IIASA International Institute for Applied Systems Analysis
www.iiasa.ac.at

# A Method for Evaluating R & D Proposals in Large Research Organizations

Larichev, O.I.

IIASA Collaborative Paper
November 1982

NOT FOR QUOTATION
WITHOUT PERMISSION
OF THE AUTHOR

# A METHOD FOR EVALUATING R & D PROPOSALS IN LARGE RESEARCH ORGANIZATIONS

Oleg I. Larichev

November 1982
CP-82-75

Collaborative paper series on
*Comparative analysis on application of decision support systems in R & D decisions*



INTERNATIONAL INSTITUTE FOR APPLIED SYSTEMS ANALYSIS
2361 Laxenburg, Austria

**COLLABORATIVE PAPER SERIES ON COMPARATIVE ANALYSIS ON APPLICATION OF DECISION SUPPORT SYSTEMS IN R & D DECISIONS**

This series of papers are a product of collaborative research coordinated through IIASA's Management and Technology Area. The collaborating institutions are Hungarian State Office of Technical Development (personnel: Anna Vari, Janos Vecsenyi, Laszlo David); Decision Analysis Unit, Brunel University, England (Personnel: Patrick Humphreys, Lawrence D. Phillips); All-Union Research Institute of Systems Studies, USSR (Personnel: Oleg. I Larichev).

The papers report case studies prepared by the personnel from the collaborating institutions based on their own, and their colleagues' work in their own institutions. They worked together as a team in developing the methods for the analysis of these case studies which are described in the first paper in the series.

IIASA provided support for this work through its telecenter for communication between the investigations, and provided facilities for short term meetings between the investigations at IIASA for development of case studies and their comparative analysis. Particular MMT staff were Ronald M. Lee, Nora Avedisians, and Miyoko Yamada, who is the editor of this series.

A summary of this comparative analysis, based on the first four case studies in this series was presented at the IFIP/IIASA Conference on *Processes and Tools for Decision Support*, Laxenburg, Austria, July, 1982.

The papers in this series are

1. Humphreys, P.C., A. Vari and J. Vecsenyi: Methods for analyzing the effects of application of Decision Support Systems in R & D decisions (CP-82-69).
2. Vari, A. and L. David: R & D planning involving multicriteria decision analytic methods at the branch level. (CP-82-73).
3. Vecsenyi, J.: Product mix development: strategy making at the enterprise level. (CP-82-74).
4. Larichev, O.I.: A method for evaluating R & D proposals in large research organizations. (CP-82-75).
5. Humphreys, P.C. and L.D. Phillips: Resolution of conflicting objectives in evaluating R & D projects involving collaboration between industry and higher education. (CP-82-xxx, forthcoming).

The paper presented at the IFIP/IIASA conference will be published as Humphreys, P.C., O.I. Larichev, A. Vari, and J. Vecsenyi, Comparative analysis of decision support systems in R & D decisions, in H.G. Sol (ed.), *Processes and Tools for Decision Support*, Amsterdam: North Holland, 1982. Another study in this series was published separately as L.D. Phillips: Requisite decision modeling: a case study. *Journal of the Operations Research Society*, 1982, 33:303-311.

**CONTENTS**

# A METHOD FOR EVALUATING R & D PROPOSALS IN LARGE RESEARCH ORGANIZATIONS

Oleg I. Larichev

## I. PROBLEM DEFINITION

Among various problems of R & D planning there is one most often faced (Zuev et al. 1979, Larichev et al. 1979), notably:

- there are individuals or organizations submitting proposals on R & D. These are potential executants or clients, interested in R & D results;
- there is a decision maker (DM) responsible for choosing the best R & D alternatives.

In this paper, a DM is understood to be the head of a planning office or the Chief Executive Officer of an organization. It is assumed that the DM follows a certain policy in choosing among the R & D proposals realized through a set of his criteria.

