THE SYSTEMATIC USE OF EXPERT JUDGMENT 
IN OPERATIONS RESEARCH

Olaf Helmer

September 1963

P-2795
THE SYSTEMATIC USE OF EXPERT JUDGMENT
IN OPERATIONS RESEARCH

Olaf Helmer *

The RAND Corporation, Santa Monica, California

This paper, despite its title, is concerned with some rather fundamental problems of operations research. As is well known, there has been a long-standing controversy as to whether operations research may be regarded as a scientific activity in the full sense of that phrase. It is my contention that the answer to this question is a qualified "yes"; there being two qualifications, one of purpose and one of method.

In comparing a scientific investigation with an operational analysis we note, first of all, a difference in emphasis: As pure scientists we are concerned primarily with the pursuit of truth and thus with a better understanding of the world we live in; any application to dealing with the real world, which may flow from such increased understanding, assumes secondary importance. As operations analysts, on the other hand, we are pragmatists, our primary concern is with more effective manipulation of the real world, even if this may have to be accomplished without the desirable degree of understanding of all the underlying phenomena.

This difference in attitude between the purist striving for understanding and the pragmatists striving for control of his surroundings brings with it a methodological difference, which is the second qualification to which I referred above. Both the exact scientist and the operations analyst tend to make use of what is sometimes called a mathematical model of the subject matter; in the

* Any views expressed in this paper are those of the author. They should not be interpreted as reflecting the views of The RAND Corporation or the official opinion or policy of any of its governmental or private research sponsors. Papers are reproduced by The RAND Corporation as a courtesy to members of its staff.

The paper was prepared for presentation at the Third International Conference on Operational Research, Oslo, Norway, July 1963.
case of the scientist such a model is apt to be part of
the well-confirmed body of our scientific knowledge,
whereas an operations research model is of a more tentative,
ad hoc, character. In other words, even if the current
status of science provides no well-established theory for
the phenomena to be dealt with by the operations analyst,
the latter must nevertheless construct a model as best he
can, where both the structure of the model and its
numerical inputs may be based merely on intuitive insight
and limited practical experience by the analyst himself or
by whatever expert advisers on the subject matter may be
available to him. As further insights accrue and more
experimental data become available, the operations analyst
has to be ready to discard his first model and replace it
with an improved one. This tentative procedure, dictated
by pragmatic considerations, is thus essentially one of
successive approximation. In this regard, operations
research has a status similar to that of the so-called
inexact sciences, of which medicine, engineering, and most
of the social sciences are examples.

Therefore, in comparing operations research with an
exact science, it is with regard to exactness that operations
research falls short, but not necessarily with regard to
the scientific character of its methods. As I have dis-
cussed elsewhere in detail, namely in my paper with Nicholas
Rescher, "On the Epistemology of the Inexact Sciences,"* the
methods of the inexact sciences (including operations
research), despite their reliance on less-than-universal
laws and on intuitive judgment, can be just as "scientific"
as those of the exact sciences, as long as certain standards
of objectivity in confirming predictions are observed.

*Management Science, 6 (1959); also issued by The
RAND Corporation as P-1513.
In this paper, I am particularly concerned with one methodological aspect of operations research, namely the reliance on expert judgment. It seems to me that the operations analyst, far from denying the need for intuitive expertise, should consciously acknowledge it and indeed make the most of it. Only by replacing the surreptitious use of expertise by the explicit and systematic application of it can operations research hope to acquire the status of a science, albeit an inexact one.

The first question which must be raised is how objectivity can in fact be safeguarded if direct empirical evidence is replaced, even in part, by opinion, however expert. The justification for such a procedure may be seen in looking upon the expert as an objective indicator, comparable to a measuring instrument. That is, the expert's predictive pronouncements have to be treated as an integral, intrinsic part of the subject matter, in the sense that information on his reliability--as evidenced by his past predictive performance--must be added to our knowledge about the field in question. By incorporating in our model the performance of this instrument for measuring probabilities--i.e., the predictive expert--our data are supplemented by the expert's pronouncements and our theory is supplemented by objectively confirmable hypotheses on his reliability.

The often unavoidable reliance on expert judgment in operations research therefore does not necessarily imply that objectivity is being thrown to the winds. However, in order to raise the level of objectivity and to exploit expert judgment most effectively when available, an effort has to be made to develop specific techniques for identifying expert performance and for processing data in the form of expert pronouncements into predictions of the greatest possible reliability.
The simplest way of scoring an expert's ability is in terms of his past performance, provided a record of the latter is available. His "degree of reliability" may be defined as the relative frequency of cases in which, when confronted with several alternative hypotheses, he ascribed to the correct one a greater probability than to the others. Since some predictions are clearly easier to make than others—either because the statement content may be logically less precise or because more theoretical knowledge may be available—it is not so much the absolute degree of reliability that counts but the relative degree of reliability as compared to that of the average well-informed person. Another, possibly more subtle, measure of an expert's performance is his "degree of accuracy," that is, the degree to which his "personal" probability estimates match his reliability in the class of those hypotheses to which he ascribed the probability p.

