# PREDICTING THE FUTURE

N. C. Dalkey

October 1968

P-3948

# PREDICTING THE FUTURE*

N. C. Dalkey

The RAND Corporation, Santa Monica, California

What is going to happen twenty or fifty years from now has become of interest to planning agencies throughout government, industry and business. There are several reasons. The most important is undoubtedly the looming disasters that could destroy this nation if steps are not taken to sidestep them—nuclear war, overpopulation of the world, pollution and depletion of the atmosphere, mounting racial conflict...if at all possible we want to be aware of such dangers, and to find strategies to minimize them. But there are other reasons. Science, technology, and advanced administrative methods are rapidly extending the scope of public and private control over economic and social developments. To use these new powers wisely, it is necessary to have a clearer picture of the future—or more exactly of the possible futures. And for business and industry, the realm of competition is moving ever farther into the future.

---



This paper was prepared for presentation at the National Conference on Fluid Power, Chicago, Illinois, October 17, 1968.

There are many fascinating challenges to the wisdom of decision-makers produced by this swift extension of horizons. But today I want to concentrate on one—how to clear away the mists clouding events that have not yet happened.

The notion that the future is hidden—that prediction is the realm of seers, necromancers, and other unsavory types—is part of our cultural heritage. It makes the engineer qualify any comment about twenty years from now with great caution; and when it comes to predicting social change, it keeps the eye of the social scientist on day after tomorrow. Of course, there are better reasons than the traditional distrust of the fortune teller for circumspection. Technological breakthroughs, or major social events, involve enormously complex processes for most of which we have only a dim understanding. We do not have neatly packaged theories, or formulae in handbooks, for generating synthetic histories ahead of time. For some areas there are partial models. We can draw the envelope of the curves of increasing performance of aircraft and extrapolate on an accelerating slope and be fairly confident that we have traced out the potentialities over the next twenty years. A fairly complete model exists for drawing the graph of changing population, accepting some shaky guesses at what size families women will demand over the years. The increase in Gross National Product has been so steady (and so astonishing!) over the past two decades that economists no longer blanch at estimating even more astonishing growth for the next two or three decades. These, and a few other indices—increase in urbanization, education levels, and the like—are a few faint beams lighting tiny ovals in the dark ahead.

For the vast majority of possible events, and above all, for putting the synthetic history together in a coherent fashion, we must still rely on the judgment of knowledgeable men.

Now the word "judgment" is a little tricky. If we take a hard look at the kind of information which goes into prediction, we can see at one extreme information that is highly substantiated by solid evidence of one sort or another. That kind of information, where there is very good reason for believing it, is called knowledge. The prototype is the well organized, and experimentally confirmed material of the natural sciences. However, it is not necessary to have the highly developed structure of a science to have information that is solid and dependable. Many of the items called "common sense" are just as reliable. For example if I hold a pencil out, and assert that if I release it, it will fall to the platform, that assertion is verified by countless day-to-day experiences with gravity and does not depend upon Newton's laws of motion for its believability. A very large amount of the prescriptions of practical technology are equally solid. Scientists are often likely to forget that the steam engine came before thermodynamics and the telephone before information theory. The technologist's criterion for a machine—does it work?—is at least as fruitful in eliminating unfounded ideas as the scientist's criterion—is it confirmed by experiment?

At the other extreme is material that has little or no foundation—usually called speculation. Policymakers in industry and government would prefer, of course, that no such material get into their decisions. However, it seems unlikely that any major policy decision—especially one with a long-term horizon—gets made without some admixture of speculative inputs.

In between is a grey area that has no good name. One has to regret the common tendency to dichotomize—to refer to everything that is not knowledge as speculation. That leaves out the vast mass of information for which there is some evidence, but not enough to say it is solid. I call this intermediate material opinion. In the literature of long range forecasting, this kind of material is often called "wisdom" or "insight" or "informed judgment" or "experience." From the point of view of the amount of supporting evidence these are merely flattering ways to say "opinion."

From the definition, you can see that this very crude scale of merit is closely related to the likelihood of a statement being true. For knowledge the likelihood is high. For speculation it is low. And for opinion it is middling. That means—and let me stress this because it is vital—in the area of opinion, no matter how plausible, and no matter how prestigious the speaker, there is always a significant probability that what is being said is false.

Unfortunately, we do not have a good practical measure for degree of confirmation. It would be a great advance in prediction technology if we could say just where a given assertion stood in the scale of reliability. You are all familiar with the Moh's hardness scale. We could make good use of even such a crude scale for the hardness of information. I hesitate to think what would happen to political debate if a candidate could be challenged to scratch a good hard fact with his opinions.