The specific characteristic of the problem is that the decision rules have to be developed before any of the R & D proposal is submitted, so that the DM is able to assess the proposals as they reach him. For, as the last proposals are furnished, the choice must actually be completed. The latter consideration provides the DM no opportunity for employing the characteristics of the submitted proposals in the formulation of the R & D plan. He must fix the concepts of his policy before the proposals start arriving and merely adjust it soon after. Another feature of the problem considered here is the absence of rigid limits on resources necessary for conducting the R & D. That is, the problem is not consistent with the general problem of portfolio optimization (e.g., Francis and Archer 1971). The idea is that the authors of proposals are in a position to secure the required resources in the case where the DM approves their R & D proposals (e.g., through state budget organizations). The rejection of a proposal forces its authors to formulate new approaches.

The DM's principal task is to make a choice of a set of the best alternatives to be integrated into the R & D plan. His second task is to compare both the accepted and rejected proposals in order to define the merits of the proposal developers. Hence, the DM is interested in forming a certain ranking of the R & D alternatives with respect to their utility. In the case study described in this paper, the problem featuring the characteristics listed above was approached both from the standpoint of a large interdisciplinary research institute and from the point of view of a planning office heading a number of research institutes.

The problem, as presented, was generated by the desire of executives (director of an institute; head of a planning office) to exert a stronger influence on the process of selecting the best R & D proposals. As a rule, the process involved the following steps.

The authors formulated the proposals so that to emphasize their merits. Inasmuch as the proposals were quite different from one another and multidisciplinary in nature, they were too complex for the DM to evaluate them directly. Consequently, to evaluate the alternatives he had to resort to experts' assistance. The latter, however, were not required to make a general evaluation of the proposals but had to answer explicit questions reflecting some or other aspects of the DM's scientific policy. Besides, a need arose to develop a decision rule integrating the scientific policy and the experts' judgements. The resulting method was supposed to be utilized by the DM.

The problem under study constituted a choice of the best R & D alternatives to be included in the 3-5 year plan (one decision). Elaboration of the plan involved contributions from the authors of proposals, the DM, and decision consultants. The information concerning the set of criteria (see below) was available to everybody. The decision rule was developed by the consultants and the DM for the latter's use.

The DM expected the consultants to submit explicit verifiable recommendations consistent with his policy. This placed specific constraints on the decision rule elaboration technique. The traditional process for formulating the R & D plan can be presented in the way shown in Figure 1.

*Figure 1*. The traditional process of R & D plan formulation.

| Participants | | Authors of proposals | Decision maker | Experts |
|---|---|---|---|---|
| *Stage contents* | | | | |
| Stage 1: | R&D proposals formulation | X | | |
| Stage 2: | Proposal evaluation | | | X |
| Stage 3: | Decision making | | X | |

The new plan formulation procedure differed from the old one in that the second and third stages were changed: the experts would now receive a special questionnaire and the DM would take decisions on the basis of the formulated decision rule.

The third stage was also contributed by consultants. Hence, as far as the authors of proposals are concerned, the old and new procedures do not differ. It is the DM who is most affected by the new procedure as it qualitatively changes the entire style of his work.

In practice, the number of proposals ranged from several hundreds to several thousands. The number of criteria used in the evaluation of proposals generally did not exceed 10 and most often amounted to five to seven.

Special emphasis must be given to the nature of these criteria. Choice among R & D alternatives at this level of decision making is considerably affected by factors which are hardly formalized such as: scale of R & D, kind of scientific backup, versatility of expected results, skill of potential researchers and developers, etc. In a word, the criteria are *qualitative* in nature.

The R & D alternatives which were subjected to evaluation largely represented applied research, i.e., they were oriented towards the solution of specific problems.

## II. A METHOD FOR R & D CHOICE SITUATION DESCRIPTION

The technique developed in this case is distinguished by a specific way of describing the R & D choice situation, and a special way of obtaining a general estimate of R & D proposals through a technique involving multiple criteria evaluation.

The description of the R & D choice problem should be articulated in a language that would allow the structuring of many real-life problems. This means developing *qualitative* criteria scales with verbal (as opposed to numerical) estimates of grades of quality on these scales. Figure 2 gives example of a feasible scale of evaluation of this type, expressing complex criteria.

