International Journal of Information Technology & Decision Making
Vol. 2, No. 1 (2003) 29–40

HUMAN BEHAVIOR IN A MULTI-CRITERIA CHOICE PROBLEM WITH INDIVIDUAL TASKS OF DIFFERENT DIFFICULTIES

E. M. FUREMS,* O. I. LARICHEV† and G. V. ROIZENSON‡

Institute for Systems Analysis, Russian Academy of Sciences
9, pr. 60 let Octjabrja, Moscow, 117312, Russia
Tel: 8-902-172-44-54
**furems-em@mtu-net.ru*
†oil@isa.ru
‡rgv@isa.ru

A. V. LOTOV

Dorodnicyn Computing Center, Russian Academy of Sciences
40, Vavilova str., Moscow, 117967 Russia
lotov@ccas.ru

K. MIETTINEN

University of Jyvaskyla Department of Mathematical Information Technology
P. O. Box 35 (Agora), FIN-40351 Jyvaskyla, Finland
miettine@mit.jyu.fi

This paper is devoted to a laboratory study of human behavior in a multi-criteria choice problem. The specific feature of the experimental study is the creation of an individually adjusted instance of a general task for each subject in accordance with his/her preferences over each criterion. Human behavior is studied in a specially constructed choice situation based on the decomposition of the alternatives of a multi-criteria problem. The procedure is based on multiple steps of pair-wise comparisons involving only some (two or three) of the original components of the alternatives. Abilities of subjects to use such comparisons and to answer the questions in a logical way are tested. The experiment was carried out in two countries: Finland and Russia.

Keywords: Human behavior; multi-criteria problems; qualitative comparison; stability of preferences; individually adjusted instance of a general task.

1. Introduction

The problem of choosing the best alternative from a small set of alternatives is a typical human activity. Examples of such problems include the selection of a good by a consumer, choosing an apartment for tenancy or purchasing, etc. While solving such problems people consider the different features of the alternatives by taking into account a number of aspects (choice criteria) significant to them.

Many business and personal problems may be defined as ill-structured ones.[9] The presence of multiple choice criteria as well as verbal description of both alternatives and criteria is inherent to such problems. In this paper, we consider human behavior in ill-structured problems. To what extent are human beings consistent and reasonable in such decision-making? This question has been studied in many papers, for example, in Russo and Rosen,[7] Montgomery and Svenson,[5] Payne *et al.*,[6] Larichev,[2] Korhonen *et al.*[1] The results of descriptive studies may be summarized as follows: the multi-criteria comparison of alternatives and, in particular, the choice of the best alternative, is difficult for a human system to process information, and the more criteria, the more complexity. While solving such problems, human beings make errors as well as use simplifying strategies to adapt the problem to their capabilities.

Experiments show that the comparison of alternatives that differ in only two components is relatively easy for human beings. For example, such a technique of obtaining information is the basis of the method ZAPROS.[3] When the alternatives are described with ordinal criteria with 3–4 possible verbal values each, individuals make 1–2 contradictions from 50–60 pair-wise comparisons of two criteria combinations. In contrast, the abilities of human beings in comparing alternatives that are different in three or more (verbal) components are not sufficiently known.

As far as a human information processing system is concerned, H. Simon[9] noted: "The evidence is overwhelming that the system is basically serial in its operation: that it can process only a few symbols at a time and that the symbols being processed must be held in special, limited memory structures whose content can be changed rapidly." The strongest restrictions on the ability of humans to find effective strategies are caused by the fact that the capacity of the short-term memory structure is very small. Due to such a feature of a human information processing system, it should be expected that an individual most likely would, firstly, pay attention to a small number of components, and then involve other components one by one.

Various techniques have been proposed that take the limited capacity of the short-term human memory into account. For example, the so-called "phase narrowing" techniques[4] are an example of adapting experimental situations to the features of human behavior. It is known that when an individual has to choose the best alternative from a set, he/she narrows such a set step-by-step to determine preferable subsets of alternatives. That is why in "phase narrowing" experiments subjects were assigned to choose, firstly, the best 6 alternatives from 20, then the best 3 from 6 and, at last, 1 from 3.

