Journal of Multi-Criteria Decision Analysis, Vol. 1, 127–138 (1992)

Cognitive Validity in Design of Decision-Aiding Techniques

O. I. LARICHEV
Institute for Systems Analysis, Moscow, Russian Federation

ABSTRACT

Issues of the cognitive validity of prescriptive decision-aiding methods are discussed in this paper. In view of the difficulty of external validation, the focus is on internal validation of the user interface. The following criteria are applied to a variety of decision aid designs: (1) psychologically valid measurement of features impacted by the decision; (2) psychologically valid elicitation of decision maker's preferences incorporated into a decision rule which processes those features; (3) possibility of checking for decision maker's consistency; (4) possibility of providing him/her with an explanation of the aids findings.

KEY WORDS Normative and prescriptive decision methods ZAPROS

INTRODUCTION

How can one justify the utilization of decision techniques in practical applications? How can one determine the scientific validity of these methods, distinguishing theoretically valid techniques from heuristics? These are questions that have long concerned both practising analysts and theoreticians developing new decision methods. The specific difficulty in searching for answers to these questions is determined by the nature of decision problems *per se*, i.e. ill-structured problems dominated by qualitative, barely formalizable factors. The major difference between decision problems and operations research problems lies in the much greater role played by subjective factors in the problem statement and solution of the former.

As is well known, in the decision area it is common to distinguish between descriptive research (focusing on how people make decisions) and normative decision methods (specifying how to make logically coherent decisions). It is natural, we believe, to include in normative techniques both expected utility theory (Von Neumann and Morgenstern, 1953) and multiattribute utility theory (Fishburn, 1964). The contradictions between descriptive research findings and the human requirements of normative decision techniques have become glaring in recent years. Amid the responses to the contradictions is one aimed at devising techniques which take into consideration the results of descriptive studies. In Brown (1989) such methods are referred to

1057-9214/92/030127-12$11.00

Received 6 September 1991
Accepted 9 November 1992

as prescriptive. How must the prescriptive techniques be built? What must the scientific validation thereof be like? These are the questions with which the present paper deals.

SEARCH FOR VALIDATION OF DECISION METHODS

Is it possible to test a decision method's effectiveness? The prime and obvious means is to exemplify a set of test situations where the objectively correct answers are known in advance to the observer but unknown to the decision maker and analyst. It is easy to see how difficult it is to find such situations. However, even in such cases, most informative for testing, good results (a high percentage of correct answers) can often be explained not only by the effectiveness of the approach but also by the quality of information provided by the decision maker. Indeed, the outcome of any given decision method application can be explained by what kind of information was supplied by the decision maker and the way it was processed and used within the framework of the decision method.

The second way of testing a decision method is closer to real life: to apply the method to a set of real problems, wait for an objective indication of the success of their solution (to establish the quality of outcome of every decision after some time) and evaluate the method by the percentage of successful solutions. The success could be indicated by the unanimous approval of the decision outcome. We have applied this approach to a problem of evaluating R & D projects by the 'ZAPROS' method (Larichev *et al.*, 1978; Larichev and Moshkovich, 1991). The ZAPROS method allows one to construct a quasi-order of alternatives using only psychologically valid ways of eliciting information. Despite good results (coincidence after 5 years of project results with the initial ZAPROS estimates in 82% of a set of 750 projects), we understood that the success could be explained by both the quality of the method and the inherent skill of the decision maker materialized through this method. Also, contingencies may occur after the decision and before its implementation which may affect that implementation. (Note that given a large number of independent decisions, this factor is less important.)

The above two approaches are evident. Should they be used for comparing various techniques, an analogous remark holds: how effective (correct, reliable, error-free) was the decision maker using a given method? Hence the difficulties of 'external' validation of decision methods are evident.

There is yet another validation approach: assessing decision techniques by their intrinsic qualities with respect to organization of 'decision maker–method' interaction. As will be seen later, this kind of assessment may be very beneficial.