An expert is, of course, best selected on the basis of his past reliability. If, moreover, data on his past accuracy are available, it may be possible to determine whether his probability estimates exhibit a systematic bias; if so, it is possible to compensate for this and thus to obtain more accurate probability forecasts. (This assumes a certain amount of statistical stability in an expert's performance, a hypothesis for which some evidence exists.) It might also be worth while examining systematic possibilities of improving an expert's performance—aside from continued training and experience in his field of expertise—by providing him with appropriately selected aids, say, in the form of tabulated data, theoretical information, check lists of relevant considerations, and so on.

Often it is even better than having one expert available to have two or more experts on hand. There are
two ways in which groups of experts can be utilized. One is to have them perform individually and to seek methods for the best combined use of their pronouncements; this is the area of consensus techniques. The other is to have them perform jointly in a group exercise and to obtain expert judgments from the group as a whole; this includes everything from simple round-table discussions among the experts, possibly resulting in a vote on the issue in question, to sophisticated pseudo-experiments. A method which falls somewhere between individual and group action is the so-called Delphi technique.

If several rather than one expert are available for consultation on a given question, the simplest means of forming a consensus of their judgments is to take the median of their responses. A variant of this procedure is to take weighted means, where the weights may either be based on a record of past performance or on self-appraisal by the expert as to his relative competence in answering a particular question. More sophisticated weighting methods might take into consideration not only the individual expert's past performance or present self-appraised competence but also correlations between these indices for several experts.

Among successful past applications of a consensus technique I may mention the following: A study of the prediction of social and technological events by Kaplan, Skogstad and Girshick (in Public Opinion Quarterly, 1951); an unpublished study carried out at RAND many years ago of the predictability of the outcomes of horse races, using a consensus of handicappers' forecasts; a consensus based on an opinion survey conducted by the Raytheon Corporation in conjunction with its very comprehensive strategic computer game; and a consensus of opinions on political progress in various South-American countries conducted by Fitzgibbon and Johnson (in American Political Science Review, 1960).
Next let me say a few words about what we at RAND have called the Delphi technique, which attempts to add to the basic consensus method (of unweighted or weighted medians) the virtue of exposing the experts' views to one another's critique, without actual confrontation and all its psychological shortcomings (such as specious persuasion, an unwillingness to abandon publicly expressed opinions, and the bandwagon effect of majority opinion). This Delphi technique replaces direct debate by a carefully designed program of sequential individual interrogations (best conducted by questionnaires) interspersed with information and opinion feedback derived by computed consensus from the earlier parts of the program. Some of the questions directed to the respondents may, for instance, inquire into the "reasons" for previously expressed opinions, and a collection of such reasons may then be presented to each respondent in the group, together with an invitation to reconsider and possibly revise his earlier estimates.

Both the inquiry into the reasons and subsequent feedback of the reasons aduced by others may serve to stimulate the experts into taking into due account considerations they might through inadvertence have neglected, and to give due weight to factors they were inclined to dismiss as unimportant on first thought.

The only published account of an application of the Delphi technique is that by Dalkey and myself this year in Management Science.* A large-scale experiment in the use of the Delphi technique is currently under way, in which a number of international panels of respondents are being used in an effort to arrive at long-range contingency forecasts of the state of the world 25 years hence.

---

Another procedure which is of considerable aid to the individual expert in his performance, especially if he is to act in concert with other experts, is the application of simulation techniques. This is a form of pseudo-experimentation, in which a suitable model of the real world rather than the real world itself is subjected to experimental manipulation. This procedure is particularly promising when it is desirable to employ several experts with varying specialties in a context in which their forecasts cannot be entered independently but where they are likely to interact with one another. Here the model furnishes the experts with an artificial, simulated environment, within which they can jointly and simultaneously experiment, responding to the changes in the environment induced by their actions, and acquiring through feedback the insights necessary to make successful predictions within the model and thus indirectly about the real world.

Various simulation experiments have been carried out in the past. As an example of special current interest, I would like to mention our hopes of applying simulation techniques to the study of the economies of underdeveloped nations, a project on which E. S. Quade reported at a recent meeting in Paris on the application of operations research to underdeveloped countries. *

A particular case of simulation, involving role-playing by the participating experts, is known as operational gaming. A simulation model may said to be gaming a real-life situation if the latter concerns decision-makers in a context involving conflicting interests. In operational gaming, the simulated environment is particularly effective in reminding the expert to take all the factors into account.

---

* P-2718, "An Approach to the Study of a Developing Economy," by O. Helmer and E. S. Quade.
in making his predictions that are potentially relevant; for if he does not, and chooses a tactic or strategy which overlooks an essential factor, an astute opponent will soon enough teach him not to make such an omission again.

Aside from well-known applications of operational gaming to military combat, I may mention such applications as to procurement planning, to cold-war diplomacy, to economic bargaining, and to industrial competition.

By way of a summary, let me restate as follows:

1. I consider operations research a science, but an inexact one.

2. The operations analyst, as opposed to the pure scientist, emphasizes control rather than understanding.

3. The use of judgment in constructing and applying operations-analytical models is inescapable.

4. Since the use of judgment is a sine-qua-non, such judgment should be as expert and its application as systematic as possible.

5. It would be in the interest of raising the scientific standards of operations research to seek improved methods of identifying and measuring expertness and of employing experts efficiently both solitarily and in groups.

6. In particular, the utilization of groups of experts by consensus techniques, by the Delphi technique, by simulation procedures and operational gaming, should be further refined to the point where they will be generally acceptable as part of the stock-in-trade of operations research.