At all events, we do not have the scale at the present. But we do have a rough feeling for the scale. I take for granted that in the area of knowledge ways of dealing with information are in pretty good order. That doesn't mean there aren't some difficulties. If you read the letters to the editor of a journal such as Science, you know that one scientist's "reasonable extension" is another's "unwarranted assumption."

About the area of speculation I have nothing wise to say. Like most decision makers, I wish it would go away. But it won't. There are very real problems there. They need attention.

In the case of opinion there is a great deal to be said. There are some traditional ways of dealing with opinion that are as old as recorded history, and probably much older. Most of these are some variant of the rule "two heads are better than one," or, more generally, "several heads are better than one." Committees, commissions, boards, panels, juries, the voting public are but a few examples of a technique that has ramified through all areas where decisions must be made on partial information.

The basis of the two heads rule is not hard to find. It certainly is the case that on any subject, there is more information—or at least as much information—in any two heads as there is in one of them alone. Unfortunately, it is equally true that in two heads there is at least as much misinformation as there is in one. And there is no certainty that there is a way of extracting just the information from the two and putting it together to form a single more reliable opinion. You may be extracting and putting together the misinformation.

The traditional procedure for putting the information in several minds together is face-to-face discussion. Now there have been a large number of experiments in recent years by small group psychologists that have shown that there are serious problems with face-to-face interaction.

In fact, they have verified a great deal of the lore about committees. As you know, a committee is something which, when you want a horse, designs a camel. The trouble with such lore is that much of it is true, but also amusing, and therefore can be ignored with a smile.

More to the point, some recent experiments we have performed at The RAND Corp. indicate that when opinions are involved, face-to-face discussion may, more often than not, result in a group opinion which is less accurate than simply the average of the individual opinions without discussion.

In these experiments, some of the biasing effects that have stood out are: (1) The influence of dominant individuals. Often, the group opinion is essentially determined by the opinion of the individual who talks the most, and there is little relation between volubility and ability. (2) Noise. A group is a social situation, and much of the interchange is more involved in maintaining the group than in furthering the problem to be solved. (3) Group pressure for conformity. The timid member may have little or no influence on the group, irrespective of his competence.

Notice, I emphasized that these considerations are important with respect to the formulation of a group opinion. In the area of speculation, unless there is enough information within the group to bring the judgment to at least the region of opinion, putting together several speculations at best creates a negligible improvement. In the area of knowledge, there are more powerful ways to employ information—systems analysis, program budgeting, and the like. Group interaction can perform a desirable function in the transfer of information, when what is being transferred is knowledge; the problems arise when what is being transferred is opinion.

To get around the difficulties with face-to-face interaction that I mentioned, some of us at RAND have over the years developed a set of procedures that have come to be known as Delphi. The name was proposed by the philosopher Abraham Kaplan. In some ways it is unfortunate—it connotes something oracular, something smacking a little of the occult—whereas as a matter of fact, precisely the opposite is involved; it primarily is concerned with making the best you can of a less than perfect fund of information. The basic characteristics of the Delphi procedures are: (a) Anononymity, (b) Iteration with controlled feedback, (c) Statistical group response. Anonymity is achieved by using questionnaires or other formal channels of communication, where specific responses are not associated with individual members of the group. This is a way of cutting down on the effects of dominant individuals and reducing group pressure. Iteration consists in performing the interaction among members of the group in several stages;

typically, at the beginning of each stage the results of the previous stage are summarized and fed back to the members of the group, and they are then asked to reasses their answers in light of what the entire group thought on the previous round. Controlled feed-back allows interaction with a large reduction in noise. Finally, rather than asking the group to arrive at a common opinion, a consensus, the group opinion is taken to be a statistical average of the final opinions of individual members of the group. In the experiments we have conducted, the median opinion——that is the middle estimate where half the group is on one side, and half on the other——has turned out to be the most accurate. By using a statistical group opinion, group pressure toward conformity is further reduced, and probably more important, the opinion of every member is reflected in the group response.

Before discussing the properties of this procedure I should say a few words about applications. As I remarked at the beginning of my talk, a growing segment of industry and government is concerned with extending the horizon of policy planning. In the past three years a large number of industrial firms have begun to apply some of the techniques that have been developed in the Delphi work. The list includes Thompson Ramo Wooldridge, IBM, Sandia Corp., The National Industrial Conference Board, Xerox, and numerous other. Some of you may have encountered the

Futures game devised by Helmer and Gordon and distributed by Kaiser Aluminum. The point of this is that interest in, and application of, these techniques is growing at a rapid rate. The widespread interest has led, in fact, to the formation of a new organization, The Institute for the Future, dedicated to generating long range forecasts of technological and social developments.