PSYCHOLOGICAL VALIDATION OF DECISION METHODS

The shortcomings of normative decision techniques are due largely to the desire of their authors to confine themselves to mathematical validation. The operations research paradigm is applied to decision making; for example, maximizing the expected utility using the axiomatic approach to utility function construction (Von Neumann and Morgenstern, 1953) is treated as the only way of scientifically validating a decision method.

The basic source of information, however, permitting the evaluation of a decision alternative is a human being, i.e. the decision maker. Hence information elicitation must attend to the specifics and constraints of the human information-processing system. Therefore decision methods may and indeed must be evaluated against the requirements of human beings and their difficulties in information problems. It follows from this that the scientific criteria of decision method construction must be psychological criteria of 'decision maker–method' interaction arrangement within a given decision method.

One should not infer from the above that the mathematical criteria of method validation are to be rejected altogether. For a number of methods (but not all) they hold, but in some other form and alongside psychological criteria.

Then, what psychological criteria are applicable to decision method construction? We suggested the following criteria (Larichev, 1987):

(1) psychologically valid measurement of factors which are important for the decision;
(2) psychologically valid way of eliciting information in the construction of a decision rule;
(3) possibility of checking the decision maker's consistency;
(4) possibility of getting explanations.

Let us consider the criteria in more detail.

MEASUREMENT

Normative decision techniques are dominated by quantitative methods measuring multiple criteria on each decision alternative. This is no accident, since the fundamentals of quantitative measurements have been developed in books and papers on utility theory. The theory maintains that the assessment of alternatives and decision making require measurement of probability and utility.

It suggests that both be measured in quantitative form, though human capabilities to do that have been somewhat doubted. Von Neumann and Morgenstern (1953) were fully aware that utility as a quantitatively measurable value could well be questioned. They understood that 'a direct sense of preference of some object or a set thereof as compared to others provides the necessary foundation'. However, they introduced an assumption, fundamental to their theory, that 'our individual can compare not only events but also combinations of events with assigned probabilities'. What was behind this assumption? It was in the first place an analogy with physical measurements (measurement of heat, length, light, etc.) where in the initial stages of human activity there were no methods of quantitative measurement, but they emerged later.

Accordingly, they make the following conclusion: 'even if utilities look rather non-quantitative today, the history and case of heat theory may repeat again, and with unpredictable consequences at that'. Decision theory development during the 35 years after the publication of the book *Theory of Games and Economic Behavior* indicates that the expectations of its authors were too optimistic. Both in the past and now we lack reliable methods of quantitative measurement of probabilities and utilities. What is more, we have a great deal of evidence of the unreliability of available methods for quantitative measurement of utility and probability.

There are two reasons for this phenomenon. First of all, many factors characteristic of decision making are of quite a different nature compared to physical values. Factors such as 'organization's prestige', 'originality of clothes fashion' and 'job attractiveness' are of extremely subjective character and the personal values of the assessor have a considerable impact on the measurement outcome.

Thus there is no general objective foundation with a chance of devising an objective measurement technique.

It is hard to imagine that any objective quantitative meters like a thermometer, common for all people, will ever be available for such notions, since one can measure them only through people (just imagine that we would like to measure objectively the quality of paintings or symphonies). In summing up much descriptive research (Johnson and Schkade, 1985; Slovic and Lichtenstein, 1971; Kaheman *et al.*, 1982), we are safe in saying that even for one subject there is no means for a reliable quantitative construction of a utility function, since there are no reliable methods of quantitative measurement of factors affecting decision making.

It is commonly agreed that for many years preceding the emergence of methods for reliable quantitative measurement of physical properties of objects people had carried out qualitative measurements. Today we may judge those techniques as too primitive, since much more reliable quantitative methods have been devised. However, the fact that qualitative (pre-quantitative) methods of measuring physical values did exist is not questioned. When qualitative methods were replaced by quantitative ones, a slighting attitude to the earlier ways of measurement emerged as something 'unscientific' or obsolescent. Advances in physics have led to the well-known pronouncement that everywhere science appears there appears a digit or a quantity.

Human decision making is accompanied by a host of factors affecting assessment and choice of alternatives. Each of the factors has its own language of measurement (of course, there are factors such as cost where a quantitative language of measurement is quite adequate).

