in the results is unknown. The danger is compounded when the mistique of the Delphi process confers legitimacy on the results.

When judgmental forecasts are necessary, a multi-method approach to forecasting, including methods for making the reasons for judgment explicit, should be used. This requirement suggests a methodological standard that can be used to evaluate judgmental forecasts, whether produced by Delphi or some other method:

An expert judgment forecast is likely to be useful to the extent that the designers a) have used a variety of procedures in order to expose method-induced error and b) have included methods designed to make the basis for judgment explicit and to challenge respondents to think more deeply through critical examination of the basis for their judgments.

Such a standard is consistent with the spirit of the Delphi method as described by its major proponents. For example, Goldschmidt (1975) stresses that Delphi studies must explore the underlying assumptions that produced the forecast. Nevertheless, few, if any, actual studies meet the standard. Since the methods for improving judgmental forecasting are available, the "traditional" Delphi technique, used alone, is not acceptable for serious forecasts.

Acknowledgments

I wish to thank Mary Downton, Dale Jamieson, Rick Katz, Will Kellogg, Maria Krenz, and Jonathan Taylor for their helpful comments on an earlier version of this paper.

References

Armstrong, J. Scott: 1978, *Long-Range Forecasting from Crystal Ball to Computer*, New York, Wiley.

Auboy, J. R.: 1981, 'A New Tool for Fishery Managers: The Delphi Technique', *North American Journal of Fisheries Management*, **1**, 55–59.

Axelrod, R. (Ed.): 1976, *Structure of Decision: The Cognitive Maps of Political Elites*, Princeton, Princeton University Press.

Baldwin, G. H.: 1982, 'The Delphi Technique and the Forecasting of Specific Fringe Benefits', *Futures*, **14**, 319–325.

Bardecki, M. J.: 1984, 'Participants' Response to the Delphi Method: An Attitudinal Perspective', *Technological Forecasting and Social Change*, **5**, 281–292.

Camerer, C. F.: 1981, 'General Conditions for the Success of Bootstrapping Models', *Organizational Behavior and Human Performance*, **27**, 411–422.

Chan, S.: 1982, 'Expert Judgments Under Uncertainty: Some Evidence and Suggestions', *Social Science Quarterly*, **63**, 428–444.

Cicarelli, J.: 1984, 'The Future of Economics: A Delphi Study', *Forecasting and Social Change*, **25**, 139–157.

Coates, J. F.: 1975, 'In Defense of Delphi', *Technological Forecasting and Social Change*, **7**, 193–194.

Dalkey, N.: 1969, 'An Experimental Study of Group Opinion: The Delphi Method', *Futures*, **1**, 408–426.

Dawes, R. M., and B. Corrigan: 1974, 'Linear Models in Decision Making', *Psychological Bulletin*, **81**, 95–106.

Delbecq, A. L., A. H. Van de Ven, and D. H. Gustafson: 1975, *Group Techniques for Program Planning*, Glenview, IL, Scott, Foresman.

Ebbesen, E. B., and V. J. Konecni: 1980, 'On the External Validity of Decision-Making Research: What Do We Know About Decisions in the Real World?', In S. Wallsten (Ed.), *Cognitive Processes in Choice and Decision Behavior*, Hillsdale, New Jersey, Erlbaum.

Einhorn, H., and R. M. Hogarth: 1981, 'Behavioral Decision Theory: Processes of Judgment and Choice', *Annual Review of Psychology*, **32**, 53–88.

Einhorn, H. J., and R. M. Hogarth: 1982, 'Prediction, Diagnosis, and Causal Thinking in Forecasting', *Journal of Forecasting*, **1**, 23–36.

Einhorn, H. J., D. N. Kleinmuntz, and B. Kleinmuntz: 1979, 'Linear Regression and Processing-Tracing Models of Judgment', *Psychological Review*, **86**, 465–485.

Ericsson, K. A., and H. A. Simon: 1980, 'Verbal Reports as Data', *Psychological Review*, **87**, 215–251.

Evans, J. St. B. T.: 1982, 'Psychological Pitfalls in Forecasting', *Futures*, **14**, 258–265.