In our experimental study, we examine a different technique that also takes into account the restricted abilities of human beings. A multi-criteria choice problem is studied where the alternatives are given by verbal statements and each alternative involves more than two criteria (four criteria, in our case). It is known that the qualitative comparison of alternatives involving two criteria is a relatively easy operation for a human being.[2] The method under study is based on reducing the number of criteria in human comparisons. We assume that the problem has four criteria. Subjects have first to compare parts of the original alternatives involving two criteria only (criterion dyads). In a criterion dyad, the verbal values of only two criteria of an alternative are displayed, while the values of the other two criteria are hidden. As a result of such comparisons, feasible criterion dyads are ordered. Usually, these comparisons do not provide sufficient information for ordering the original alternatives. For this reason, a second stage is needed. In the approach suggested, the next step is to compare parts of the original alternatives involving three criteria (criterion triads). Finally, a comparison of the original (full) alternatives with four criteria may be needed, but the number of such comparisons is small.

The questions that are posed to a subject in the first stage (two-criterion comparison) depend on the ordering of criteria. Different criterion orders can be used. Does the criterion order influence the choice? To answer this question, two rounds of the procedure were carried out that were based on two different criterion orders. The second question is related to the ability of human beings to use comparisons of triads supported by the specially prepared graphic tool.

It should be noted that in many psychological experiments the same task is presented to all the subjects to study human behavior in a multi-criteria choice problem. However, such a task may be easy to one subject (from the point of view of his/her preferences) and complex to another. We believe that it is of great importance to give each subject an individually adjusted instance of a general task, taking into account his/her desirable levels upon some criteria, communicated by him/her preliminary.

In Sec. 2, the problem is described and the method for the generation of individually adjusted instances is introduced. Section 3 deals with the organization of the experiment. In Sec. 4, the experiment and its results are described. Section 5 contains a discussion of the results.

2. The Problem and Developing Individually Adjusted Instances of A General Task

For our experiments, we selected the problem of choosing a part-time job by a student to be engaged concurrently with his/her studies. *Salary*, *Work time*, *Position* and *Time to workplace* were used as criteria. Such a problem is typical for both Finnish and Russian students involved in our experiments as subjects.

Our experiments started with the development of individually adjusted instances of the general task. Initially, a student was presented a description of the general task (choosing a part-time job), and was prompted to enter a desirable salary and a time period of his/her study in the university, which he/she preferred not to omit. Then, a set of alternatives was generated individually for each student. He/she had to order them with respect to each criterion according to his/her preferences, excluding the *Salary* (obviously, more salary is preferable to anybody). Let us note, that the number of values on the *Salary* scale could be different for different students depending on the number of alternative jobs in his/her individually adjusted instance of the task.

It is clear that each student may have preferences of his/her own with respect to the criteria *Position*, *Time to workplace* and *Work time*. For example, one student specified the following order for the criterion *Position* (from the best to the worst): Bank employee, Private tutor, Teacher of mathematics in the lyceum, System programmer, Security. The order specified by another student with respect to the same criterion was as follows: Teacher of mathematics in the lyceum, Private tutor, System programmer, Bank employee, Security. Moreover, as far as the criterion *Time to workplace* was concerned, one student preferred 30 minutes by bicycle over 15 minutes by bicycle and explained it by the lack of spare time for other kinds of exercise. Here, we gave an opportunity for each student to create personal ordinal scales on the criteria, arranging the alternatives from the best to the worst.

An individually adjusted problem was generated for each student on the basis of his/her desirable levels upon the criteria *Salary* and *Work time* as well as taking into account his/her orders for the components of the alternatives with respect to the criteria *Position* and *Time to workplace*, respectively.

We generated the individual sets of jobs using the criteria *Salary* and *Work time* as a basis to creating tasks of two levels of difficulty:

(a) Difficult Choice (DC) — Initially, a salary that was more than, equal to or less than the desirable one, was assigned to each job. Then the following principle was used to specify a work time to each job: the bigger salary, the more overlapping of work time and study time and, accordingly, the less time to study. Thus, each student was put in a situation of a difficult choice between lecture and/or seminar attendance and additional payment.