Accuracy and reliability of measurement are rather essential for sound decision making, since the selection of best alternative depends on them. Naturally, a decision maker influencing the destinies of people is inclined to put questions in a language he/she understands and receive an answer in the same language. We believe that the only feasible method of measurement for the majority of factors is qualitative estimates located on ordinal scales.

VALID INFORMATION ELICITATION FOR CONSTRUCTING A DECISION RULE

We shall differentiate two types of measurement. Above, measurements of the main factors influencing decision were discussed. We shall call them the primary measurements. In some normative methods, primary measurements allow one to reach the final decision. For example, measurements of utility and probability in quantitative form allow one to calculate the expected utility of every alternative.

However, for a large majority of normative methods this is not enough. Some cognitive operations of information elicitation are needed for constructing a decision rule. We shall call them the secondary measurements. For example, one needs to measure criteria weights to decide on the form (additive or multiplicative) of a utility function (Keeney and Raiffa, 1976).

Normative decision methods set quite different requirements for man (e.g. assignment of criteria weights, determination of certainty equivalent for a lottery, etc.; build a probability distribution of this outcome, etc.). In meeting these requirements, an individual exercises various information-processing operations. These may be composite, i.e. incorporating others, or simple, elementary and non-decomposable into elementary operations. Take, for example, a problem of building a utility function for a single criterion exercised within the framework of multiattribute utility theory (MAUT) (Keeney and Raiffa, 1976). It involves a number of similar problems of finding a certainty equivalent for lotteries.

In analysing different normative techniques, one may distinguish three groups of information-processing operations: operations with criteria as items (e.g. ordering the criteria by importance); operations with one alternative assessment by many criteria (e.g. to compare two estimates on two criteria of one alternative); operations with alternatives as items (e.g. selecting the best alternative from several).

Let us refer to an operation as elementary if it cannot be fragmented into other, simpler operations relating to the units of the same group (i.e. to criteria, alternatives and alternative estimation by criteria).

We suggest the following approach to estimating the validity of different decision rule constructions: to collect psychological research data indicating how confident and reliable a human being is in exercising some or other information-processing operation. If the data can be collected, the psychological validity of this or that normative technique can be characterized in terms of the psychological validity of the integrated elementary operations of information processing. We classify elementary operations into the following categories.

(1) *Complex* (C) if psychological research indicates that in performing such operations the decision maker displays many inconsistencies and makes use of simplified strategies (e.g. drops a number of criteria).
(2) *Admissible* (A) if psychological research indicates that the decision maker is capable of performing them with small inconsistencies and using complex strategies (e.g. using combinations of criteria estimates).
(3) *Admissible but for small dimension* (ASD) if there are data indicating that given a small number of objects (criteria, alternatives, multiattribute estimates) man performs the operation rather confidently but as the quantity thereof enlarges it becomes increasingly difficult.

Table I

No. of operation	Name of elementary operation	Evaluation
01	*Operations with criteria as items*	
011	Ordering with respect to utility (value)	A
012	Assigning quantitative criteria weights	C
013	Decomposition of complex criterion into simple ones	C
02	*Operations with separate alternative assessments by criteria*	
021	Assigning a quantitative equivalent for qualitative estimate by a criterion	UC
022	Determination of quantitative equivalent of a lottery	C
023	Qualitative comparison of two estimates taken from two criteria scales	A
024	Determination of quantitative trade-off value for two criteria estimates	UC
025	Determination of a satisfactory level by one criterion	UA
026	Nomination of probability for criteria estimate	C
03	*Operations with alternatives as items*	
031	Comparison of two alternatives viewed as a set of estimates by criteria and selection of the best one	ASD
032	Comparison of two alternatives viewed as something whole and selection of the best one	UA
033	Nomination of probabilistic estimates of alternatives	C
034	Attribution of alternatives to decision classes	ASD
035	Quantitative estimation of utility	C
036	Decomposition of complex alternatives into simple ones	C
037	Qualitative comparison of the probabilities of two alternatives	A

(4) *Uncertain* (U, UC, UA) if not enough psychological research on these operations has been conducted. However, one may be able to make a preliminary conclusion on admissability (UA) or complexity of the operation (UC). Description of the three groups of evaluation operations with their estimates is given below.