The question whether there is anything in the technique beyond a rather rapid and convenient way to skim the cream off the top of the heads of some knowledgeable people becomes of immediate practical import.

With something like this in mind, we undertook an intensive series of experiments at RAND, beginning this spring, to see if we could get a clearer picture of what is going on in the group formulation of opinion, and in particular, to evaluate whether the questionnaire, controlled feedback procedure led to an improvement of group judgments. We used upper class and graduate students from UCLA as subjects. They were paid for participating.

The basic task for the experiment was answering twenty questions of an almanac sort. Typical questions were "How many women Marines were there at the close of World War II? "How many telephones were in operation in Africa in 1967? "How many votes did Kennedy receive in the 1960 presidential election in the state of Texas?" We used this kind of question for two reasons. First of all, we wanted questions where the subjects did not know the answers, but we did, so we would have a precise check on accuracy. Secondly, we wanted questions where the subjects, even though they did not know the exact answers still had enough general information relevant to the answers so that they could make an informed estimate. You will notice that is precisely the nature of opinion—not enough is known to determine the answer, but enough is known so that something better than a sheer guess can be made.

The basic procedure was rather austere. A group of subjects were asked to answer individually twenty questions of the almanac sort. On the second round each subject was informed concerning the median—the middle response—of the group's answers on the first round, and the limits within which 50% of the answers lay. They were then asked, in light of the group answer on the first round, to reassess their individual answers. This procedure might be repeated for several rounds of feedback and reassessment. A typical experiment would involve thirty subjects, a control group of 15 who underwent the basic procedure, and an experimental group who performed some variant of it.

The general results of these experiments can be summarized in this way: On the opening round, typically there was a wide spread of answers. On succeeding rounds, the answers came closer and closer together—the group "converged" toward a common answer. However, the convergence was never complete, there was still a significant spread after several rounds. Most important, the group response—defined as the median of the individual responses—did tend to become more accurate on iteration. Not always. In about two-thirds of the cases the group response improved. In about one-third of the cases it became less accurate.

In contrast to the results of our experiments with face-to-face interaction, then, the Delphi procedures more often than not improved the group response.

This is a happy result, and, we believe, gives a firmer ground for use of the procedures in decision areas where it is necessary to rely on opinion.

In addition, we have gained some insight into the nature of opinion—it turns out to be a different kind of thing from knowledge in more ways than one—and we have found some illuminating items concerning social interaction that raise large questions about the ways we now make decisions.

I pointed out that one of the results of the Delphi exercises was that on the opening round there was a large spread in the estimates. The variation was so large from question to question that it was necessary to examine over 4000 answers spread over 250 questions before a clear pattern emerged. With this very large set of data, the picture of the kinds of responses that can occur settled down to a neat and well-defined pattern—in fact it resembled a diagram well-known to communications engineers—the log-normal curve. What the curve told us was that underneath the erratic replies of our subjects there is an underlying order, and this order is similar to a communication process. It also gave us some feeling for the <u>range</u> of responses, which turns out to be truly astonishing. It is not unusual for answers to the same question to differ by a factor of 1000 or 10,000. It also told us that in answering questions of the almanac sort, the subjects were dealing in ratios—they were thinking in terms of orders of magnitude, of relative size of numbers, not in terms of absolute size.

The admonition I made earlier—with opinion there is always a likelihood that what you are saying is false—is so to speak, true in spades. There is a likelihood that you are enormously mistaken. Secondly, trying to pin down the answer too precisely is at least a waste of time. It may be worse than that, and I'll have a further comment as soon as I have discussed some of the other results of the experiments.

I mentioned that from one round to the next, the estimates of the subjects converged, came closer together. This was the most visible effect of using the procedure. When we examined the basis for changing opinions from one round to the next, the effect could almost entirely be explained simply in terms of how far away the subject's answer was from the median. The farther away his answer was from the group the more likely he was to change his opinion. This is a simple way to quantify group pressure. It tells us that even with the very thin type of interaction we were dealing with, group pressure is very much a part of the picture, and the effect of group pressure can be described in a very simple fashion.

There is a reasonable suspicion that the individual who resisted this pressure was more sure of his answer than the one who changed. When we compared the accuracy of the first round answers of those who did not change their estimates at all—the holdouts—their median answer was more accurate than the answer of those who changed—a mathematician working on the project dubbed them the "swingers." And in fact, the holdouts were more accurate than the entire group. But on the second round, after the changes, even though the swingers were still less accurate, the holdouts were not as accurate as the total group.