    Finally, each job created as above was assigned the values upon the criteria *Position* and *Time to workplace*. They were generated on the basis of the principle: the better position to a student, the longer the time to get to a workplace was. These values were assigned to the jobs in an arbitrary order.

(b) Moderately Difficult Choice (MDC) — In addition to the jobs created according to DC, we created jobs that had the best value upon either of the two criteria *Salary* or *Work time* and the second best value on the other, or the best values upon both the criteria.

We used only non-dominated alternatives in each individually adjusted instance of the general task. Each instance consisted of 4 to 10 alternative jobs. It should be noted that the students did not know what kind of tasks (DC, MDC) they had to solve.

3. Organization of the Experiment

The students solved their individually adjusted problems twice, in two rounds. Each round consisted of several stages. In the first stage, parts of the existing alternatives involving only two criteria (criterion dyads) were presented. Correspondingly, the second stage involved three-criterion parts of the alternatives (criterion triads), etc. Thus, the difference between the stages was the number of criteria to be taken into account. Two different orders of criteria were specified for the two rounds. In the first round we presented the criteria in the order (1) *Position*, (2) *Time to workplace*, (3) *Work time* and (4) *Salary*. In the second round, we used the order (1) *Salary*, (2) *Work time*, (3) *Time to workplace* and (4) *Position*.

The two rounds were needed to check the following hypothesis: the subjects could make the same choice in both the rounds.

In order to select the best alternative, each student had to carry out a series of comparisons between the alternatives available. During the first stage, each student was presented all feasible dyads, except for those that could automatically be ordered on the basis of ordinal scales of the criteria. Each time, while making the comparison, the student could answer in one of the following ways:

- I prefer the first dyad to the second one, $(1 > 2)$
- I prefer the second dyad to the first one, $(1 < 2)$
- I am indifferent to both dyads, $(1 = 2)$
- I do not know.

Let us note that the answer "I do not know" was available only for the sake of completeness of the outcomes of the comparison. It was used by students only in 0.8% of the cases.

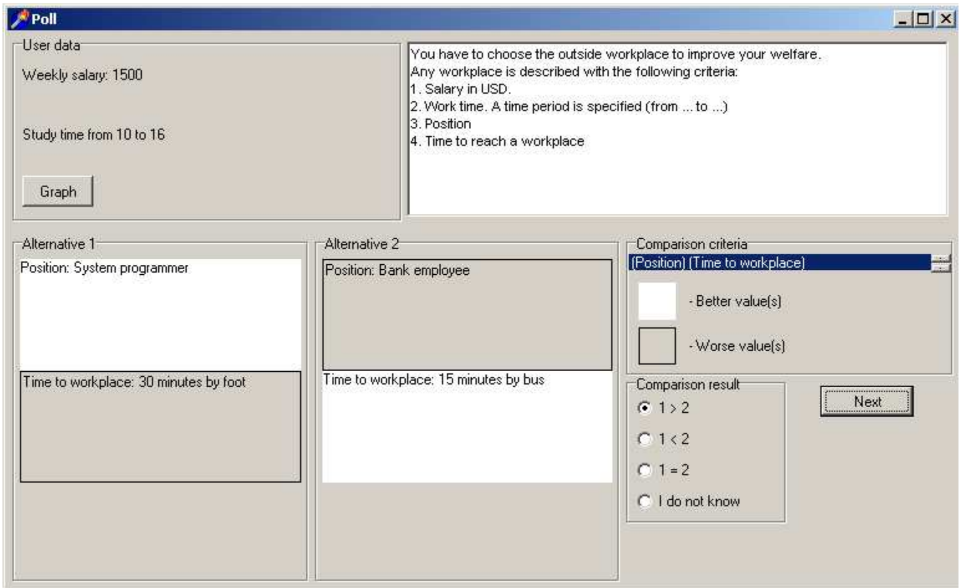

Fig. 1. Example of comparing a pair of feasible criterion dyads.