It is easy to see that any elementary operation must fall into one of the four categories as set out in Table I.

Let us describe each elementary operation in more detail.

Operations with criteria as items

Operations 011 and 012 constitute measurements of the criteria ranking for the decision maker. There are papers (Nikiforov *et al.*, 1984; Larichev *et al.*, 1980) which demonstrate general consistency of subjects in ranking criteria. Given seven criteria with binary estimates, the subjects do rank criteria in a relatively stable and consistent manner (Larichev *et al.*, 1980). The subjects consistently rank the criteria most important for them, though allow for exchange of secondary criteria ranks (Nikiforov *et al.*, 1984). Several papers (Fischer, 1991; Stuwart and Ely, 1984; Weber *et al.*, 1988) have been published in recent years which indicate, however, that subjects make essential errors in quantitative measurement of criteria weights. Indeed, the quantitative measurement of weights is an operation to which humans are not accustomed. Man hardly realizes the consequences of its implementation (e.g. can an insignificant variation in criterion weight lead to the choice of some other alternative?). Although this operation is present in many normative techniques (and sometimes viewed as 'natural'), the research of recent years shows that the weights assigned by the subjects are not reliable and stable information.

Operation 013 was studied in building the criteria hierarchy frequently employed in MAUT. The results indicate (Borcherding and Winterfeldt, 1988; Weber *et al.*, 1988) that given a considerable number of criteria, decomposition is not an operation resistant to decision maker's errors. At the same time, decomposition of a complex criterion into two or three obvious subcriteria (from the standpoint of meaning) may, we believe, be carried out quite reliably.

Operations with separate alternative assessments by criteria

Operation 021 is, in our opinion, an invalid replacement of qualitative notions on scales by arbitrarily assigned numbers. Although we lack data of detailed psychological research, this operation seems difficult for the decision maker. Just as in assigning quantitative weights, an insignificant modification of numbers can affect the ratio of alternatives.

The reliability of preference measurement with lotteries was studied in depth by Dolbear and Lave (1967) and produced negative results. Operation 023 was subjected to a systematic study in the course of ZAPROS method development. The results indicate that the decision maker performs this operation successfully with a small number of inconsistencies (Larichev *et al.*, 1978).

Operation 024, i.e. quantitative determination of changes in a single criterion estimate, is equivalent to modification of the estimate of the other. Unfortunately, no systematic verification of the reliability of this operation has been carried out. The point here is again about measurements difficult for a human being. It is safe to say that operation 025, as indicated by a number of studies, is a routine operation involving the replacement of the criterion by constraint. It is typically exercised while looking for satisfactory values and satisfactory (but not optimal) decisions (Kaheman *et al.*, 1982).

Operations with alternatives as items

Operation 031 was systematically studied by Russo and Rosen (1975), who show that this operation is difficult for a human being. Even with four criteria, people use simplifying heuristics which may lead to errors and inconsistencies. It is worth recalling that it was precisely the pairwise comparison of criteria estimates that allowed Tversky (1969) to build a 'money pump', i.e. a transitive ring of alternatives. At the same time, given two or three criteria, the pressure on short-term memory in exercising this operation is insignificant, which makes it possible to expect rather stable comparisons. We treat the problem of the pairwise comparison of two whole images of alternatives (e.g. presented in the form of pictures, slides, etc.) somewhat differently. We think that the decision maker develops a 'gestalt' of alternatives which essentially increases the reliability of such comparisons. Unfortunately, we are unfamiliar with any systematic studies on the reliability of such comparisons. Operation 033 urges the decision maker to generate probabilistic estimates. Kaheman *et al.*'s (1982) experiments have convincingly shown the unreliability of such estimates.