The distinguishing features of such an approach to describing the decision situation are as follows:

- It can help introduce complex qualitative notions into consideration and, in so doing, obtain a complete description of all the factors relating to the real-life situation being considered.
- The formulations can take account of the uncertainty arising from incomplete knowledge of implications of the decision at the time of decision making. The formulations are rather flexible to emphasize the risk involved in choosing some criteria estimates.

*Figure 2.* Example of a qualitative scale using verbal estimates of grades of quality.

---

**"Availability of research backup for the R & D exemptions"**

☐ 1. The executant has completed a major portion of the given R & D. The remaining part of research poses no problems.

☐ 2. The R & D activities face a number of problems. There are some ideas concerning their solution and defined lines of research.

☐ 3. The R & D project depends on the solution of a number of difficult problems. There are no ideas concerning their solution.

---

The estimate formulations can easily acquire a predictive connotation.

- The description of the situation in terms of qualitative criteria (like that shown in Figure 2) is a ***verbal decision model.*** The formulations of estimates of quality reflect those grades of quality which the planners take account of in decision making. In fact, they represent a language for communication between planners and experts, and for obtaining the relevant information. The estimate formulations are quite usual for the experts as they contain words and expressions typical of that used in the environment in which both the planning authorities and experts work. At the same time, the formulations allows the experts to spot the substantial difference between adjacent quality grades. The planners trust the descriptions as they were made on the basis of their preferences.

The method proposed for describing the decision situation helps significantly to increase the reliability of information furnished by the experts. The latter tend to be biased to the greatest extent when they are offered the opportunity of evaluating the decision alternatives as a whole and allowed to determine their strengths and weaknesses on an *ad hoc* basis. In the case where the set of criteria and estimate formulations are made available at the start, the expert has to consider the appraised objects (i.e., proposals) from the point of view of the *planner's* preferences. In evaluating a proposal on each criterion, the expert selects one out of several submitted formulations on the criterion as appropriate in characterizing the proposal. Should he be biased, and would like to "correct" the actual R & D estimate, his assessments on individual criteria are easy to verify. And for the expert himself this raises the danger of being considered "professionally incompetent."

It is worth noting that the set of criteria was defined on the basis of decision maker's desire to emphasize those qualities substantial for a comprehensive evaluation of R & D. The verbal formulation of estimates of quality grade on each of the criteria were developed with the DM's assistance. Their quantity was determined by the DM's intention to single out certain distinct quality levels to be subject to measurement. Each formulation was thoroughly reviewed in a session with a group of potential experts.

All incoming proposals were divided into groups clustered by subject matter. Experts were nominated to evaluate the proposals by multiple criteria. First, each alternative was evaluated by one expert and then his estimates were verified by some other more competent expert. The complete set of criteria employed in various cases where the method was applied is presented in Zuev et al. (1979) and Larichev et al. (1979).

## III. A METHOD FOR DECISION RULE FORMULATION

The description of a decision situation in a DM's usual language considerably increases his trust in the outcome of the analysis. To maintain the trust, it is necessary to use this language throughout the decision rule formulation.

The R & D general utility model can be treated as a rule according to which every combination of criteria estimates is consistent with a certain class of *quality* (the decision rule).

On the basis of research on human behavior in choosing among complex alternatives, we can formulate certain requirements for the procedures for eliciting the DM's preferences. First of all, these procedures must provide for verification of the DM's preferences for stability and consistency. They must involve primarily questions where the probability of obtaining reliable information is the greatest. The method of eliciting the DM's preferences may consist of formulating hypotheses concerning the possibility of obtaining some data or other from a person, and of the verification of these hypotheses.

Each combination of estimates on criteria is an image of a certain alternative for the DM. The two most bright, "contrasting" images correspond to the combinations of the best and the worst estimates by all criteria (let us call them reference situations).