In Fig. 1, a display is provided that helps a student to compare a pair of feasible criterion dyads upon the criteria *Position* and *Time to workplace*. The better values (in accordance to ordinal scales) are presented on a white background and the worse values on a gray one.

As a result, each student produced an order of all feasible criterion dyads of existing alternatives. There were cases when such information was sufficient to order several pairs of full alternatives. Sometime it was even possible to identify the best alternative.

In most of the cases, however, two-criterion comparisons were not sufficient to identify the best alternative. Therefore, the procedure moved to the next stage. In the second stage, a student had to compare all possible pairs of criterion triads (except for those that could be automatically ordered on the basis of the results of two-criterion comparisons). One of the displays is given in Fig. 2 where a student is asked to make a comparison of a pair of feasible criterion triads.

Figure 2 gives an example of the display that supports the comparison of a pair of criterion triads having different values in criteria *Salary*, *Position* and *Time to workplace*. To simplify the comparison, information about the result of the comparison carried out on the previous stage (comparing criterion dyads) is used here. The three criteria were divided into two groups: (*Salary & Position*) and (*Time to workplace*). Let us note that a criterion triad may be split into a criterion dyad and a one-criterion part in three different ways. In Fig. 2, one of the possible splittings is shown. Since this particular student decided on the previous stage that

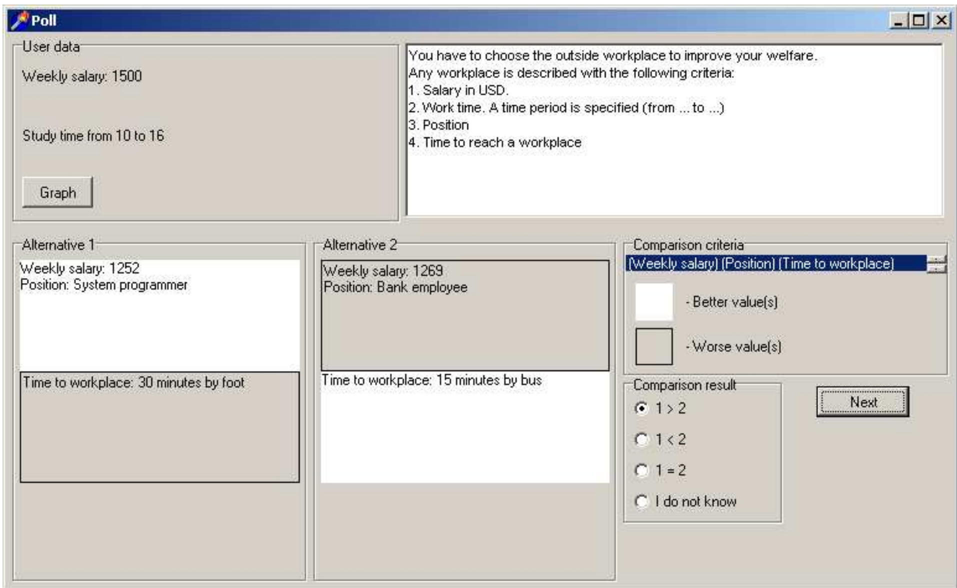

Fig. 2. Example of comparing a pair of feasible criterion triads.

the criterion dyad "System programmer, 1252 USD" is better, than the criterion dyad "Bank employee, 1269 USD", the corresponding combinations of components are displayed in the dialog window on a white and on a gray background, respectively. The preferred value of the criterion "*Time to workplace*" is displayed on a white background. It is important that all possible splits of the criterion triad into a criterion dyad and a one-criterion part were given to the student to check the consistency of the answers.

The results of the second stage gave an opportunity to order all the feasible criterion triads. If the information obtained on this stage was sufficient to identify the best alternative, the procedure was completed. In the opposite case, it was moved to the final stage.