Operation 034 is employed in problems of direct classification, when a person places the alternative, described as a combination of criteria estimates, into one of several ordered decision classes. We conducted a special study of human capacities in handling such problems (Larichev *et al.*, 1988). It revealed that the reliability and stability of such information are dependent upon the number of criteria, the number of estimates on scales and the number of criteria classes. The parameter value domain under which people manage this problem was determined. Beyond this domain, human behaviour changes drastically—the number of inconsistencies amasses and simplifying heuristics emerge.

Operation 035 is accompanied by a quantitative measurement of the utility of alternatives. The quantitative estimates of utilities of various alternatives typical for decision-making problems (ecological projects, scientific research, etc.) are unreliable (Johnson and Schkade, 1985) owing to the inadequate language of measurement (see above). Recent works on the study of decomposition of a complex event into simple ones (Borcherding and Winterfeldt 1988) ascertain that operation 036 is reliable only when decomposing into a small number of relatively clear-cut events. There are no special devices for the systematic control of completeness of enumerating events during decision tree and fault tree construction. Note that this operation needs additional psychological research.

Operation 037 is a reliable one (Huber and Huber, 1987); in fact, children as well as adults do this operation in a reliable way.

The operations found in Table I are elementary. They may form more complex ones though, and evaluations of simple operations may be transferred to evaluations of complex ones. Thus, for example, the choice of a best alternative of a group may be represented as a combination of operations of pairwise comparison of alternatives (031) on the respective estimates. An operation of identifying the most unsatisfactory alternative estimate by criterion may be represented as a combination of operations 023, i.e. is admissible.

Evaluation of elementary operations allows us to characterize the psychological validity of one or other normative technique. This approach was employed by Larichev *et al.* (1987) in analysing several multiattribute linear programming techniques. Note that a mere seven out of 20 techniques considered are based on admissible operations.

Let us sum up. There are two main reasons according to which elementary operations of information processing come to be difficult for a human being.

The first reason, as mentioned above, is inadequate primary or secondary measurements, i.e. attempts to use numbers in cases when man confidently employs only several qualitative categories. We stress once again: the assignment of numbers to verbal ordered estimates reduces the reliability of measurements because of two causes. First of all, by assigning numbers (points, percentages, etc.), man constructs a subjective quantitative scale which can never be precise. By rating a qualitative estimate as 0.6 rather than 0.65, man can influence the final ratio of alternatives, i.e. decision. Secondly, man is unable to foresee how the numbers will be used in future (and they are used quite differently by different normative methods). It is the inadequate measurements that underlie difficulties with the elementary operations 012, 021, 022, 024, 026, 033 and 035.

The second reason is directly related to the limits of short-term (working) memory. Operations 031 and 034 can be used with qualitative variables received by psychologically valid primary measurements. Operations 031 and 034 make it necessary to keep in working memory the criterion alternative estimates, as well as the decision classes for operation 034. It is precisely because of the limited capacity of the short-term memory that these operations are performed rather reliably under small values of some variables and unreliably given large values (see Larichev *et al.*, 1988).

Which elementary operations can then be considered as valid and admissible? They are far from many and all of them, except for 025, are qualitative. We may somewhat reliably use operations 011, 023, 032 and 057. We may, within some limits, use operations 025, 031, 034 and 036. The admissible operations reduce to qualitative comparisons (of the type 'better', 'worse', 'approximately equal') of criteria, pairs of estimates on two criteria scales and holistic alternatives. Also, we may assign satisfactory values and exercise a simple decomposition of criteria and alternatives. Given a small number of criteria, we may compare two alternatives. With not too large a number of criteria, decision classes and divisions on scales we may refer alternatives to decision classes. All this together seems to be an essential constraint for an analyst working on prescriptive techniques. However, psychologically valid decision methods can be developed.

CONSISTENCY TEST

One of the inherent characteristics of human behaviour is error. In transmitting and processing information, people make errors. They make fewer and even considerably fewer errors when using the valid information elicitation procedures described above, but all the same they do make errors. The latter may be caused by the distraction of human attention, a person's fatigue, or other reasons. Errors are observed both in practice and in psychological experiments.

In other words, an individual can make unavoidable errors from time to time. Hence information obtained from a person must be subjected to verification rather than be used uncontrollably.