What was happening was a little difficult to trace down, but some detective work dug out the story. Hidden under the large changes produced by group pressure was a smaller effect which is most simply described like this: The group median was one "center of attraction" for changes, but also the true answer was a weaker center of attraction. Those who changed their mind had a residual amount of information that had not been exploited in the first round, and this improved the overall group answer. In this case, the less knowledgeable members of the group still played a vital role in improving the group response. It almost makes you believe in democracy!

One of the traditional rules in what I have come to call opinion technology has been that you should collect only the most expert or the most knowledgeable individuals on a given subject. There is clearly more than a hint here that possibly—unless you are sure that the experts encompass <u>all</u> relevant information—you may want to widen the group.

One of the very important questions in this business that we were able only to touch on in our experiments is—how do you recognize an expert? Again, in the field of knowledge there are fairly rough but probably sound techniques. In particular, for most of the scholarly fields, the community roughly grades its members. At times the measurement techniques are disturbingly reminiscent of the point I made previously concerning the dominant individual in a group—namely grading by volume of publication. But we can be sure that the situation is less satisfactory in the area of opinion.

One obvious approach is to let the subjects rate their own competence. Accordingly, in about half of our experiments we included scales which enable the respondents to state how confident they were in their answer, or how much they thought they knew about the question. One of the real lowlights of our study was the discovery that there was very little correlation between either the confidence an individual had in his answer—or how much he thought he knew about the question—and the accuracy of his response. There was a weak relationship between the average group confidence and the group score, but all attempts we made to select an elite group of more knowledgeable subjects on the basis of their self-rating, failed.

Another approach is to select individuals on the basis of their general intelligence. In about half of the experiments we administered a standard intelligence test. Another lowlight came when we found little or no relation between the intelligence scores and the accuracy of the individuals' responses. There was a somewhat clearer relationship between the average intelligence score of a group, and the average individual performance in that group. This indicated that, to some extent, intelligence was playing a role in determining the overall accuracy of individual scores. But this relationship disappeared entirely when we tried relating the average intelligence of the group with the average group scores. In short, at the rather high levels of intelligence that we were dealing with, the group washed out the effect of intelligence.

We examined the role of men and women in determining the group response. We verified two of the basic cliches about women—namely that they do not have a head for numbers, and they are more changeable. Uniformly, the women subjects were less accurate than the men. And uniformly, the women were more likely to change their opinions than the men. Your first reaction might be that women should not be allowed to have opinions—given any opinion by a woman you can find a man with a better opinion—but that would be wrong. Consider that in order to have improvement in the group estimate it is necessary for some members to change their opinion. The women, by being more responsive to the group, played a significant role in making the group more accurate.

One final consideration with respect to what is an expert. We examined the field in which each subject was majoring at the university. Frankly, we expected that students majoring in the so-called hard sciences, physics, mathematics, engineering, and the like, would turn in better performances than students majoring in the soft sciences and the humanities. In fact, it was pretty much the opposite. At least for men. For women, the conjecture held.

As I say, these results of our experiments are only partly related to the basic question of how you recognize experts. They are very suggestive, however, that the standards we have set, and the procedures for employing experts, are more appropriate for the area of knowledge than they are for the area of opinion.

Let me tell you about one other result that was both surprising, and highly suggestive. We had the general impression that the Delphi techniques are more efficient than face-to-face interaction, and one way to track this down was to discover how long it takes to make the best estimate for one of our questions. Accordingly, we conducted three exercises in which we presented the questions one by one, allowing different times to respond. The time ranged from four minutes per question down to as little as fifteen seconds. That last is barely time enough for the subjects to read the question and write a number. If you consider one of the almanac type questions carefully, it involves a rather complex judgment. Take the question about the vote for Kennedy in Texas. The subjects know something about Texas--its population, the fact that it is a southern state and preponderantly Democratic, but conservative, predominantly Protestant, etc. Nevertheless, to sum up the results of our experiments, the subjects made their best estimates in about thirty seconds. Some of you may recall having heard the advice that in taking an objective examination in school, you are best off to trust your first answer. We seem to have found good objective evidence for this advice. The result also supports a point I made earlier; with opinion, attempts to <u>over</u>refine may result in making the opinion worse.

Let me sum up: Opinion plays a basic role in long range forecasting of technological and social development. The traditional ways of dealing with opinion have significant drawbacks. Our experiments have shown that it is possible to design techniques for putting the opinions of a group of individuals together that avoid some of these drawbacks. However, the improvement we obtain is small.

The next step is to see whether we can dampen the effects of group pressure, in this case excessive convergence, and amplify the movement toward the true answer. We have to wait on further experiments to see whether these are possible, but we have some grounds for optimism.