The final stage is the comparison of the pairs of full alternatives involving all the four criteria in order to identify the most preferred alternative. However, due to the previous comparisons of all the feasible criterion dyads and triads, the required number of comparisons of full alternatives was fairly small. In most of the experiments, it was not needed at all. The comparisons of full alternatives were supported by computer displays based on the same idea of splitting the list of criteria as in the case of the comparisons involving three criteria. We do not describe them here in details. Note that each pair of alternatives with four criteria that required a comparison was presented to a student in four different ways.

In the framework of the procedure, it was possible to collect information on how successful and non-contradictory the students were in comparing the criterion triads. We wanted to check the following hypothesis: *human beings are able to compare consistently the criterion triads, presented as criterion dyads and one-criterion parts, when the information on ordering of criterion dyads and single-criterion parts is used.*

In contrast to the comparisons involving three criteria, only few comparisons of full alternatives were carried out. For this reason, we cannot provide any statistically sound result on human behavior in the four-criterion case. Nevertheless, some observations are possible.

4. Experiment and its Results

The experiments were conducted with students (13 persons) of the Department of Mathematical Information Technology at the University of Jyvaskyla in Finland and students (25 persons) of the Department for Computational Mathematics and Cybernetics at the Lomonosov Moscow State University in Russia. A computer program was developed to implement the procedure described earlier and to compute the statistics.

First of all, each student had to order the alternatives upon the criteria. On the basis of such orderings, the program constructed an individually adjusted instance of the general task to each student. The type of such individually adjusted task (DC or MDC) was assigned arbitrarily to each student. In the experiments, 24 students had to solve a DC task and 14 students an MDC task.

Then, each student had to solve his/her own individually adjusted task twice, in two rounds. Each round, as it was said earlier, consisted of two or three stages: comparisons of dyads on the first stage, comparisons triads on the second stage, and comparisons of full alternatives at the final stage (if needed). It is important that while presenting a pair of objects for a comparison, the software reminded a student, with the help of colors, which component or combination of components of the first object was better or worse than the corresponding component(s) of the second one (according to his/her previous answers). It turned out in the experiment that only 34 students were involved in the second stage. It means that only 34 students had to compare the criterion triads.

Let us return to the two questions formulated earlier.

Question 1: *Can the subjects make the same choice in both of the rounds*?

To answer the question, let us consider the results given in Table 1. The columns "Round I" and "Round II" contain the alternatives chosen by each student in the first and the second rounds, respectively. The criteria are arranged in the following order: *Salary*, *Work time*, *Position* and *Time to workplace* (notwithstanding different orders of criteria in the two rounds) and the figures represent the ranks of the components of each alternative according to the student's originally specified preferences (1 is the best and so on).

The column "Stability of choice" is related to the stability of the student's choices. In order to analyze the results obtained, we define stability characteristics in the following ways:

(a) the choice is claimed to be absolutely stable (AS) if the same alternative was selected in both the rounds;
(b) the choice is claimed to be semi-stable (SS) if the alternatives selected in the first and the second rounds have the first value(s) upon at least one of the criteria;
(c) the choice is claimed to be unstable (US) in all the other cases.

Item (b) requires some comments. When considering Table 1, the choice of the first student appears to be semi-stable, since the first value upon the first criterion is repeated in the second round (it can mean that the student was interested in salary both the times). The values upon the other criteria seem to be casual. The choice of the sixth student is semi-stable because of the same reasons.

As one can see from Table 1, the choices of 14 of the 38 students turned out to be absolutely stable (they selected the same job in both the first and in the second rounds) and the choices of 6 students were semi-stable. Thus, we can say that about one half of the students were stable in their choices.

To our opinion such an effect has the following reason. The first round seems to be used by students to study the problem, to develop a compromise between the criteria, and thereby to form a strategy of choice (say, to find a job with a salary as high as possible, or with the most convenient work time, or with the most preferable position, etc.). The presentation of the results of the first round provides

Table 1. Choices of the students.