What are the methods of human information verification? It is much easier to reiterate the question to the person. Even a simple question such as 'are you sure of your last answer?' frequently makes the person more attentive in a man-machine dialogue. On the other hand, a question put for the second time in a literal manner may force man to automatically repeat the last erroneous answer. Therefore various psychological experiments repeat questions at different points of the query differentiated in time (e.g. the first question is repeated following the 20th, etc.). Clearly, this method of control has its own shortcomings, notably a large number of questions put to the person and no full guarantee of getting two infallible answers to one and the same question. Given a comparatively short succession of questions, the decision maker is capable of remembering the answer to the question put earlier.

Much more efficient means are the closed procedures (Larichev *et al.*, 1978; Larichev and Moshkovich, 1991) under which the earlier collected information is subjected to an indirect rather than direct test. The questioning procedure is built so that the questions are duplicated, but the duplication is exercised implicitly through other questions logically associated with the former. Here is an example. We are willing to compare four values: A, B, C and D. The easiest way of solving this task is to perform a successive pairwise comparison. Let the comparison produce $A>B$, $B>C$ and $C>D$, which allows us to order the values as $A>B>C>D$.

There can be errors in decision maker's responses, however. It seems more logical, therefore, to compare each value with each of the others: A with B, C and D; B with C and D; C with D.

Such a procedure will to some extent reduce the probability of essential errors in ordering. Let the person in the first round make an error in comparing C and D ($D>C$). (We simply did not compare B and D.) Now, under the closed procedure, let us get $C>D$ (the same wrong answer) but $D>B$—this makes us doubt the correctness of the produced answer, disclose the logical contradiction of the form $B>C>D>B$ and give the person an opportunity to analyse it and find consistent answers. Note that pinpointing a logical inconsistency must, in general, lead not to the automatic exclusion of an error but to the creation of premises for a logical analysis.

So, what is a proper way of building closed procedures for psychologically valid information-processing operations? For any operations concerned with comparison of several values we may introduce additional objects of comparison. For operation 011 it is easy to arrange comparison of each criterion with every other and analyse the inconsistencies if they emerge in the course of ordering the criteria on the basis of this comparison. In a similar way operation 032 can be checked. The following closed procedure was first suggested for operation 023 by Larichev *et al.*, 1978. Let there be N criteria with ordinal scales with a small (two to five) number of estimates. It is necessary to order the estimates of all criteria, i.e. place them on a single ordinal scale. To do this, it was suggested to exercise pairwise comparisons of criteria scales. Note that as the number of criteria increases (i.e. as the problem complexity grows), the abundance of information produced by such comparison potentially increases. A closed procedure of this type has been successfully employed in practical applications of the ZAPROS method.

The information produced by operation 034 was verified with a Pareto dominance relation. For alternative A1, referred by the decision maker to a certain class of decisions, a Pareto dominance cone was built in the criteria space, where the alternatives that happened to be within this cone must belong to a class not worse than A1. The violation of this condition produced a contradiction offered to the decision maker for analysis. This type of verification procedure was used in the CLASS method (Larichev and Moshkovich, 1990). Verification of information generated by operation 031 may be arranged in two ways. The first one requires utilization of operation 023, i.e. comparison of modifications in a pair of criteria estimates with a view to determining the supremacy of one alternative over the other. (Note that in this case the alternatives are not always comparable and the incomparability may turn to be one of the results.)

The second approach may be applied along with the first one when there are several alternatives. In this case it is possible to try to arrange a closed procedure by pairwise comparison of alternatives. It is sometimes reasonable to introduce an extra alternative (which is not in the dominance relation with the two basic ones) whose comparison provides an additional verification of the decision maker's preferences. We do not know the way of arranging verification of decomposition operations 013 and 036; therefore we believe it makes sense to use them when decomposition is meaningful and the number of the next-level elements is small.