A hypothesis was put forward that under 7-8 criteria with 2-6 estimates (quality grades) on each of the criteria scales, the decision maker can sequentially and consistently determine the utility superiority of alternatives differing in their estimates on two of the criteria, while on all other criteria the estimates of these alternatives are similar and belong to the reference situation. This hypothesis was based on the assumption that the DM treats the images created by the best and the worst estimates on criteria as something whole in comparing the two alternatives, every one of which receives *one* best (or worst) estimate and another — arbitrary estimate on the two remaining criteria.

We shall illustrate with an example what information is required in this case from the DM.

Let the estimates of an alternative be first set at the *best* on all (N) criteria, but on two criteria, A and B, they can change. The transition from the best to the next estimate on each of these criterion is related to quality deterioration. Let us put the following question to DM:

> Let the alternative have the estimates $A_1$ and $B_1$ on criteria A and B, respectively, and the highest estimate on all other (N-2) criteria. Let us consider two cases: (a) quality deterioration has occurred by criterion A, corresponding to transition from estimate $A_1$ to $A_2$; (b) quality deterioration has occurred by criterion B, corresponding to transition from estimate $B_1$ to $B_2$. Which of the two cases corresponds to the greatest quality deterioration?

When answering this question the DM compares the *deteriorations of quality* $x_1$ and $y_1$ (Figure 3). The result of the comparison can be represented by a graph arc directed from the best estimate to the worst one (in Figure 3, $y_1 < x_1$).

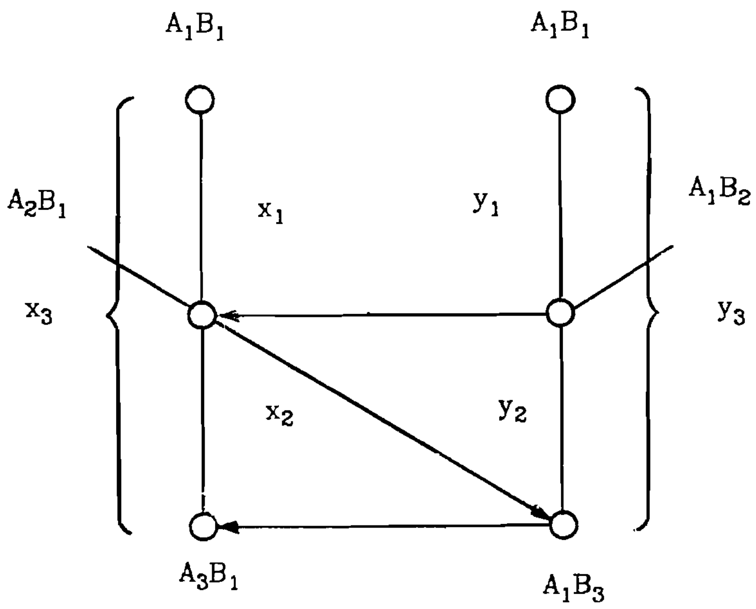


*Figure 3.* Comparison of deteriorations of quality on criteria, A and B.

Then comparison is made of deterioration of qualities $x_1$ and $y_3$ (a set of two successive deteriorations of qualities $y_1 + y_2$), etc. On the basis of DM's answers one can build a unified scale of criteria A and B, shown in Figure 4. Similarly, any enquiry can be made of the DM, using the other reference point (the worst estimates on (N-2) criteria). The DM's possible responses in each enquiry are of the form "more," "less," "approximately equal."

It is necessary to make 0.5 (N-1) comparisons of criteria pairs. This leads to a closed procedure implying tests for consistency. The results of each comparison provide a basis for building a unified scale comprising the two criteria. Obviously, with $N \geq 2$ the information required for construction of a unified scale is at least duplicated (e.g., with $N = 3$ the scale