#	Round I	Round II	Type of task	Stability of choice
1.	1, 1, 2, 4	1, 4, 5, 1	MDC	SS
2.	4, 2, 3, 3	4, 1, 5, 1	DC	US
3.	2, 1, 2, 4	1, 2, 5, 1	MDC	US
4.	2, 4, 2, 4	1, 3, 4, 2	DC	US
5.	1, 2, 1, 5	1, 2, 1, 5	MDC	AS
6.	3, 1, 1, 5	2, 2, 1, 5	MDC	SS
7.	2, 2, 3, 3	1, 4, 5, 1	MDC	US
8.	1, 1, 1, 5	1, 1, 1, 5	MDC	AS
9.	2, 1, 1, 5	2, 1, 1, 5	MDC	AS
10.	1, 3, 1, 5	1, 3, 2, 4	DC	SS
11.	1, 1, 5, 1	1, 1, 5, 1	MDC	AS
12.	2, 4, 1, 5	1, 3, 3, 3	DC	US
13.	2, 2, 1, 5	2, 2, 4, 2	MDC	US
14.	6, 1, 5, 1	1, 4, 3, 3	DC	US
15.	2, 3, 1, 5	1, 3, 2, 4	DC	US
16.	5, 2, 4, 2	1, 4, 2, 4	DC	US
17.	3, 1, 2, 4	1, 2, 3, 3	MDC	US
18.	1, 3, 3, 3	1, 3, 3, 3	DC	AS
19.	3, 1, 1, 5	1, 1, 2, 5	MDC	SS
20.	2, 1, 2, 4	2, 1, 2, 4	MDC	AS
21.	1, 3, 4, 2	1, 3, 1, 5	DC	SS
22.	1, 2, 2, 4	1, 2, 2, 4	MDC	AS
23.	6, 1, 2, 4	2, 2, 4, 2	MDC	US
24.	4, 1, 5, 1	1, 3, 4, 2	DC	US
25.	2, 3, 1, 5	2, 3, 1, 5	DC	AS
26.	1, 4, 1, 5	1, 4, 1, 5	DC	AS
27.	3, 2, 1, 5	2, 3, 2, 4	DC	US
28.	3, 2, 4, 2	1, 4, 1, 5	DC	US
29.	1, 4, 3, 3	1, 4, 3, 3	DC	AS
30.	4, 1, 3, 3	1, 4, 1, 5	DC	US
31.	2, 3, 3, 3	2, 3, 3, 3	DC	AS
32.	1, 4, 1, 5	1, 4, 1, 5	DC	AS
33.	4, 1, 2, 4	2, 3, 4, 2	DC	US
34.	1, 4, 4, 2	1, 4, 4, 2	DC	AS
35.	1, 4, 1, 5	3, 2, 1, 5	DC	SS
36.	2, 3, 1, 5	1, 4, 2, 4	DC	US
37.	4, 1, 4, 2	1, 4, 2, 4	DC	US
38.	1, 4, 3, 3	1, 4, 3, 3	DC	AS

a possibility to learn about the alternatives. The second round allowed then the students to apply the selected strategy or to confirm the previous choice in the case it satisfied them. The existence of several rounds coincides with real life decision-making where people often collect information, explore a problem and only after that make the final choice.

Question 2: *Can the subjects compare pairs of triads*?

To answer this question, we have to study the behavior of the students on the second stage where they had to compare parts of the alternatives involving three criteria. During the process of such a comparison, each student had to answer three different formulations of the same question. Among 34 students that had to compare three-criterion parts of the alternatives, 29 gave identical answers on the three different representations of the question. Thus, only 5 students failed to give an identical answer (it is interesting that one of them managed to give three different answers). When considering this result, we can say that most of the students are able to compare criterion triads. So, the second question can be answered positively.

However, it is possible to support this inference based on the common sense with the methods of mathematical statistics, to be more precise, methods of non-parametric statistics.[8] We test the hypothesis that the domination of students who managed to deal with the triad comparisons was obtained by a chance. Let P_1 be a probability of students' ability to compare criterion triads, and P_2 be a probability of students' inability to do so. Let us consider two hypotheses:

Hypothesis H_0: comparison of criterion triads is a difficult task to most of the students, which is given in mathematical form as $P_1 \leq P_2$, and the alternative

Hypothesis H_1: most of the students are able to compare criterion triads, i.e. $P_1 > P_2$.