GENERATION OF EXPLANATIONS

From a behavioural point of view, one of the requirements of the application of any method is its explainability. In making a crucial decision, the decision maker would like to know why alternative A turned out better than B, and both of them are better than C. This decision maker's requirement is quite reasonable. The stage of information elicitation from the decision maker (measurements) and the stage of final results presentation are separated by a stage of information transformation. Understandably, the decision maker wants to be sure that it is precisely his/her own preferences without any distortions that are behind the assessment of alternatives. In order to meet this requirement, the decision method must be 'transparent': it must be conducive to finding an unambiguous correspondence between the information elicited from the decision maker and the final evaluations of alternatives. Only then does there appear an opportunity to obtain explanations by the decision maker.

The possibility of obtaining explanations in a natural language is one of the characteristics of expert systems. This characteristic allows one to make expert systems user-friendly.

In decision problems it is the decision maker who is the user and whose preferences provide a foundation for the decision rule. He/she is also eager to get explanations, however, since the information he/she generates relates to partial elements of the problem (separate estimates on criteria scales, individual alternatives) and the results of the method application are ordinarily of a more general nature.

It is easy to see that this decision maker's requirement cannot be met by axiomatic approaches where he/she must take for granted not only the axioms but also the built-in information transformations (lotteries, weight multiplication by criteria estimates, etc.). Just inadequate from the standpoint of this requirement are some heuristic methods.

However, the psychologically valid techniques dealt with in this paper meet the explanations requirement. Under the valid measurement of variables an explanation is possible and appears as follows: 'such and such relationships hold on the basis of such and such information elicited from the decision maker and tested for consistency'.

CONCLUSIONS

Axiomatic methods begin with proof of existence of a utility function when certain axioms are satisfied and certain measurements are assumed to be possible. From the point of view of our approach one cannot begin with this. The measurements are unreliable and the axioms do not hold. Completeness cannot be required because incomparable opinions obviously exist; they are incomparable because no valid ways are available to obtain data essential for comparison. The transitivity axiom cannot be used, since people are not necessarily transitive. The continuity assumption is useless for us, since there are sensitivity thresholds in people.

We intend to develop psychologically valid methods leading to intelligent decisions.

There are three main problems in decision making: (1) to order alternatives; (2) to classify alternatives into decision classes; (3) to select one best alternative.

Three psychologically valid methods have been developed for the problems. The ZAPROS method (Larichev *et al.*, 1978; Larichev and Moshkovich, 1991) allows one to construct a quasi-order of alternatives using operation 023. The CLASS method (Larichev and Moshkovich, 1990) allows one to construct a classification of the alternatives using operation 034. The ASTRIDA method (Berkeley *et al.*, 1990; Moshkovich, 1991) is the first attempt to select one best alternative using different reformulations of the problem and the operation 023. All methods provide the means for consistency testing and the possibility of getting explanations.

The ZAPROS and CLASS methods have been applied successfully (Zuev *et al.*, 1979; Larichev *et al.*, 1989).

The practical application of the above techniques indicates that they have much more chance of success in organizations as applied to real decision-making procedures.

ACKNOWLEDGEMENT

The author would like to express his gratitude to Professor Rex Brown for detail reading and valuable suggestions.

REFERENCES

Berkeley, D., Humphreys, P., Larichev, O. and Moshkovich, H., 'Aiding strategic decision making: derivation and development of ASTRIDA', in Vecsenyi, Y. and Sol, H. (eds), *Environment for Supporting Decision Processes, Proc. IFIP TC-8.3 Conf.*, Budapest, 1990.

Borcherding, K. and Winterfeldt, D., 'The effects of varying value trees on multiattributive evaluations', *Acta Psychol.*, **68**, 153-170 (1988).

Brown, R. V., 'Toward a prescriptive science and technology of decision aiding', *Ann. Oper. Res.*, **19**, 467-483 (1989).

Dolbear, F. I. and Lave, L. B., 'Inconsistent behavior in lottery choice experiments', *Behav. Sci.*, **12**, 14-23 (1967).

Fischer, G., 'Range sensitivity of attribute weights in multiattribute utility assessment', Duke University, 1991 (unpublished).

Fishburn, P. C., *Decisions and Value Theory*, New York: Wiley, 1964.