Fischer, G. W.: 1981, 'When Oracles Fail—A Comparison of Four procedures for Aggregating Subjective Probability Forecasts', *Organizational Behavior and Human Performance*, **28**, 96–110.

Fischhoff, B.: 1982, 'Debiasing', In D. Kahneman, P. Slovic, and A. Tversky (Eds.), *Judgment Under Uncertainty: Heuristics and Biases*, New York, Cambridge University Press.

Fischhoff, B.: 1983, 'Predicting Frames', *Journal of Experimental Psychology: Learning Memory and Cognition*, **9**, 103–116.

Fry, F. L.: 1980, 'The End of Affirmative Action', *Business Horizons*, **23**, 34–40.

Goldschmidt, P. G.: 1975, 'Scientific Inquiry or Political Critique?', *Technological Forecasting and Social Change*, **7**, 195–213.

Hammond, K. R.: 1976, 'Externalizing the Parameters of Quasirational Thought', In M. Zeleny (Ed.), *Multiple Criteria Decision Making, Kyoto 1975*, New York, Springer-Verlag, 1976.

Hammond, K. R.: 1981, 'Principles of Organization in Intuitive and Analytical Cognition', University of Colorado at Boulder, Center for Research on Judgment and Policy, Report No. 231.

Hammond, K. R., and L. Adelman: 1976, 'Science, Values, and Human Judgment', *Science*, **194**, 389–396.

Hammond, K. R., B. F. Anderson, J. Sutherland, and B. Marvin: 1984, 'Improving Scientists' Judgments of Risk', *Risk Analysis*, **4**, 69–78.

Hammond, K. R., G. H. McClelland, and J. Mumpower: 1980, *Human Judgment and Decision Making: Theories, Methods, and Procedures*, New York, Praeger.

Hammond, K. R., T. R. Stewart, B. Brehmer, and D. O. Steinman: 1975, 'Social Judgment Theory', In M. F. Kaplan and S. Schwartz (Eds.), *Human Judgment and Decision Processes*, New York, Academic Press.

Hill, K. Q. and J. Fowles: 1975, 'The Methodological Worth of the Delphi Forecasting Technique', *Technological Forecasting and Social Change*, **7**, 179–192.

Hogarth, R. M.: 1980, *Judgement and Choice: The Psychology of Decision*, Chichester, England, Wiley.

Hogarth, R. M. and S. Makridakis: 1981, 'Forecasting and Planning: An Evaluation', *Management Science*, **27**, 115–138.

Kahneman, D., and A. Tversky: 1979, 'Intuitive Prediction: Biases and Corrective Procedures', In S. Makridakis and S. C. Wheelwright (Eds.), *Forecasting, TIMS, Studies in Management Science*, **12**, 313–327.

Kahneman, D., P. Slovic, and A. Tversky (Eds.): 1982, *Judgment under Uncertainty: Heuristics and Biases*, New York, Cambridge University Press.

Keeney, R. L.: 1982, 'Decision Analysis: An Overview', *Operations Research*, **30**, 803–838.

Keeney, R. L., and H. Raiffa: 1976, *Decisions with Multiple Objectives: Preferences and Value Tradeoffs*, New York, Wiley.

Kleinmuntz, B. (Ed.): 1968, *Formal Representation of Human Judgment*, New York, Wiley.

Kruglanski, A. W., N. Friedland, and E. Farkash: 1984, 'Lay Persons' Sensitivity to Statistical Information: The Case of High Perceived Applicability', *Journal of Personality and Social Psychology*, **46**, 503–518.

Levin, I. P., J. J. Louviere, A. A. Schepanski, and K. L. Norman: 1983, 'External Validity Tests of Laboratory Studies of Information Integration', *Organizational Behavior and Human Performance*, **31**, 173–193.

Linstone, H. A.: 1975, 'Eight Basic Pitfalls: A Checklist', In Linstone, H. D., and M. Turoff (Eds), *The Delphi Method: Techniques and Applications*, Reading, MA, Addison-Wesley, 573–586.