In order to reject the hypothesis H_0 under significance $\alpha = 0.05$ and taking into account that 34 subjects were involved in comparing the triads, we obtain using the binomial statistical test (see Table A in Ref. 8) that the number of students that are able to compare the triads must not be less than 23. Since 29 subjects demonstrated this ability, we have to reject the hypothesis H_0 and to admit the hypothesis H_1, that is, $P_1 > P_2$. Thus, the statistical analysis proves that most of the students were able to use the comparison of triads.

5. Discussion

The main result of our study is related to the application of individually adjusted instances of a general task in psychological experiments. It was visible that all the students had to solve psychologically complicated problems in the process of comparing alternatives. While solving such tasks, the students made meaningful and multi-criteria choices. That is why we can recommend the application of individually adjusted instances for future research of human abilities.

Even though we do not suggest a Decision Support System (DSS) here, some of our results may be of interest in the process of designing them. An important finding of our experiment is that a properly supported comparison of parts of alternatives involving three criteria can be used in a DSS. This result is statistically significant and confirms the human ability to compare alternatives with three criteria in the case of proper graphic support based on comparisons involving two criteria completed earlier.

The third finding is the usefulness of applying two rounds in DSSs. Indeed, it seems to be reasonable to assume that the first round allowed a subject to study his/her problem, to find trade-offs, and to reach a meaningful choice. In the case of further experiments, a structure with three rounds could be reasonable.

It is interesting to note that in the second round, the comparison of dyads was sufficient for all the students to select the best alternative. This makes us suppose the presence of a learning effect.

Finally, we deem that it would be desirable to elaborate a DSS that could help a human being to solve the problems of a choice from a small number of alternatives. To our opinion, such a DSS must meet the following requirements:

(a) Qualitative technique for preference elicitation.
(b) On-line check of subject answers with respect to inconsistencies.
(c) Display of intermediate results to provide effective feedback.
(d) Arrangement using two rounds.

We expect the findings of our experiments to be useful in developing such a decision support tool.

Acknowledgment

The authors wish to express their gratitude to Mr. A. Kistanov for his help in organizing the experiments at the Lomonosov Moscow State University. The research was supported by the joint project #3 of the Academy of Finland and the Russian Academy of Sciences, by the Academy of Finland, grant #65760 and by the grants of the Russian Foundation for Basic Research 0015-96053 and 01-01-00514.

References

1. P. Korhonen, O. Larichev, H. Moshkovich, A. Mechitov and J. Wallenius, Choice behaviour in a computer-aided multiattribute decision task, *J. Multicriteria Decision Analysis* **6** (1997) 233–246.
2. O. I. Larichev, Cognitive validity in design of decision-aiding techniques, *J. Multicriteria Decision Analysis* **1**, 3 (1992) 127–138.
3. O. I. Larichev, Ranking multicriteria alternatives: The method ZAPROS III, *European J. Operational Research* **131** (2001) 550–558.
4. I. P. Levin and J. D. Jasper, Phased narrowing: A new process tracing method for decision making, *Organizational Behavior and Human Decision Processes* **64** (1995) 1–8.
5. H. Montgomery and O. Svenson, A think-aloud study of dominance structuring in decision processes, *Process and Structure on Human Decision Making*, eds. H. Montgomery and O. Svenson (J. Wiley and Sons, Chichester, 1989) 135–150.
6. J. W. Payne, J. R. Bettman, E. Coupey and E. J. Johnson, A constructive process view of decision making: Multiple strategies in judgment and choice, *Current Themes in Psychological Decision Research*, eds. O. Huber, J. Mumpower, J. van der Pligt and P. Koele (North Holland, Amsterdam, 1993) 107–142.

7. J. E. Russo and L. D. Rosen, An eye fixation analysis of multiattribute choice, memory and cognition **3** (1975) 267–276.
8. R. P. Runyon, *Nonparametric Statistics*: *A Contemporary Approach* (Addison Wesley, Reading, 1977).
9. H. Simon, *The Sciences of the Artificial* (The MIT Press, Cambridge, 1985).