Hobbs, B., 'What can we learn from experiments in multiobjective decision analysis', *IEEE Trans. Syst., Man, Cyber.*, **SMC-16**, 384-394 (1986).

Huber, B. L. and Huber, O., 'Development of the concept of comparative subjective probability', *J. Exp. Child Psychol.*, **44**, 304-316 (1987).

Johnson, E. and Schkade, D., 'Bias in utility measurements: further evidence and expectations', *Manag. Sci.*, **36**, 406-424 (1985).

Kaheman, D., Slovic, P. and Tversky, A. (eds), *Judgement under Uncertainty: Heuristic and Biases*, Cambridge: Cambridge University Press, 1982.

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

Larichev, O. I., *Objective Models and Subjective Decisions*, Moscow: Nauka, 1987 (in Russian).

Larichev, O., Boychenko, V., Moshkovich, H. and Sheptalova, I., 'Modelling multiattribute information processing strategies in a binary decision task', *Organiz. Behav. Human Perform.*, **26**, 278-291 (1980).

Larichev, O. and Moshkovich, H., 'Decision support system CLASS for R and D planning', *Proc. First Int. Conf. on Expert Planning Systems*, Brighton, 1990, *Conf. Publ. 322*, The Institute of Electrical Engineers, 227-232.

Larichev, O. and Moshkovich, H., 'ZAPROS: a method and system for ordering multiattribute alternatives on the base of a decision-maker's preferences', *Preprint*, Institute for Systems Studies, Moscow, 1991.

Larichev, A., Moshkovich, H. and Rebric, S., 'Systematic research into human behavior in multiattribute object classification problems', *Acta Psychol.*, **68**, 171-182 (1988).

Larichev, O., Polyakov, O. and Nikiforov, A., 'Multicriterion linear programming problems (analytic survey)', *J. Econ. Psychol.*, **8**, 389-407 (1987).

Larichev, O., Zuev, J. and Gnedenko, L., 'The method ZAPROS (closed procedures near reference situations) for the analysis of variants of complex decisions', in Emelyanov, S. (ed.), *Multicriteria Choice in Solving Ill-Structured Problems, VNIISI Proc.*, 1978 (in Russian).

Larichev, O., Moshkovich, H., Drozdova, I., Starvinsky, V. and Kovalev, B., 'The system of estimation of research in health care', *Sov. Med.*, N1, 32-37 (1990).

Moshkovich, H., 'Qualitative methods in decision making', *Proc. 13th Conf. on SPUDM*, Fribourg, Fribourg University, 44-79, 1991.

Nikiforov, A., Rebrik, S. and Sheptelova, L. P., 'Experimental investigation of stability preference in some problems of decision making', *VNIISI Proc.*, Vol. 9, 69-80, VNIISI Publ., Moscow, 1984 (in Russian).

Russo, J. E. and Rosen, L. D., 'An eye fixation analysis of multialternative choice', *Memory Cognit.* **3**, 267-276 (1975).

Slovic, P. and Lichtenstein, S., 'Comparison of Bayesian and regression approaches to the study of information processing in judgement', *Organiz. Behav. Human Perform.*, **6**, 649-744 (1971).

Stuwart, T. R. and Ely, D. W., 'Range sensitivity: a necessary condition and a test for the validity of weights', paper prepared for *Multiple Criteria Decision Making Conference*. Cleveland, OH, 1984.

Tversky, A., 'Intransitivity of preferences', *Psychol. Rev.*, **76**, 31–48 (1969).

Von Neumann, J. and Morgenstern, O., *Theory of Games and Economic Behavior*, Princeton, NJ: Princeton University Press, 1953.

Weber, M., Eisenfuhr, F., and von Winterfeldt, D., 'The effects of splitting attributes on weights in multiattribute utility measurement', *Manag. Sci.*, **34**, 431–445 (1988).

Zuev, Y., Larichev, O., Filippov, Chuev, Y., 'The problem of R and D projects estimation', *Bull. USSR Acad. Sci.* **8**, 29–39 (1979) (in Russian).