Linstone, H. A. and M. Turoff: 1975, 'Introduction', In Linstone, H. D., and M. Turoff (Eds), *The Delphi Method: Techniques and Applications*, Reading, MA, Addison-Wesley, 3–12.

Martino, J. P.: 1983, *Technological Forecasting for Decision Making*, (Second edition), New York, North-Holland.

Mitroff, I. A. and M. Turoff: 1975, 'Philosophical and Methodological Foundations of Delphi', In Linstone, H. D., and M. Turoff (Eds), *The Delphi Method: Techniques and Applications*, Reading, MA, Addison-Wesley, 17–36.

Morgan, D. R., J. P. Pelissero, and R. E. England: 1979, 'Urban Planning: Using Delphi as a Decision-Making Aid', *Public Administration Review*, **39**, 380–384.

Nisbett, R. E., D. H. Krantz, C. Jepson, and Z. Kunda: 1983, 'The Use of Statistical Heuristics in Everyday Inductive Reasoning', *Psychological Review*, **90**, 339–363.

Paliwoda, S. J.: 1983, 'Predicting the Future Using Delphi', *Management Decision*, **21**, 31–38.

Parente, F. J., J. K. Anderson, P. Myers, and T. O'Brien: 1984, 'An Examination of Factors Contributing to Delphi Accuracy', *Journal of Forecasting*, **3**, 173–182.

Pill, J.: 1971, 'The Delphi Method: Substance, Context, a Critique and an Annotated Bibliography', *Socio-economic Planning Sciences*, **5**, 57–71.

Preble, J. F.: 1983, 'Public Sector Use of the Delphi Technique', *Technological Forecasting and Social Change*, **23**, 75–88.

Raiffa, Howard: 1968, *Decision Analysis: Introductory Lectures on Choices Under Uncertainty*, Reading, MA, Addison-Wesley.

Riggs, W. E.: 1983, 'The Delphi Technique: An Experimental Evaluation', *Technological Forecasting and Social Change*, **23**, 89–94.

Rohrbaugh, J.: 1979. 'Improving the Quality of Group Judgment: Social Judgment Analysis and the Delphi Technique', *Organizational Behavior and Human Performance*, **24**, 79–92.

Sackman, H.: 1975, *Delphi Critique*, Lexington, MA, D.C. Heath and Company.

Sahel, D. and K. Yee: 1975, 'Delphi: An Investigation from a Bayesian Viewpoint', *Technological Forecasting and Social Change*, **7**, 165–178.

Scheele, D. S.: 1975, 'Consumerism Comes to Delphi', *Technological Forecasting and Social Change*, **7**, 215–219.

Sjoberg, L.: 1982, 'Aided and Unaided Decision Making: Improving Intuitive Judgment', *Journal of Forecasting*, **1**, 349–363.

Slovic, P., B. Fischhoff, and S. Lichtenstein: 1977, 'Behavioral Decision Theory', *Annual Review of Psychology*, **28**, 1–39.

Slovic, P., and S. Lichtenstein: 1971, 'Comparison of Bayesian and Regression Approaches to the Study of Information Processing in Judgment', *Organizational Behavior and Human Performance*, **6**, 649–744.

Spinelli, T.: 1983, 'The Delphi Decision-Making Process', *The Journal of Psychology*, **113**, 73–80.

Steiner, I. D.: 1972, *Group Processes and Productivity*, New York, Academic Press.

Stewart, T. R., and M. H. Glantz: 1985, 'Expert Judgment and Climate Forecasting: A Methodological Critique of "Climate Change to the Year 2000" ', *Climatic Change*, **7**, 159–183.

Tversky, A., and D. Kahneman: 1974, 'Judgment Under Uncertainty: Heuristics and Biases', *Science*, **185**, 1124–1131.

Tversky, A., and D. Kahneman: 1981, 'The Framing of Decisions and the Rationality of Choice', *Science*, **211**, 453–458.

Tversky, A., and D. Kahneman: 1983, 'Extensional *vs.* Intuitive Reasoning: The Conjunction Fallacy in Probability Judgment', *Psychological Review*, **90**, 293–315.