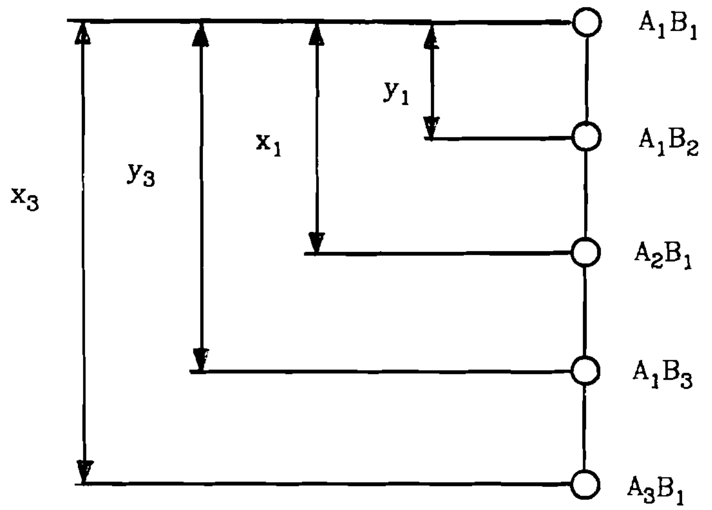


*Figure 4.* Construction of a unified scale of criteria A and B corresponding to the comparisons shown in Figure 1.

of criteria A and B can in a number of cases be built both directly and on the basis of comparing criteria A and C, and criteria B and C) and the amount of abundant information is growing with the growth of N. Hence, building a unified scale of estimates from N criteria makes it possible to test the DM's preferences for consistency and transitivity.

It is worth noting that the greater the number of criteria, the more difficult it is for the DM to take account of the reference situation. The method for preference elicitation provides for the simultaneous increase in the amount of abundant information employed for testing the DM's preferences.

The closed procedure for preference elicitation and construction of a unified scale of criteria estimates described above was tested on a number of model situations, and in practice, with the DM operating with four criteria (three times), with six criteria (twice), and with seven criteria (twice). The data received from the DM was almost always consistent. Thus, when interrogating three decision makers concerning four criteria with three to five estimates on each of the criterion scales, there was not a single violation of transitivity. When questioning on six to seven criteria with three to six estimates on each of the scales, there were one to three inconsistent answers out of 50-70. A second questioning of the DM allowed him to remove any inconsistencies (the time taken to question the DM did not exceed one hour). It can be assumed that with three to four estimates on each of the criteria scales a small number of inconsistencies will remain on the level of N = 10.

At the same time, there were substantially different strategies of comparison (different unified scales) in various reference situations. Hence, the reference situation directly affected the DM's strategy. A small number of contradictions indicated that the DM can consistently and reliably express his strategy near the reference situations.

The experiments described above also tested a hypothesis that the DM can reliably compare magnitudes of quality deterioration along separate criteria scales near the reference situations. In the course of comparisons, the DM is asked the questions like the following:

> Let there be best estimates on (N - 1) criteria. Which out of the two quality deteriorations (a) transition from estimate $A_1$ to estimate $A_2$; (b) transition from estimate $A_2$ to $A_3$ is the greatest?

The DM's answers can look like "more," "less," "approximately equal." The responses to questions of this type have always been consistent, though this could well be expected, given a small number of estimates on criteria scales. It should be noted that in some cases the comparison results varied in different reference situations, i.e., the DM's strategies were complex.

The existence of reliable and tested information near the reference situations allowed to build two unified scales of estimates near the respective reference situations.

Further on, this information was employed for checking the preference independence between criteria. In the case of criteria dependence, this must first of all manifest itself in different comparisons near the reference situations. Indeed, criteria dependence implies that the quality deterioration along the criterion scale depends on the reference situation. Under the ordinal criteria, the dependence emerged due to the "bright," "absorbing" estimates characterizing some reference situation or other (emergence of new quality).

Testing for independence implies comparisons between one and the same pair of criteria, made near different reference situations.

Testing for independence is conveniently carried out with a table like that furnished for criteria A and B in Figure 5. All comparisons on the scales of criteria A and B (Figure 3) and the two scale estimate comparisons conducted near the reference situations are entered in the table.

| | $x_1$ | $x_2$ | $x_3$ | $y_1$ | $y_2$ | $y_3$ |
|---|---|---|---|---|---|---|
| $x_1$ | = | < | < | > | > | < |
| $x_2$ | | = | < | | > | < |
| $x_3$ | | | = | > | > | > |
| $y_1$ | | | | = | < | < |
| $y_2$ | | | | | = | < |
| $y_3$ | | | | | | = |

*Figure 5.* Comparisons of the scales of criteria A and B ($x_1$, $x_2$, $x_3$, $y_1$, $y_2$, and $y_3$ are defined in Figure 1).

In a case where the comparisons are contradictory, some square of the table in Figure 5 will contain conflicting data.

Should the information received from the DM near different reference situations be consistent, then it is inferred that criteria A and B are independent (this does not rule out the possibility of dependence for the other pairs of criteria). If the test revealed the criteria dependence, then the relevant pairs of criteria are immediately identified. The cause of dependence is easily determined: these are the estimates on some criteria near one of the reference situations which result in a "new quality." We can neglect these estimates, i.e., pass to the adjacent (higher or lower) estimates on the criteria scale and thus formulate a new reference situation.

If, for example, we managed to derive that concerning four criteria (A, B, C, D) the comparisons of quality deterioration for the two criteria A and B are conflicting with $C_1D_1$ and $C_3D_3$ (here these are three estimates on each criterion scale) and that dependence emerges due to the estimate $C_3$, then a new reference situation $A_3B_3C_2D_3$ is formed wherein the DM's preferences are derived by the method described above. Consequently, we test the consistency of information received from the DM by the new (third) reference situation and by the situation $A_1B_1C_1D_1$. The tests are performed with a view to establishing the independence subspace in the multidimensional space of criteria estimate combinations. The comparisons of quality deterioration in the independence subspace do not depend on the reference situations, i.e., remain unchanged with any reference situation.

The transition from the DM's information to the decision rule is carried out in the following manner.

Two combinations of estimates $L_1$ and $L_2$ can, in conformity with the decision maker's preferences, be in one of the following relations:

$L_1 \rightarrow L_2$ (combination $L_1$ is preferable);

$L_1 \leftrightarrow L_2$ (combinations are equivalent);

$L_1 \sim L_2$ (combinations are non-comparable).

Let us define the binary relation for any pair of estimate combinations in the following way:

1. If all the estimates relate to criteria independence subspace:

(a) $L_1 \rightarrow L_2$, if the aggregate information obtained from the DM in the form of tables, similar to the one in Figure 3, allows us to find no smaller quality deteriorations of $L_2$ for each of the quality deteriorations of respective $L_1$ and, at least for one — a larger deterioration. ($L_1 \leftrightarrow L_2$, if all quality deteriorations are equal pairwise).

(b) $L_1 \sim L_2$ in all cases when the previously mentioned correspondence between the quality deteriorations cannot be spotted.

2. In cases when the estimates relate to subspace wherein a part of criteria are dependent.

(a) $L_1 \rightarrow L_2$:

— if this results from a direct comparison near one of the reference situations;

— if, on dependent criteria, estimates $L_1$ dominate estimates $L_2$, (i.e., not worse on all criteria, and better at least on one criterion), and on all other quality deterioration criteria $L_2$ (in conformity with the DM's information) are not inferior to quality deteriorations $L_1$.

(b) $L_1 \leftrightarrow L_2$ if all quality deteriorations are equal pairwise.

(c) $L_1 \sim L_2$ —in all other cases.

Transition from a binary relation to a quasi-order on the combinations of criteria estimates is carried out as follows.

Let us single out, on the basis of the binary relation mentioned above, all undominated estimate combinations and refer to them as the first nucleus. Then form the second nucleus out of the remaining combinations, and so on. Each of the estimate combinations is ranked "i" in the case that it is dominated by a combination ranked (i - 1) and itself dominates the combination ranked (i + 1). If a combination is dominated by some other combination from Kth nucleus, and itself dominates a combination from the (K + P) nucleus, then its rank is "fuzzy" within the range from (K + 1) to (K + P - 1).

An example of a possible quasi-order for 20 alternatives is shown in Figure 6.

It was proved that the quasi-order obtained with the given method is acyclic.

The nuclei and alternative ranks thus obtained can be directly employed by the DM with a view to defining the best, worst, medium and intermediate quality groups of alternatives. It should be noted that at all stages of decision rule elaboration the DM's data were used undistorted.

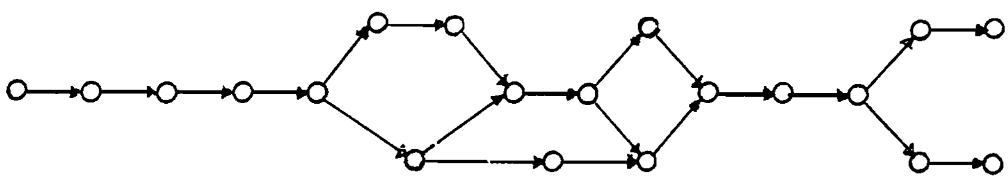

*Figure 6.* A possible quasi-order for 20 alternatives.

The algorithm of obtaining information from the DM and construction of a unified scale is realized through an interactive procedure on a PDP-11/70 computer.

## IV. EVALUATION OF THE METHOD'S APPLICATION

A positive factor of the method's application was the trust shown in the results on the part of the DM. Indeed, all the resulting estimates were directly based on the DM's verbal information without any transformation thereof. It was possible to define the relative position of any pair of alternatives directly on the basis of data obtained from DMs near the reference situations.

Thus, judging by the reliability criteria of the decision makers' information and the trust towards it, the method ranked high in practice. Besides, in one of practical cases the method was evaluated by a third criterion: the forecasting ability of proposal evaluation for R & D. The recommendations obtained for a group of 700 proposals with the help of the method were, for a number of reasons, not implemented. Retrospective examination of the actual results of the R & D proposals which were subsequently chosen revealed the correctness of estimates, obtained with this method, for 80% of R & D.

The method had several successful applications where the estimates obtained with its help were actually employed in decision making. Naturally, the criteria, decision rules and number of alternatives changed between applications.

## V. ORGANIZATIONAL SYSTEMS AND DECISION METHODS

It is very important, in regard to the applicability of any particular decision method, that the planners be ready to apply them. Of course, a more reliable and methodologically validated technique has a greater chance of successful application. The point, however, is not only in the merits or shortcomings of a procedure or a method.

First of all, the new methods and procedures must be adapted to the existing organizational structures and to the traditional ways of gathering and considering the proposals. Penetrating such systems, the method changes their essence, sharply increasing the rationality and centralization of decision making. At the same time, there is no need for drastic changes of such systems, which are rather difficult for planning agencies.

The problem of applying the new method and procedures is also of a psychological nature. DMs tend to share a number of old-fashioned views hampering the improvements in the traditional forms of work. One of them is a consideration implying that a great number of R & D themes (up to several thousands) can well be directly analyzed. It is clear that with complex and different R & D proposals such notions are far from realistic. Another notion is that a choice can be avoided either through a proportional allocation of resources to all the options, or by securing additional resources. Experience shows that this unrealistic assumption can result in dissipation of resources. The third notion holds that the application of the new methods and procedures must lead to a reduced DM's influence on decision making. Quite the reverse thing occurs with adequate methods. It should be stressed once again that, on the basis of some estimates or other, the final decision is always taken by the decision maker with due account of the existing constraints.

The complex problems characterizing R & D planning do not tolerate either an approach with is too simple or extreme formalization. The practical utility of a method consists of its assistance to planners. Only then with the new methods become a useful tool for improving the existing systems of long-range research planning.

**REFERENCES**

Francis, J.C. and S.H. Archer. 1971. *Portfolio Analysis*. Englewood Cliffs, New Jersey: Prentice-Hall.

Larichev, O.I., Ju. A. Zuev, and L.S. Gnedenko. 1979. Method ZAPROS (closed procedures near reference situations) for Solving the Ill-Structured Problems of Choice with Many Criteria. Preprint VNIISI, M. (in Russian).

Zuev, Ju. A., O.I. Larichev, V.A. Filippov, and Ju. V. Chujev. 1979. The Problems of Estimation of R and D Projects. Vestnik Akademii nauk SSSR, No. 8 (in Russian).